A new family of tandem repetitive early histone genes in the sea urchin *Lytechinus pictus*: evidence for concerted evolution within tandem arrays

Curtis A.Holt and Geoffrey Childs

Departments of Genetics and Molecular Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Received 2 May 1984; Revised and Accepted 23 July 1984

ABSTRACT

We have isolated and characterized a third nonallelic tandemly arrayed histone cluster (LpE) from the sea urchin *Lytechinus pictus*. Although this tandem array is not intermingled with the other two early histone gene families also found in the *L. pictus* genome, the order and polarity of the five histone coding sequences in this family are the same as every other well characterized sea urchin early histone gene family. Heteroduplex analysis and restriction endonuclease mapping experiments indicate that the LpE family is more closely related to the B-C than the A-D family of early histone genes. Examination of several individual sperm DNA samples has revealed considerable polymorphism in each of the three tandem repeat families. Within an individual, however, each family is remarkably homogeneous. Thus, our results indicate that rapid fixation of variants acts to homogenize the members of a single tandem array at a considerably faster rate within a family than between families. However, at least some exchange of sequences between families is evident based on the conservation of many restriction endonuclease recognition sites and from analysis of a cosmid clone in which the A-D and E tandem repeats are found adjacent to one another. These differences in the rate of fixation of variants within and between these families are likely to be responsible for the maintenance of diversity between the different families.

INTRODUCTION

The sea urchin genome contains several multigene families that are regulated in a developmental and tissue specific fashion. One of these families, the "early" histone genes, is arranged in 300-500 tandem repeats (1,2). Each tandem repeat contains one copy of each of the five histone structural genes separated by noncoding spacer DNA. The order of the genes as well as their polarity and the relative length of each noncoding spacer DNA segment have been conserved during the millions of years since different sea urchin species last shared a common ancestor (1,2). The sequences of the histone coding segments in the early genes of different species has remained highly conserved whereas the spacer DNA, excluding regulatory sequences immediately flanking the genes, has diverged considerably (3,4). One difference that has been observed between the early histone genes of different

sea urchin species is the presence of nonallelic clusters in Lytechinus pictus (5,6) and Psammechinus milaris (3) but apparently not in Strongylocentrotus purpuratus (7). In P. milaris the two nonintermingled clusters, h19 and h22, are present in greatly differing quantities, with h19 repeats represented only a few times per haploid genome (3). In L. pictus, two nonintermingled tandem repeat families (the A-D and B-C families) are present in roughly equal abundance (5).

In this report we identify a third class of histone early repeat family in the sea urchin L. pictus. In order to understand the structure and evolution of tandem repeated gene families we have cloned and characterized the LpE repeat family and compared it with the other families present in L. pictus. We also show that many variants present in individual animals are fixed in each member of a tandem repeat family but spread only at a slower rate to the nonallelic histone tandem repeats.

MATERIALS AND METHODS

Materials

Enzymes were purchased from New England Biolabs and radioisotopes from Amersham.

Preparation of Phage and Cosmid Libraries and Cloning of Histone DNA.

The partial BamHI DNA library made from the sperm DNA of L. pictus individual #8 using the isolated arms of the phage λ 1059 (8) was prepared and described in Childs *et al.* (9).

Cosmid libraries were constructed in the vector pTL5 (10) from an individual L. pictus sperm DNA sample (#14). Sperm DNA was partially digested with Sau3A as described in Maniatis *et al.* (11), and 35-45 kb DNA was collected by fractionation on 1.5M-5M NaCl gradients (12). Individual gradient fractions were then separately ligated into Bgl II digested vector DNA (1:10 molar ratio) which had been treated with bacterial alkaline phosphatase and the ligated DNA was then packaged *in vitro* (13). Cosmids were amplified on the bacterial strain DH-1 (recA⁻) and colonies were screened using the procedure of Hanahan and Meselson (14).

For selection of phage clone λ LpE, 100,000 phage were spread on NZY plates (22 x 22cm) and plaque lifts were performed by the procedure of Benton and Davis (15). The probe used was the purified EcoRI insert of the plasmid pLpC (5). This probe contains the early H3 and H2B coding sequences as well as surrounding spacer DNA (5,6). Cosmids containing early histone genes were selected using a 1.3 kb Eco RI-Bam HI fragment of the plasmid pSpL3 (16)

cloned into the M13mp8 vector. This probe contains a late H2B gene from the sea urchin S. purpuratus.

Isolation of DNA.

Phage were grown as liquid lysates in NZY broth. Plasmid and cosmid DNAs were isolated by CsCl₂ gradient centrifugation by the alkaline lysis procedure of Birnbaum and Doly (17). L. pictus sperm DNA was isolated as described by Kedes et al. (18).

DNA/DNA hybridizations

Southern blots (19) were hybridized exactly as described by Childs et al. (20). Isolated DNA fragments or whole plasmid DNA was nick translated to specific activities of 1-3 x 10⁸ cpm/μg (21). M13 clones were labeled by elongation from the annealed sequencing primer as detailed in the P/L Biochemicals sequencing manual, except dideoxynucleotide triphosphate was omitted and ³²P-dCTP (3000 Ci/mM) was used.

Electron Microscopic Analysis

Heteroduplexes were made by the procedure of Davis et al. (22), exactly as detailed by Cohn and Kedes (6). R-loops (23) were formed exactly as described by Cohn and Kedes (6). We used L. pictus blastula histone mRNA which had been purified using histone DNA cellulose (24) for the formation of R-loops. Measurements were performed with a numonics graphic calculator on enlarged prints or projected negatives. Double stranded circular pSC101 was used as a length standard (9 kb).

RESULTS

The L. pictus histone genes that are expressed in early embryogenesis are tandemly repeated and contain no BamHI or HindIII restriction endonuclease recognition sites (9,20). We have previously exploited this finding by using these enzymes to separate, by either gel electrophoresis or cloning, the tandem repetitive sequences from other histone gene families which are not in tandem arrays (9). Genomic DNA that is digested to completion with either of these endonucleases results in the tandem repetitive genes being present only on very large DNA fragments. In addition, junction DNA fragments corresponding to the ends of clusters or infrequent interruptions within the clusters of transposable elements will remain associated with the high molecular weight repeat units. An example of such an experiment using sperm DNA collected from five individual animals is shown in figure 1. Tandemly repeated histone sequences can be found in the high molecular weight region of the gel and single copy sequences corresponding either to transposed early

Figure 1: Autoradiogram of a Southern Blot of BamHI-digested Sea Urchin DNA Hybridized to an Early Histone Gene Probe.

Sperm DNA from five *L. pictus* individuals (named 8,12,13,14 and 15) was digested with Bam HI and 5 μ g samples were resolved on a 0.65% agarose gel, blotted, hybridized with 10^6 cpm/ml nick translated insert of the plasmid pLpC and washed as described in the methods section. The panel on the right is a shorter exposure of the lane with individual #8 DNA to indicate the absence of hybridization at intervals of 7.2 kb.

histone pseudogene "orphons" (20) or late histone genes (9,16) can be found in lower molecular weight fragments. Individual #8, however, contained a tandemly repetitive histone gene family with a single BamHI site every 7.2 kb, the same size as the early histone gene repeat unit in this sea urchin species (5,6). Thus, it seems that a significant fraction of the early histone genes of male #8 contain a BamHI site and that these repeats are not intermingled with other tandem arrays that are resistant to cleavage by this enzyme. Remarkably, all or almost all the repeats of this family contain a BamHI site as seen by the absence of bands at multiples of 7.2 kb on the Southern blot. This fixation of the variant cluster containing the BamHI site has occurred in a nonintermingled fraction of tandem repeats of a single

individual within a wild population of animals of the same species. This appears to be an example of concerted evolution between members of multigene families within an individual genome.

Cloning and Characterization of the BamHI sensitive tandem repeat family of animal #8.

Previous analysis of the early histone genes of L. pictus has identified two distinct nonintermingled classes of nonallelic tandem repeats (referred to as the A-D and B-C repeats) which differ primarily in their spacer DNA sequences (6). We have cloned the repeat family containing a BamHI site from individual #8 in order to identify from which nonallelic repeat family it is derived. Several phage clones were identified (see Methods section) that contained the same variant histone gene cluster and one, λ LpE, was chosen for detailed study. λ LpE contains a 14.4 kb insert consisting of two complete tandem repeat units (data not shown). One of these tandem arrays was subcloned into the plasmid vector pBR322 (25) and this repeat was further characterized. Early histone gene families from every sea urchin species examined are comprised of repeat units containing each of the five classes of histone genes. These genes are arranged in the same order and polarity and are separated by noncoding spacer DNA. To determine the position and number of coding regions on the LpE repeat unit, we examined BamHI digested plasmid DNA molecules for their ability to form R-loops with histone mRNAs (6,26). As predicted for the structure of histone repeat units, the R-loops consist of regions of DNA complementary to early histone mRNA and are separated by noncoding spacer DNA (figure 2A). As others have shown (6,26), the lengths of the R-loops are smaller than the known sizes of the histone genes probably due to branch migration. Accordingly, in figure 2C we have extrapolated from the center of each measured R loop to the known size of each histone gene. The order of the genes shown in figure 2B was determined from multiple restriction endonuclease digestions followed by hybridization to each of the five histone genes (see figure 5 and data not shown) as well as heteroduplex analysis with other L. pictus histone repeats with a known gene order and polarity (see below). The histone gene organization shown in figure 2 is typical of sea urchin early histone genes (1,2) and agrees well with the sizes of the total repeat length and spacer DNA for L. pictus determined by Cohn and Kedes on two nonallelic repeats present on the plasmids pLpA, pLpB and pLpC (5,6).

Analysis of heterduplexes formed between the histone gene segments of the A-D and B-C repeat units revealed areas of homology in the coding segments but extensive segments of nonhomology in each of the spacer regions (6). We

Figure 2: Electron Micrographic Analysis of the Histone Gene Segment of pLpE after R-loop Formation with Purified Histone mRNA.

A&B) Electron micrographs of the histone gene segment from pLpE after R loop formation with mRNA. The 7.06 kb segment has five histone gene coding regions separated by spacer DNA. These genes are 3' H2A-H1-H4-H2B-H3 5'. No single molecules displaying all five R loops were found. C) Summary of R loop analysis. 1) Measurements and statistics of spacers and R loops, presented as number of base pairs + standard deviation from the mean. N is the number of molecules measured. 2) Physical map of the histone genes in pLpE. The sizes of the coding regions measured from R loop analysis were changed to actual mRNA lengths using the centers of the R loops as the centers of the mRNA coding regions.

confirmed these results for the H3 coding segments and the spacer DNA separating the H3 and H2A genes by comparing their nucleotide sequences (27). The coding segments are 97% conserved while the spacers show areas of non-homology interdigitated with short regions of as much as 90% homology. We decided to examine the homology of the entire histone DNA segment of pLpE with the clones described by Cohn and Kedes (5,6). Figures 3A and B show an electron micrograph of a heteroduplex formed between BamHI digested pLpE and EcoRI digested pLpA. We observed four regions of homology separated by nonhomologous regions of DNA. The homologous segments map to the coding segments of the H1, H4, H2B and H3 genes as well as a limited amount of noncoding spacer DNA immediately flanking each gene (Figure 3C). The spacer

Figure 3: Electron micrograph of a Heteroduplex formed Between the Eukaryotic Insert of pLpE and pLpA.

A) Electron micrograph of a typical pLpE-pLpA heteroduplex. B) Schematic diagram of the heteroduplex indicating the positions of the five histone genes. C) Schematic summary of heteroduplex analysis. Bold lines represent double-stranded regions of homology between the histone gene coding regions of pLpE and pLpA; fine lines indicate non-homologous single-stranded regions in spacer DNA. The single-stranded areas are contracted approximately 10% from their actual lengths due to the conditions of the spread (39). This is not corrected for in the numbers given. For this reason we refer to the length of the LpE repeat as 7.2 kb. Measurements are presented as number of bases $\pm$ standard deviation from the mean.

DNA homology probably occurs in the conserved 200 base pair segment immediately adjacent to the 5' end of each gene (2,3) but our measurements are not precise enough to be certain at which end of the gene the additional homology lies. Because pLpA contains only four of the five histone genes of the A-D repeat unit, (5,6 and figure 5) (the H2A gene is on the 1.2 kb D fragment), the H2A segment of pLpE is, as expected, not homologous with pLpA. These results demonstrate that the histone genes contained on pLpE are in the same order and polarity as those of all other sea urchin early tandem repeat units.

Using the spreading conditions, described, we have never observed nonhomologous regions in heteroduplexes formed between BamHI digested pLpE and

Figure 4: Restriction Endonuclease Maps of the Three Tandem Repetitive Early Histone Gene Families in *L. pictus*.

Restriction endonuclease maps were generated by examination of single and double digestion of whole plasmid, cosmids and isolated eukaryotic segments. The B-C and A-D tandem repeats are aligned with the two tandem repeats found on λ LpE. The position of the histone coding segments are indicated by the heavy lines while those of the spacer DNA by the thin lines. Brackets indicate the position and length of the Eco RI fragments characteristic of each repeat unit.

EcoRI digested pLpB or EcoRI pLpC (data not shown). Comparisons of restriction endonuclease maps show differences between pLpE and the B-C tandem repeat family (figure 4). This is best illustrated by examination of the location of EcoRI sites in the three cloned repeats. Both the LpB-C and LpE repeats contain 4.1 kb and 3.1 kb EcoRI fragments. The pLpB fragment of 4.1 kb contains the H2A, H1 and H4 genes (5) while the equivalent sized fragment in pLpE contains the H4, H2B and H3 genes. Similarly, the 3.1 kb EcoRI insert of pLpC contains the remaining two histone genes of the B-C repeat unit, the H2B and H3 genes (5), while the equivalent sized fragment in pLpE contains the remainder of the E repeat including the H2A and H1 genes. These results are presented in figure 5 A-F. These data support the conclusion that *L. pictus* contains three distinct nonintermingled, nonallelic classes of early histone genes: the LpA-D, LpB-C, and LpE repeats. Because the sizes of the EcoRI sites are identical in the LpB-C and LpE repeats, it is perhaps understandable that the LpE repeat has remained undetected until now. Cohn and Kedes (5,6) isolated only two clones, pLpB and pLpC, which by chance contained the two fragments that are adjacent in a tandem repeat. Many restriction endonuclease

sites, especially those within coding segments, are conserved between the three repeat units but extensive differences are also apparent (figure 4 and ref. 5).

Identification of E type tandem repeats in all individual sperm DNA's.

It remains possible that individual #8 was unusual and most L. pictus contain only the LpA-D and LpB-C repeats. We tested this idea in two different ways. First, sperm DNA of five individual L. pictus males were digested to completion with EcoRI, blotted onto nitrocellulose (19), and hybridized to a subcloned DNA fragment which contains only the early H3 gene of L. pictus (27). The histone H3 gene is contained on the 6 kb fragment corresponding to pLpA (A-D repeats) and the 3.1 kb fragment corresponding to pLpC (B-C repeats). Only if each individual sperm DNA sample contains the pLpE repeat will a 4.1 kb fragment hybridize to the histone H3 subclone. The results clearly demonstrate the 4.1 kb EcoRI fragment hybridizing to the H3 gene in each of the five individual DNA samples (Figure 6). Individuals #12-15 therefore contain pLpE repeats, however, they have no BamHI site. These results also demonstrate the extent of length and sequence polymorphism between individuals as well as the rapid fixation of the variants within the same repeat family. There is some heterogeneity in the length of the 4.1 kb EcoRI fragment of LpE repeats, as well as differences in the abundance of the LpE repeat relative to the LpA-D and LpB-C repeats (figure 6). Several additional polymorphisms are also present in these five individuals. Animal #13 contains an EcoRI (A) fragment that is considerably shorter than the 6.0 kb standard sized fragment. This length variant also represents a larger proportion of the total number of early histone genes than the LpA-D family present in other individuals. Alternatively, the (A) fragment is the "normal" 6.0 kb in length in animal #14 but it represents a much smaller fraction of the repeats. This individual also has an unusual arrangement of the LpB-C repeat unit. The 3.1 kb fragment is completely absent but instead an intense band of 7.2 kb is found. The most likely explanation is that each member of the LpB-C repeat in animal #14 has lost the EcoRI site defining the B-C junction, resulting in a 7.2 kb band. It is also apparent that animal #14 is homozygous for this variant repeat structure. Animal #12 contains a group of tandem repeats that are resistant to EcoRI cleavage. We have not identified which type of repeat class these EcoRI resistant histone genes are derived from. This individual (#12) also contains 6.0 kb, 4.1 kb and 3.1 kb species which hybridize to the H3 subclone. The EcoRI resistant histone genes could represent an allele in a heterozygous individual that has lost the EcoRI sites of any one of the three types of

Figure 5: Analysis of the Individual Histone Genes Found on the EcoRI Fragments of Cloned Early Histone Genes.

Cloned plasmid, cosmid or phage DNA (0.5 µg/gal) was digested to completion with EcoRI. Five identical agarose gels (0.8%) were prepared and samples were blotted onto nitrocellulose and hybridized as described below. A) Ethidium-bromide stained DNA of one of the agarose gels used for the hybridization analysis. The DNA sample in each lane is indicated at the top of the figure. Sizes of the histone DNA inserts are shown to the right of the figure. B) Hybridization of a gel (shown in panel A) to a histone H1 gene specific probe. The 1.37 kb HincII fragment containing the H1 gene (5) was isolated, nick translated, and hybridized at 150,000 cpm/ml to a nitrocellulose replica of panel A. C) Hybridization to an H2A specific probe. The probe used (500,000 cpm/ml) was an M13 subclone of the 800 bp EcoRI-BamHI fragment of the *S. purpuratus* late histone gene clone SpL3 (16). This probe contains an entire late H2A gene and some surrounding nonhistone DNA. D) Hybridization to an H2B gene specific probe. The probe used (500,000 cpm/ml) was a 1.2 kb EcoRI-BamHI fragment subcloned into m13 mp8. The fragment is

derived from the *S. purpuratus* late histone clone SpL3 (16) and contains the entire late H2B gene plus surrounding nonhistone DNA. E) Hybridization to an H3 gene specific probe. The probe used (500,000 cpm/ml) was an m13 mp8 subclone (A5H) encoding amino acids 45-93 of the early H3 gene found on pLpA (27). F) Hybridization to an H4 gene specific probe. The probe used (500,000 cpm/ml) was m13 H4319 (399 bp) (27) which contains the entire *L. pictus* late H4 coding sequence and 5' leader region. The (*) denotes minor contaminating EcoRI* activity that produces bands which hybridize but are not easily visible on the stained gel.

repeats. This would be consistent with our finding that many of the variants which we detect are spread to each member of a tandem repeat family. However, it is possible that only a fraction of the repeats in a continuous segment of one or more nonallelic families are affected in this manner. That we can see examples of individuals which are apparently homozygous for the B-C repeat unit and others that are heterozygous for one of the repeat families and yet still contain E repeats strongly argues for the independent nature of the E gene family.

Isolation of Cosmids Containing the Early Histone genes from Animal #14

We have constructed a cosmid library from a partial Sau3A digest of the sperm DNA of animal #14 (see materials and methods). We screened this library with the late H2B gene of the *S. purpuratus* clone SpL3 (16) and purified several cosmid clones. Among the clones that hybridized to this probe were cosmids containing early histone genes. None of the three clones tested, CosLpA-D, CosLpE and CosLpA-D-E, contained a BamHI site within the eukaryotic inserts (data not shown). The cosmid vector pTL5 contains a single BamHI site (10) that is the only BamHI site in each of these cosmid clones. The total length of DNA in each clone indicates that about five individual tandem repeats are present. The sizes of the EcoRI fragments contained on each clone were compared with the inserts of the clones pLpA, pLpB, pLpC and λ LpE (figure 5A). CosLpA-D contains in addition to the vector sequences, bands of 6.0 kb and 1.2 kb. The 6.0 kb fragment hybridizes with H1, H2B, H3 and H4 probes while the 1.2 kb fragment hybridizes only to the H2A probe (figure 5B-F). These results are consistent with the structure of the LpA-D tandem repeat unit characterized by Cohn and Kedes (5,6). CosLpE contains, in addition to the vector sequences, bands of 4.1 kb and 3.1 kb. Each of these fragments hybridize to the same individual histone gene probes as the corresponding sized fragments in λ LpE but not pLpB and pLpC (figures 5 B-F) proving that animal #14 contains E repeats but without BamHI recognition sites. CosLpA-D-E surprisingly contains both A-D and E type repeat units based on the sizes of the EcoRI fragments and the individual histone genes contained on each fragment (figure 5). A 2.7 kb EcoRI

Figure 6: Autoradiogram of a Southern Blot of Eco RI Digated Sea Urchin DNA hybridized with an H3 subclone.

Sperm DNA from five *L. pictus* individuals was digested with Eco RI and 5 μ g samples were resolved on a 0.65% agarose gel, blotted, and washed as described in the Methods section. The probe used (10^6 cpm/ml) was an ml3mp8 subclone (A5H) containing a Hae III fragment of the early H3 gene of pLpA encoding amino acids 45-93 (27).

Figure 7: Structure of Cosmid LpA-D-E.

A restriction endonuclease map was generated by examination of single and double digests of cosmid DNA. The positions of each histone coding segment were determined by hybridization to Southern blots and from the known orientation of histone genes in the LpA-D and LpE tandem repeats.

Figure 8: Autoradiogram of a Southern blot of EcoRI digested sea urchin individual #14 DNA probed with pLpA.

Sperm DNA from individual #14 (5 μ g) was digested with EcoRI and resolved on a 0.65% agarose gel, blotted and then hybridized with nick translated pLpA DNA (10^6 cpm/ml). pLpA contains the early H1, H2B, H3 and H4 genes. The sizes of the fragments shown were determined using HindIII digested lambda DNA present on the same gel.

fragment which hybridizes to the H4 and H2B genes spans the A-D and E repeats of this cloned DNA fragment (figures 5 and 7). Digestion with other restriction endonucleases confirms that the A-D and E tandem repeats are not intermingled in this cosmid clone (data not shown). The identification of a cosmid clone containing both E and A-D tandem repeats argues that at least a portion of the A-D and E tandem arrays can be adjacent. We cannot rule out the remote possibility that CosLpA-D-E resulted from a cloning artifact. It is, however, unlikely that two histone DNA fragments would ligate to one another amid all the nonhistone genomic DNAs present during construction of the cosmid library. We are, however, able to identify an EcoRI fragment of the same size (2.7 kb) as the one that spans the two repeats found in CosLpA-D-E in the genomic DNA of individual #14 (figure 8). The presence of as many as 50 "orphan" early histone genes (20) and late histone genes (9,16) not present in the tandem repeats (see all the additional bands present in single copy in figure 1) precludes any definitive conclusions about the origins of the 2.7 kb fragment using genomic Southern blots.

DISCUSSION

We have shown that the sea urchin *L. pictus* contains three different nonintermingled arrays of tandemly repeated early histone genes. Each of these tandem arrays is arranged in the same invariable manner as other sea urchin

tandem repeats but each family when compared with the others contain distinct sequence polymorphisms. Heteroduplex analysis between the newly identified LpE class of tandem repeats with pLpA revealed extensive regions of nonhomology in the spacer DNA (figure 3). Variations in the spacer DNA segments of LpE and the B-C repeat unit were evident only by differences in restriction endonuclease recognition sites (figure 4). These comparisons support the conclusion that the LpE repeats are more closely related to the LpB-C family than the LpA-D family. Due to the possibility of continuous exchange of sequences between early histone gene families, it is impossible to distinguish whether the LpE and LpB-C families appear to be more closely related because they exchange sequences more frequently or if they evolved more recently from a common ancestor. We have examined the sequence relatedness of short regions of the LpA and LpB-C repeats by nucleic acid sequencing. The sequences of the H3 coding segments and 3' spacer DNAs from the A-D and B-C repeats showed that the coding segments are 97% conserved while the spacer DNAs were much less conserved (27). These results implicate mechanisms such as gene conversion combined with selection for the maintenance of the highly conserved regions of coding and spacer DNA segment. We have found another example of this within the same L. pictus genome. The late histone genes expressed only at high rates in later embryogenesis also exchange sequences among family members but not at detectable rates between early and late histone genes (27). This physical separation of late histone genes from the tandem repeats may explain how they can be independently maintained in all sea urchin species for over 150 million years.

Within an individual sea urchin, most histone repeat units are virtually identical in both size and sequence (5,7). However, when several individuals of a sea urchin species are compared, both length and sequence polymorphisms can be identified (5,7). Thus, while the sequences within a given family of tandem arrays are remarkably homogeneous, they may vary considerably between different families. This concerted evolution is reminiscent of similar observations made with other tandem repeated gene families such as Drosophila rDNA and histone genes (28-30), Xenopus 5S rDNA (31) and mammalian rDNA (32). The rate of this process has been most carefully analyzed in a tandem array of APRT (adenine phosphoribosyl transferase) genes that had been transfected into L-cells (33). Roberts and Axel have found that the entire set of genes in an array can rapidly acquire a mutation originally found in only one copy of the array (33). One mechanism postulated to explain how tandem arrays in one family can remain nearly identical, leaving other families corrected only to a

lesser degree, would be to place them on separate chromosomes (28). This is apparently the basis for the concerted evolution of human and ape rDNA in portions but not the entire length of the ribosomal gene (32). Portions of rDNA repeats on different chromosomes are maintained in a polymorphic state. Similarly, the coding segment of the 18S ribosomal DNA of Drosophila melanogaster contains a one base pair change in all members present on the X chromosome when compared to each of the 18S genes on the Y chromosome (34). Since gene conversion events acting to homogenize sequences occur more frequently on closely linked sequences (35), partitioning of the tandem repeat units on different chromosomes could explain the maintenance of variation in the A-D, B-C and E repeat units as well as the independence of the late histone gene family and sperm specific histone genes. Here we propose that ability of variants to be homogenized within but not between arrays is a consequence of gene conversion processes which occur more easily between genes which are in close proximity.

The ability of independent tandem clusters to interact is documented by the cosmid clone CosLpA-D-E. The spacing of the Eco RI sites 2.7 kb apart in the DNA which spans the A-D and E families on CosLpA-D-E is inconsistent with a homologous recombination event within an equivalent region of an A-D and E tandem repeat member during unequal exchange of sequences. Instead some type of nonhomologous recombination, or deletion or inversion event like one documented within the S. purpuratus early histone genes (36), could account for the structure of CosLpA-D-E. The extensive differences in the spacer DNA segments of these two repeat families (figure 3) makes it improbable that the majority of the A-D and E families are adjacent in the L. pictus genome. If this were the case, there is no reason why variants would not spread equally among these families, homogenizing their sequences. More likely, there is a continued exchange of sequences between the three families but at a slower rate than intrafamily homogenization. Selection may also play a key role in the maintenance of three distinct families of early histone genes. There is considerable heterogeneity among early histone mRNAs (37). The h19 and h22 histone gene families in P. milaris encode different histone subtypes for the H2A, H2B and probably H1 proteins (3). It has been suggested that h19 genes were acquired by P. milaris by a recent horizontal transfer of a histone gene cluster from S. drosbachiensis (38). Perhaps the acquisition or duplication and divergence of tandem repeat families entails a selective advantage due to the presence of distinct histone subtypes.

We originally discovered the LpE repeats in individual #8 (figure 1)

because they contained a single BamHI recognition sequence. What remains striking is the fact that this variant had spread throughout the E family in this individual but not to the A-D or B-C histone gene families. Additional equally dramatic fixation of variants within only one of the three early histone gene families are seen (figure 6). Other polymorphisms frequently associated with tandem repeat families such as variations in the redundancy and repeat length can also be seen. The early histone genes of L. pictus are therefore undergoing continual expansion and contraction of their copy number. Further investigation of the structure and sequence of L. pictus histone gene families should help to identify mechanisms used to exchange information between related multigene families.

ACKNOWLEDGEMENTS

We would like to thank Cathy Nocente-McGrath and Karen Weisser for technical assistance and Dr. Scott Hawley for critical evaluation of the manuscript. G.C. would like to acknowledge the constant input and criticism of our weekly research talk group (Drs. L. Leinwand, J. Sninsky, S. Hawley, S. Emmons and J. Lenz). We would also like to thank Dr. L. Kedes for providing plasmids pLpA, pLpB and pLpC. This work was supported by a grant (GM 30333) from the National Institutes of Health. G.C. is the recipient of a Junior Faculty Award from the American Cancer Society. C.A.H. is supported by a predoctoral Training Grant from the National Institutes of Health.

REFERENCES

1. Kedes, L.H. (1979) Ann. Rev. Biochem. 48, 837-870.
2. Hentschel, C.C. and Birnstiel, M.L. (1981) Cell 25, 301-313.
3. Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucleic Acids Res. 8, 957-977.
4. Hentschel, C.C., Irminger, J.C., Bucher, P. and Birnstiel, M.L. (1980) Nature 285, 147-151.
5. Cohn, R.H. and Kedes, L.H. (1979) Cell 18, 843-853.
6. Cohn, R.H. and Kedes, L.H. (1979) Cell 18, 855-864.
7. Overton, G.C. and Weinberg, E.S. (1978) Cell 14, 247-257.
8. Karn, J., Brenner, S., Barnett, L. and Cesareni, G. (1980) Proc. Nat. Acad. Sci. USA 77, 5172-5176.
9. Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and Knowles, J.A. (1982) Cell 31, 383-393.
10. Lund, T., Grosveld, F.G. and Flavell, R.A. (1982) Proc. Nat. Acad. Sci. USA 79, 520-524.
11. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, p. 282, Cold Spring Harbor Press, New York.
12. Grosveld, F.G., Land, T., Murray, E.J., Mellor, A.L., Dahl, H.M. and Flavell, R.A. (1982) Nucleic Acids Res. 10, 6715-6732.
13. Hohn, B. and Murray, K. (1977) Proc. Nat. Acad. Sci. USA 74, 3259-3263.
14. Hanahan, D. and Meselson, M. (1980) Gene 10, 63-67.
15. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-183.

16. Maxson, R., Mohun, T., Gormezano, G., Childs, G. and Kedes, L. (1983) *Nature* 301, 120-125.
17. Birnbaum, H.C. and Doly, J. (1979) *Nucleic Acids Res.* 7, 1513-1523.
18. Kedes, L.H., Cohn, R.H., Lowry, J.C., Chang, A.C.Y. and Cohen, S.N. (1975) *Cell* 6, 359-369.
19. Southern, E.M. (1975) *J. Mol. Biol.* 98, 503-517.
20. Childs, G., Maxson, R., Cohn, R.H. and Kedes, L.H. (1981) *Cell* 23, 651-663.
21. Rigby, P.W., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
22. Davis, R.W., Simon, M. and Davidson, N. (1971) *Meth. Enzymol.* 21, 413-428.
23. White, R.L. and Hogness, D.S. (1977) *Cell* 10, 177-192.
24. Childs, G., Levy, S. and Kedes, L.H. (1979) *Biochemistry* 18, 208-213.
25. Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) *Gene* 2, 95-113.
26. Holmes, D.S., Cohn, R.H., Kedes, L.H. and Davidson, N. (1977) *Biochemistry* 16, 1504-1511.
27. Roberts, S.B., Weisser, K. and Childs, G. (1984) *J. Mol. Biol.* 174, 647-662.
28. Coen, E.S. and Dover, G.A. (1983) *Cell* 33, 849-855.
29. Coen, E.S., Thoday, J.M. and Dover, G. (1982) *Nature* 295, 564-568.
30. Coen, E.S., Strachan, T. and Dover, G. (1982) *J. Mol. Biol.* 158, 17-35.
31. Ford, P.J. and Brown, D.D. (1976) *Cell* 8, 485-493.
32. Arnheim, N., Krystal, M., Schmickel, R., Wilson, G., Ryder, O. and Zimmer, E. (1980) *Proc. Nat. Acad. Sci. USA* 77, 7323-7327.
33. Roberts, J.M. and Axel, R. (1982) *Cell* 29, 109-119.
34. Yagura, T., Yagura, M., Murematsu, M. (1979) *J. Mol. Biol.* 133, 533-542.
35. Roeder, G.S., Smith, M. and Lambie, E.J. (1984) *Molec. & Cellular Biol.* 4, 703-711.
36. Vitelli, L. and Weinberg, E.S. (1983) *Nucleic Acids Res.* 11, 2135-2153.
37. Grunstein, M., Schedl, P. and Kedes, L.H. (1976) *J. Mol. Biol.* 104, 351-369.
38. Busslinger, M., Rusconi, S. and Birnstiel, M.L. (1982) *The EMBO J.* 1, 27-33.
39. Ferguson, J. and Davis, R.W. (1978) *Advanced Techniques in Biological Electron Microscopy II*, ed. J.K. Koehler, pp. 123-171.