Evidence for nonrandom alterations in a fraction of the highly repetitive DNA of a eukaryote

Nelwyn T.Christie and Dorothy M.Skinner*

University of Tennessee-Oak Ridge Graduate School of Biomedical Sciences, and the Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830, USA

Received 18 October 1979

ABSTRACT

Although the DNA of the red crab *Geryon quinquedens* has no patent satellites, a large fraction ($\sim 40\%$) is highly repetitive. Treatment of total DNA by Hind III produces fragments comprising 5-6% of the genome. While the sizes of some of these fragments form an arithmetic series based on an 81 bp repeating unit, the amounts of the multimers differ significantly from distributions observed for multimeric series in the DNAs of other eukaryotes. In red crab DNA, the amounts of some of the multimers suggest that they may have undergone as much as four times the divergence as the others. Other data, however, are more compatible with the conclusion that there has been selective amplification of segments of highly repeated DNA which results in the enhanced amount of specific multimers. These results indicate the presence of a nonrandom process in the evolution of the highly repetitive DNA. Selective mutation alone seems insufficient to explain these results.

INTRODUCTION

The size of the eukaryotic genome greatly exceeds estimates of the amount of DNA required for growth and development of the organism; no more than 5 to 15% is thought to code for proteins. The remainder of the genome may be involved in chromosome structure; it may influence differential gene expression as well (1). Within the large noncoding fraction, attention has been focused on repetitive DNAs because of their ease of isolation and chromosomal localization. In the four brachyuran ("true crab") genomes which have been examined, highly repetitive sequences comprise 25 to 41% of the genome (2-4). In each case except the red crab *Geryon quinquedens* these sequences include patent satellite DNAs. [Patent satellites are observed by centrifugation in CsCl gradients; cryptic satellites are obscured by the major DNA component and are observed with special treatment, such as binding to ligands. Other repetitive sequences, which may or may not be separable by centrifugation, can be isolated by rapid reassociation (5,6) or treatment with restriction enzymes (4,7).] The most striking characteristic common to the DNA of all four species is that there are very small

amounts of sequences classified as middle repetitive. Typically, in other eukaryotes these sequences, with several hundred copies per genome, comprise between 15 and 50% of the total DNA. For the crab genomes the majority (greater than 90%) of the repetitive sequences can appropriately be termed "highly repetitive" with copy numbers of several thousand per genome. While this distribution of repetition frequencies may or may not be a characteristic feature of crab genomes, it facilitates fine structure analyses of this class of DNA. Detailed structural analyses should be useful in elucidating the origins of highly repetitive DNAs. The structure of a satellite fraction from the hermit crab *Pagurus pollicaris* is strong evidence that unequal crossovers occur in highly repetitive DNAs (8,9). The marked diversities in the structural features of satellite DNAs (10) suggest that more than one mechanism is involved in their formation.

In general, it has been assumed that highly repetitive DNAs originate as a homogeneous class of sequences in the eukaryotic genome. Regardless of the mechanism(s) of origin of these DNAs, it is clear that with increasing time from their amplification, alterations of various types accumulate. These alterations, which diminish sequence homogeneity, may include single base changes, deletions, and insertions, as well as more complex changes such as unequal sister chromatid exchanges and other chromosomal rearrangements. Accumulation of such modifications has the potential of transforming simple sequence DNAs, which might behave as satellites in CsCl gradients, into sets of cryptic repetitive sequences with densities similar to that of the remainder of the genome. Although the DNA of the red crab lacks patent density satellites, it contains several classes of sequences with characteristics similar to those of satellite DNAs. Both the reassociation rate and the measured mismatch (4%) of the entire class of repetitive DNA (4) are similar to values observed for satellite DNAs (11). In addition, digestion of *Geryon* DNA by several restriction endonucleases yields extensive arrays of discretely sized fragments. By digestion of total DNA with Hind III, we identified a series of fragments whose sizes were multimers of a basic repeat of 81 base pairs (bp). The present study was undertaken to explore the relationship of the sequences of the different multimers. The amount of each multimer present in several satellite DNAs is reduced compared to the next smaller multimer (12-14). This result is expected if there is random loss of restriction sites that had been spaced uniformly. In *Geryon* DNA the amounts of both

the tetramer and the octamer are larger than expected, suggesting non-random changes in those sequences which contain the Hind III sites. If an amplification step is responsible for this bias, only selected fractions have been amplified. Amplification of the entire set of sequences containing the Hind III sites would maintain multimer ratios determined by the degree of divergence since the initial amplification of the basic unit.

MATERIALS AND METHODS

Isolation and Labeling of DNA. DNA was isolated as described (4). Both labeled and unlabeled DNAs were used for restriction enzyme analyses. DNA was labeled with γ -³²P-ATP (2000-3000 Ci/mmol; from ICN) at the 5' termini after treatment with bacterial alkaline phosphatase (Worthington) by the action of T4 polynucleotide kinase (15,16; P. L. Biochemicals, Inc.).

Restriction Enzyme Analyses and Isolation of Restriction Fragments. Other than Hind III and Alu I (4), restriction endonucleases were purchased from New England Biolabs. Restriction digestions were carried out as described (8,17). Preparative digests were precipitated with ethanol and solubilized in electrophoresis buffer before application to the gels. Typically, 50 μ g of DNA was applied to wells 1.2 X 0.2 cm for preparative isolations. Isolated fragments were further purified by electrophoresis on a similar gel. Electrophoresis was carried out in a vertical apparatus with 90 mM Tris-borate buffer (pH 8.3) and 2.5 mM EDTA (18). Individual fragments were excised from the gels and recast with 7% polyacrylamide in glass tubes containing a 10% polyacrylamide plug. Fragments were electroeluted at 4°C into dialysis bags and precipitated before redigestion by other restriction endonucleases.

Quantification of the DNA in the Restriction Fragments. Quantification of the amounts of Hind III fragments was accomplished using optical density scans of total DNA (6 μ g) digested with Hind III and electrophoresed on 5% acrylamide gels. Since the film response to impinging light is represented by a sigmoidal function (19), parallel gels of ϕ X174 DNA (1 μ g) digested with Alu I were photographed simultaneously. The amount of ϕ X174 DNA and the photographic conditions (aperture settings, time of exposure) were adjusted until the ratios for the areas under selected DNA peaks agreed with independent determinations of the amount of DNA in the corresponding fragments (20). This assures that the response of the film was in the linear range and justifies the use of peak areas to determine the amount of DNA in the Hind III fragments of *Geryon*.

Measurements of the relative amounts of the fragments were made by labeling the 5' ends of a Hind III digest of total DNA and electrophoresing on 7% polyacrylamide gels. The gels were sliced into 1-mm segments, incubated overnight with 0.2 ml of HClO_4 in H_2O_2 (1:2 by volume), and counted in scintillation fluid (Amersham/Searle). The total radioactivity in the major peaks ranged from 5.2×10^4 to 2.8×10^5 cpm. Estimates of the amount of each fragment were made from the lengths of the fragments, the radioactivity in each, and the genome size. For both labeled and unlabeled DNA, all fragments of one size class were included since the contribution of unrelated fragments is not known. At the maximum this represents a two-fold overestimate, but would not affect the conclusion that certain sequences have been amplified.

Determination of the Sizes of the Restriction Fragments. The sizes of the restriction fragments were obtained by comparison to Hinf I fragments of ϕX174 which range from 24 to 726 bp. A least-squares regression line was calculated for plots of the log of the known Hinf I sizes in nucleotides (21) versus the distance migrated. The deviations between the sizes obtained from DNA sequencing data (21) and the sizes estimated from the linear regression line were only 1-5 bp for fragments less than 200 bp but increased for larger fragments. Fractional deviations in size for the larger fragments were comparable to those of the small fragments. Absolute deviations increased in proportion to fragment length (22).

RESULTS AND DISCUSSION

When total DNA is digested with Hpa II, Hha I, Taq I, Hga I, Hinf I, Hae III, or Mbo I, fragments of discrete sizes are observed. The discrete fragments coexist with a background of DNA fragments having a continuous range of sizes. Most of the enzymes used have recognition sites of 4 bp, which occur rather frequently throughout the genome. Digests of total DNA with Sma I and Bam I, whose sites are 6 bp, contained no detectable fragments of discrete size classes (data not shown).

Organization of Hind III Sites in Total DNA. Of all the enzymes tested Hind III produced the smallest number of discretely sized fragments and the least amount of heterogeneously sized DNA (Fig. 1A; lanes 2 and 4). The size of each of the five major fragments was determined as described in Materials and Methods relative to Hinf I fragments of ϕX174 DNA (Fig. 1A; lanes 1 and 3). The lengths of the five fragments conform to the proposed organization (see below) derived from restriction enzyme analyses and direct

Figure 1. Gel Electrophoresis of Hind III Digests of Total *Geryon* DNA. In (A) Hind I digests of ϕ X174 DNA (lanes 1 and 3) and Hind III digests of *Geryon* DNA (lanes 2 and 4) were labeled as described and electrophoresed on a 5% (lanes 1 and 2) or a 7% (lanes 3 and 4) polyacrylamide slab gel. In (B) is a densitometric scan of a Hind III digest of *Geryon* DNA which was electrophoresed on a 5% polyacrylamide tube gel. Fragments from monomer to octamer are displayed. Accurate sizing for specific fragments is shown in A, lanes 1, 2 and Fig. 3, 5 and 6.

DNA sequencing. The slight variations in the sizes of the fragments obtained for different electrophoresis conditions are within reasonable experimental error. The smallest fragment (V) is approximately 81 bp; larger fragments are multiples of this basic unit. Fragments III and IV are very close to the lengths expected for a tetramer and a dimer. Two fragments, I and II, are both close to the length of an octamer. There are also minor fragments at sizes expected for trimers, pentamers, hexamers, and heptamers. We show below that the sequences of all the multimers from the dimer to the heptamer and at least one of the octamers are very closely related. The decamers may be unrelated to the basic 81 bp repeat or they may be related to one of the octamers. Finally, the fragment of approximate nonamer length represented a small fraction of the total and was not analyzed.

Quantification of Hind III Fragments in Total DNA.

The relative amounts of the fragments were determined either from scans of gels containing UV-visible amounts of DNA or by counting labeled fragments displayed on a gel. A direct estimate of the amount of the fragments smaller than fragment III was made from optical density scans of a total Hind III

digest (Fig. 1B) using scans of comparably sized Alu I fragments of ϕ X174 DNA as standards. The combined amounts of major and minor fragments comprise approximately 5% of the genome. While nearly equal numbers of copies of four of the five major fragments are present (monomer, dimer, and two octamers), the tetramer is present at a somewhat lower frequency. The trimer is only a minor fragment (Fig. 2). The unexpected amounts of several multimers lead us to undertake restriction endonuclease mapping of the multimers to explore their relatedness.

If it is assumed that cleavage sites for Hind III were once spaced uniformly, a test for the random loss of sites can be performed (12,23). The proportion of each multimer ($\underline{n} = 1$ for a monomer) will be given by the probability of a run of ($\underline{n} - 1$) consecutive alterations in an array of regularly

Figure 2. Test for Random Inactivation of Hind III Sites of Total DNA. If it is assumed that a regularly spaced array of restriction sites is altered by random divergence to give dimers, trimers, and larger multimers of the basic unit, the fraction $\underline{f}$ of DNA in each multimer is given by

$$\underline{f} = n(1-p)^2 p^{(n-1)}, \text{ or}$$

$$\log \underline{f}/n = 2 \log (1-p) + (n-1) \log p,$$

where $\underline{n} = 1$ for monomers, etc., and $\underline{p}$ is the fraction of modified sites. The plot of $\log \underline{f}/n$ versus $\underline{n}-1$ is a straight line when sites are inactivated at random assuming that a change of one base pair prevents enzymatic cleavage. The slope is given by $\log p$ and the intercept by $\log (1-p)$. The term $\underline{f}/n$ has been replaced by $\underline{c}_n$, the number of copies of each multimer relative to the monomer. The latter term is determined directly as the ratio of radioactivity in each multimer. The dashed lines are three theoretical plots for different degrees of divergence in a once uniform repeat: (a) 3.7%, (b) 9.5%, (c) 15.7%. Divergence is the fraction of altered base pairs (11) calculated for the following fractions of modified sites: (a) 22%, (b) 56%, and (c) 92% by $\underline{p} \times 1/\underline{t}$ where $\underline{t}$ is the number of base pairs in the restriction site. Discontinuity is indicated on the graph between the amounts of octamer and decamer since essentially no nonamer fragments were detected.

TABLE 1. ESTIMATE OF THE AMOUNTS OF HIND III FRAGMENTS IN THE GENOME

HIND III FRAGMENT (bp)	RATIO OF COPIES/GENOME ^a	PERCENT OF THE GENOME ^b	COPIES/GENOME ^c
81	1.0	0.25	1.2×10^5
162	0.95	0.30	1.2×10^5
243	0.10	0.075	1.2×10^4
324	0.35	0.35	4.2×10^4
405	0.005	0.006	6.0×10^2
486	0.005	0.006	6.0×10^2
567	0.008	0.014	9.6×10^2
648	0.75	1.5	9.0×10^4
675	0.75	1.5	9.0×10^4
810	0.47	1.2	5.6×10^4

^aDetermined from the amount of radio-activity in each fragment relative to that in monomer DNA.

^bDetermined from optical density scans of Hind III digestion products.

^cCalculated from the % of each fragment in the genome, length of the fragment, and a genome size for *Geryon* determined by reassociation kinetic data (24). Using a ratio of the rate of reassociation for *Geryon* single copy DNA (4) to that of *Gecarcinus lateralis* (3) and the genome size of *G. lateralis* determined independently (25), the size of the *Geryon* genome was calculated to be 2.4 pg. Both single copy rates were computer-derived values of the expected rates of isolated single copy DNA as determined from reassociation of total DNA.

dimer, tetramer, and octamer are best fit by line (c), which represents 92% altered sites. This level of divergence cannot reflect the overall pattern since 62, 50, and 12%, respectively, of these fragments were digested to 47 and 40 bp fragments by Hha I. If a divergence of 16% was present in the fragments, more of the original Hha I sites would be altered.

Several explanations are possible for this nonrandom distribution of enzyme sites. Since the Hind III fragments were obtained from digests of total DNA, the presence of a multimeric series based on a single-repeating unit cannot be assumed from the sizes alone. It is necessary to demonstrate that a sequence relationship exists for this series of fragments. The 81 bp monomer contains both a Hha I site and a Hae III site (Fig. 3). The other Hind III fragments were digested with these enzymes to determine if they were multiples of this basic unit.

Analyses of the Multimers. A. The Monomer, Dimer, Trimer, and Tetramer. Initial digests of the monomer (V) and dimer (IV) with the restriction enzymes

spaced restriction sites. Such a probability distribution predicts that each successively larger multimer will be present in smaller amounts. Figure 2 shows the expected amounts of multimers for different degrees of divergence [dashed lines: (a) 3.7%, (b) 9.5%, and (c) 15.7%] and the experimentally determined amounts of Hind III fragments in *Geryon* DNA (solid line) which persist even with increases of either enzyme concentration or time of incubation. The ratios of the amounts of the different multimers (Table 1) after complete digestion are different from those of various satellite DNAs (12-14) and total DNAs (26,27). The deviation of the *Geryon* fragments from a linear plot indicates that the distribution of Hind III sites (Fig. 2; solid line) cannot be explained simply by random alteration of a formerly homogeneously spaced array of sites. The amounts of

Figure 3. Autoradiograms of Acrylamide Gels of Hha I and Hae III Digests of Monomer, Dimer, Trimer, and Tetramer DNAs. A Hinf I digest of ØX174 DNA was included in each gel (A,B: lanes 3). Measured sizes are given first; proposed sizes (Fig. 4) are in parentheses.

Mbo I, Taq I, Hpa II and Alu I left most of the DNA of each fragment undigested while with either Hha I or Hae III from 50 to 75% of each was digested. The digests with either of the latter enzymes contained two major and several minor fragments. Hha I digests of the monomer, dimer, trimer, and tetramer (III) are shown in Fig. 3A; lanes 1, 2, 4, and 5. The major products are fragments of 47 and 40 bp which represent between 30 and 60% of each digest. Direct sequencing by the method of Maxam and Gilbert (28) has shown that Hha I fragments of the same size from either the monomer, dimer, or tetramer have essentially the same sequences (29). The 47 and 40 bp fragments correspond to the two 5' termini of the monomer. An overlap of four bp for the Hind III site and two for the

Hha I site results in a repeating unit of 81 bp.

Although the reduced amounts of the trimer (Fig. 3A; lane 4) prevented its sequencing, sequence homology between it and the monomer, dimer, and tetramer is indicated by the presence of similarly sized major and minor fragments in a Hha I digest. Digests of the dimer, trimer, and tetramer contain fragments 130 and 127 bp in length while fragments of 212 bp are also present in the trimer and tetramer. The sizes of these minor fragments are close to those expected if one or two Hha I sites, respectively, were lost from a set of sites formerly spaced every 81 bp.

The proposed organization for the original spacing of Hind III, Hha I, and Hae III sites is shown in Fig. 4A from which the organization of the

Figure 4. Proposed Organization for the Original Spacings of Restriction Endonuclease Sites in *Geryon* DNA Fragments Produced by Digestion with Hind III. A diagrammatic representation of the sites is given in (A) for three restriction endonucleases: Hind III (○), Hha I (X), and Hae III (▲), in the arrangement of the original DNA. At each Hind III site, the size of a multimer is given in base pairs as measured from the left of the diagram and assuming the intervening Hind III sites have been inactivated. In (B) is given the array of fragments expected after Hha I cleavage of 5'-end-labeled Hind III fragments assuming random divergence has occurred. Fragments missing more than five sites are not shown but may be present in undetectable quantities in digests of appropriate multimers. In (C) is given the array of fragments expected for Hae III cleavage of 5'-end-labeled Hind III fragments as for Hha I sites in (B). In the text, the arrays in (B) and (C) are referred to as subset 1 fragments. A proposed organization is given for subset 2 DNA (D) from the octamer based on analyses discussed in the text.

monomer and its multiples can be deduced. Hha I fragments (Fig. 4B) other than those of 47 or 40 bp have lost one or more sites. Hha I cleavage of a population of 5'-end-labeled trimers (243 bp) would produce fragments of 47, 128, and 209 bp from the 5'-labeled termini shown at the left and fragments of 40, 121, and 202 bp from the 5'-labeled termini shown at the right. Any 81 bp fragment or its multiples would be internal, would lack a 5'-labeled end, and would escape detection in an autoradiogram. The presence in the Hha I digest of fragments whose sizes closely approximate those of fragments that have lost one site (128 bp from the 47 bp end and 121 bp from the 40 bp end) and two sites (209 bp from the 47 bp end and 202 bp from the 40 bp end) supports this organization. The sizes of the observed fragments are within reasonable limits of experimental variation expected for the multimeric series. Only for the smallest pair of fragments, 47 and 40 bp, was a 7 bp difference in size sufficient for detection by our electrophoretic con-

ditions. In the digest of the tetramer (Fig. 3A; lane 5), fragments of the sizes expected if three Hha I sites were lost (290, 283 bp), may be obscured by the large amount of undigested tetramer DNA.

The undigested fraction of each Hind III fragment presumably represents sequences that have lost all the Hha I sites, although a completely different repeating unit cannot currently be excluded. Since the Hind III fragments were obtained from total DNA, fragments of the same length may be derived from two or more distinct sequence classes. Alternatively, the undigested fraction in the larger multimers may include background fragments of heterogeneous sizes. In autoradiograms the undigested fraction will appear to comprise a larger amount of the total digest because it retains label at both 5' ends while each Hha I digestion product is labeled at only one end. That the undigested fragments represented incomplete digestion was ruled out by increasing both the amount of enzyme and the time of incubation.

Hae III digestion of the monomer and the next three consecutive multimers (Fig. 3B; lanes 1, 2, 4 and 5) produced the array of fragments predicted in Fig. 4C. The pairs of fragments present in the Hae III digests, one for each 5' end, correspond to those that have lost either no Hae III sites (75 and 10 bp), one site (156 and 91 bp), two sites (237 and 172 bp) or three sites (318 and 253 bp). The large amount of undigested tetramer (324 bp) obscures the 318 bp fragment. The 10 bp fragments are not shown in the figure. It can be seen that fragments recovered from Hha I or Hae III digests of the smallest four oligomers are those expected for an original sequence composed of repeating units of 81 bp.

Fifty to seventy-five percent of each of the Hind III fragments thus far described is digested by either Hha I or Hae III. Fractions of the monomer or dimer undigested by Hha I were also resistant to Hae III. It is unlikely that these results are due to the inhibition of the restriction enzyme since a three-fold increase in the amount of Hae III led to no further digestion. We conclude that the subset of sequences containing the Hha I sites almost completely overlaps the subset containing the Hae III sites. Since random mutation in a homogeneous population of repeating units would not be expected to lead to the elimination of both sites in one-fourth to one-half of the sequences comprising one multimer and to the presence of both sites in the remaining sequences, this is a surprising result. This distribution of Hha I and Hae III sites suggests two classes of sequences, one considerably more divergent than the other.

B. The Pentamer, Hexamer, and Heptamer. As seen in Table 1 and Fig. 2,

Figure 5. Autoradiogram of Acrylamide Gel Containing Hha I Digests of Pentamer, Hexamer, and Heptamer DNA. A Hind III digest of ØX174 DNA was included (lane 4). Proposed sizes (Fig. 4) are given in parentheses. The measured sizes of the proposed 47 and 40 bp fragments deviate more than for Hha I digests of other multimers due to the longer duration of electrophoresis, since they have migrated beyond the central portion of the gel where sizing is most accurate.

the fragments corresponding to the lengths expected for a pentamer, hexamer, or heptamer comprise very small fractions of a total Hind III digest. Because of the small amounts of each of these fragments and the relatively large background of heterogeneous fragments of similar size, quantification was less precise. At least 50% of each of these fragments is digested by Hha I to yield as major products the 47 and 40 bp fragments characteristic of the 81 bp repeating unit (Fig. 5; lanes 1, pentamer; 2, hexamer; and 3, heptamer). The minor digestion products can arise from the loss of one or more of the sites in the 81 bp spacing as found in the four smaller multimers. In summary, the organization of these fragments is similar to that of the smaller multimers.

C. The Octamers. The organization of the octamers [Fig. 1; fragments I (675 bp), and II (648 bp)] was examined by digestion of co-isolated fragments with Hha I (Fig. 6A; lane 1) and fragment I by Hae III (lane 4). The sizes of the Hha I cleavage products of co-isolated fragments were essentially the same as for digests of fragment I (Fig. 6A;

Figure 6. Restriction Endonuclease Analysis of Octamer and Decamer DNA. In (A) and (B) *Hinf* I digests of ØX174 were run in lanes 3 and 2, respectively. In (A) measured sizes are given for the octamer fragments in lanes 1 and 4; proposed sizes (Fig. 4) are given in parentheses. The deviations between measured and proposed sizes are less for the *Hae* III digest (lane 4) because a longer duration of electrophoresis allowed migration of the larger fragments to the central portion of the gel. (B) Shows octamer DNA digested first with *Hha* I, then with *Taq* I (lane 1); and decamer DNA digested with *Taq* I (lane 3).

the 675 and 648 bp fragments. Both the *Hha* I and *Hae* III digestion patterns of the fragments from the octamers are more complex than those of the smaller multimers. Apparently several sequence subsets are present; nevertheless, a fraction of the DNA of the octamers is clearly related to the smaller multimers.

Most *Hha* I digestion products of the octamers (Fig. 6A; lanes 1 and 2) correspond to the proposed organization for the smaller multimers. There are two deviations from the results expected for divergence from an original repeat length of 81 bp. First, the two fragments of 80 and 142 bp do not fit the proposed organization. Second, there is an inordinately large amount

of one of the size classes. Because of the decreased probability of altering neighboring sites by random mutation, there should be larger amounts of the fragments of sizes produced by the loss of one or two consecutive sites. In fact, quantification of the amounts of products in an Hha I digest of octamer fragments shows that 52% of the DNA is digested to a family of fragments that form a broad band with a mean of approximately 270 bp (Fig. 6A; lanes 1 and 2). This amount significantly exceeds that expected for the size class corresponding to the loss of three consecutive Hha I sites from each 5' end. While the very large amount of this fragment may represent a selective amplification of a divergent octamer, the presence of other fragments in the digest is sufficient to indicate amplification since those fragments corresponding to the 40, 47, 121, 128, 202, and 209 bp fragments comprise approximately 15% of the Hha I digest of the octamers (29). This represents 3×10^4 copies in contrast to the five copies of octamers predicted from 4% divergence.

The population of molecules comprising the octamers can be assigned to four classes according to the products of Hha I digestion: subset 1 yields restriction fragments whose sizes can be predicted from the 81 bp repeat; subset 2 yields fragments of the sizes predicted but in anomalously large amounts (i.e., the 270 bp class); subset 3 yields fragments of unexpected sizes (i.e., the 80 and 142 bp fragments); subset 4 is the fraction undigested by Hha I. The latter three categories may represent diversity in the types of alterations that have occurred in the basic repeat, including amplification, introduction of new Hha I sites, and the loss of the 81 bp spacing for Hha I sites. Fractions of subset 3 and 4 may be unrelated to the basic repeat. Subsets 1, 3, and 4 are also present in the smaller multimers; however, subset 3 accounts for a lower percentage of the smaller multimers.

The classes of fragments produced by digestion with Hae III (Fig. 6A; lane 4) are very similar to those produced by digestion with Hha I. Subset 2 is represented by fragments of 350 and 320 bp. They correspond to proposed sizes of 334 and 318 bp, both of which would be produced by cleavage at the same site. As for the Hha I/270 bp fragments, although these are of a size expected from the model, they are present in anomalously high amounts. The next six smaller fragments in the Hae III digest are clearly related to the proposed organization containing an 81 bp repeat, i.e., subset 1 DNA. The remaining fragments of sizes 55 to 24 bp do not appear to be related to the basic repeating unit and are therefore subset 3 DNA. The major digestion products are the 24 bp fragment in subset 3 and the two large fragments in subset 2. Undigested DNA is also present (subset 4 DNA). A diagram is

Figure 7. Autoradiogram of Acrylamide Gel Containing Restriction Endonuclease Fragments of Decamer DNA. Hinf I fragments of ØX174 (lanes 1 and 3) were electrophoresed with Hha I fragments of 5' end-labeled decamer DNA (lane 2) on a 7% polyacrylamide gel and Hae III fragments of the labeled decamer (lane 4) on a 10% polyacrylamide gel.

presented (Fig. 4D) for the proposed organization of subset 2 DNA in which there remains a cluster of sites from the basic sequence. Assuming that the subset 2 DNAs from both Hha I and Hae III digests overlap as in the monomer and dimer DNAs, the specific increases in the amounts of larger Hae III and Hha I products strengthen the possibility that sequences divergent from the basic 81 bp repeat have been amplified.

Further analysis of the octamers included digestion with Hha I, precipitation and treatment with Taq I (Fig. 6B; lane 1). Digestion with both enzymes produced fragments whose sizes were the same as those seen in Hha I digests of octamers. Unique to the double digest is a major fragment of 138 bp. Its presence indicates that a significant fraction of octamers has Taq I sites 138 bp from either one or both of the Hind III sites. When the decamer is digested by Taq I (Fig. 6B; lane 3), the 138 bp fragment is present but not as a major product.

Thus, if one of the octamer fragments is related to the decamer, it is not a simple relationship. It is possible that two classes of sequences are present in the octamers and that the decamer is related to only one of these. Nevertheless, the relationship of the octamer to the other multimers is strengthened by the presence of the Hha I and Hae III sites near its center. It seems unlikely that these three sites have arisen fortuitously at positions expected for an octamer of the basic 81 bp unit.

D. The Decamer. The cleavage products produced by digestion of labeled decamers with Hha I, Taq I, and Hae III were most similar to those sequences in the octamers that had diverged from the basic 81 bp repeat. In the Hha I digest (Fig. 7, lane 2) the fragment sizes characteristic of the 81 bp repeat

(viz., the 47 and 40 bp fragments) were not seen. Also missing were fragments of the sizes that would occur if one or two consecutive sites had been modified. The 270 bp fragments that comprise the bulk of the Hha I digest of the octamers account for only a minor fraction of this digest. The 80 and 142 bp fragments of the divergent subset 3 DNA are also present. Fragments of 165, 138, and 88 bp are recovered both from Taq I digests of decamers and from Hha I/Taq I double digests of octamers (Fig. 6B; lane 1 and 3). Hae III digestion of decamer DNA (Fig. 7; lane 4) did not yield the pairs of fragments expected for subset 1 DNA. Rather, a fragment of 55 bp (subset 3) was observed as a major fragment. In summary, there are several similar spacings of restriction sites in the decamer and that fraction of the octamer that differs from the basic 81 bp repeat.

Nonrandom Alterations in Highly Repetitive DNAs: Selective Mutation or Selective Amplification? The distribution of amounts of the fragments in the Hind III multimeric series digested from total *Geryon* DNA deviates from the distribution of multiples of a basic repeating unit observed in several isolated satellite DNAs (12-14) and in satellite DNAs digested from total DNAs (26,27). Either selective mutation or selective amplification is required to explain this organization. Mutation may be selective either by limited accessibility of specific DNA regions to mutagenic agents or by a propensity of specific DNA sequences to a particular alteration. For DNA regions to vary in accessibility to mutation, the nucleosome structure might dictate protected regions, e.g., sites approximately every 162 bp. Enhanced amounts of dimer (162 bp), tetramer (2 X 162 bp), hexamer (3 X 162 bp), and octamer (4 X 162 bp) might then be expected. The amount of hexamer is not enhanced but is equivalent to that expected for 4% divergence. Specific sites for mutation would require precise phasing of the repeating DNA units with the histones or other proteins in the chromatin complexes. While nucleosome phasing relative to DNA sequence has been suggested (30,31), other studies indicate random alignment of nucleosomes (32). The presence of more octamers than hexamers in *Geryon* DNA requires the additional constraint of a multinucleosome complex with every fourth nucleosome being less accessible to mutation than every third. It has been proposed that higher-order chromatin structure consists of units of eight nucleosomes (33). While the constraints within a multinucleosome may yield selective mutation, it is surprising that other repetitive DNAs have not contained a distribution of multimeric fragments similar to that of *Geryon*. For seven repetitive DNAs examined (34), there is considerable variability in the basic unit varying

from ~ 170 bp in the α -satellite of the African green monkey to 1360 bp in the calf satellite. Nevertheless, the fact that the repeating unit of many repetitive DNAs is ~ 170 bp or multiples of that length is compatible with some effect of nucleosome structure on DNA sequence, at least for repetitive DNAs. Sequences isolated by restriction endonuclease digestion of human (35) and hermit crab (6) DNAs also have a basic repeat of similar length. More recently, seven satellites have been described in calf DNA (36), all having basic repeat lengths of ~ 1360 bp. One of these, of $\rho = 1.715 \text{ g/cm}^3$, has two types of basic repeating units that differ only in the position of an internal Hha I site. Hae III digestion of this satellite shows no evidence of a 1360 bp repeat but rather suggests a short repeat of 12 bp (37). These observations coupled with those of *Geryon* DNA indicate that nucleosome phasing with specific DNA sequences would in any case be insufficient to explain the distribution of restriction sites in all repetitive DNAs.

Several other factors may contribute to sequence alterations in repetitive DNAs. One of these is the effect of repair enzymes on the nature of mutation. Repair of UV damage in certain strains of yeast causes the reversion of an ochre mutation primarily by A·T to G·C transitions (38,39). Slightly different sequences within a set of repeating units might then determine different pathways for subsequent mutation in some repeats. In addition, the mutation rate in the *rII* locus of T4 can be altered at a single site within an ochre codon by changes in neighboring base pairs (40); this specificity can explain the "hot spots" observed for the mutability of this locus (41,42). Effects of this type reflect single nucleotide changes and will differ for individual repetitive DNAs depending on their sequence. Even these instances of specific mutation seem inadequate to explain the distribution of Hind III sites in *Geryon* DNA. In particular, the relative amounts of Hind III multimers is difficult to explain by models of selective mutation alone. An alternate model of selective amplification requires only that the amplified segment of DNA be sufficiently small to be unrepresentative of the entire class in its relative amounts of multimers. Examples of non-random distributions of restriction sites have been seen in other DNAs as well (27). For the 1.688 g/cm^3 satellite of *Drosophila melanogaster* whose basic unit is ~ 365 bp, Carlson and Brutlag (43) show that some regions have fewer restriction sites than would be predicted on the basis of random inactivation. They also found a fragment of 15 monomer units, a highly improbable result for random mutation. In sequence analyses of African green monkey α -satellite whose basic unit is ~ 170 bp, particular variants were

detected which could not be explained by random mutation (44). These examples of nonrandom alterations coupled with those in *Geryon* DNA suggest that a selective amplification superimposed on a biased pattern of mutation would be a plausible scheme for the formation of at least some complex repetitive DNAs.

Selective Amplification of Small Segments of DNA. Selective amplification could occur in at least two ways. First, amplification of a small subset of repetitive sequences might occur after translocation to a different chromosome. Alternatively, only a small fraction of a repeated sequence at a single location might be selected for amplification. In *Geryon* the highly repeated sequences captured from total DNA as Hind III fragments cannot be assumed to be contiguous. Rather, they may be interspersed with single-copy DNA, which abuts two highly repetitive components of very different frequencies in this genome (4).

There could also be several types of amplification. It could occur in a subset that had diverged from the basic repeat by a saltatory mechanism (45) or by an unequal crossover mechanism (8,9). In either case, to account for the relative amounts of the Hind III multimers, it must be stipulated that only a small fraction of related repetitive sequences has undergone amplification. Replication of a segment 2 or 3 X 10³ bp long might easily produce enhanced amounts of particular multimers, since the total DNA in the Hind III multimers is 3 X 10⁸ bp. The apparent amplification of a divergent octamer indicates that the amplified segment is greater than 650 bp. A segment of that size or several times that size would still be sufficiently small such that the distribution of multimers would not reflect that of the entire population of sequences. Divergence of the original sequence and subsequent amplification of all related sequences would preserve ratios expected for simple random inactivation of regularly spaced restriction endonuclease sites.

Pattern of Evolutionary Alterations in *Geryon* DNA. A scheme of evolutionary changes in highly repetitive sequences that is compatible with the characteristics of *Geryon* DNA is a cyclic pattern of amplification followed by divergence and reamplification of the divergent sequences. This assessment is based on the reduced amounts of some multimers such as the trimer and enhanced amounts of others such as the tetramer and octamer. Except for the limitations that can be placed on the size of the segment selected for amplification, this evolutionary pattern is as described by Southern (46) for guinea pig α -satellite. In contrast, Fry and Salser (47) have suggested

that a library of satellite sequences is conserved by their association with genetic mechanisms fostering sympatric speciation. (In the sympatric model of evolution, two biological races adapt to different ecological niches within the same geographic range and form distinct species. Alternatively, in the allopatric model, geographic isolation precedes genetic isolation.) Chromosomal rearrangements, especially those facilitated by heterochromatic regions, have also been implicated in speciation (48-51). Conservation of specific sequences is probably not necessary for a role of repetitive DNAs in sympatric speciation. Heterochromatin and thus the highly repetitive DNAs localized to this region may facilitate chromosomal rearrangements that increase the potential for variant species. Increases in genetic variability are directly related to increased survival of a particular taxonomic group, and that survival will perpetuate a class of highly repeated DNAs. In addition, it is not necessary that the amplification of a particular DNA sequence initiates speciation but only that it fosters it.

CONCLUSIONS

Our observations are compatible with a pattern of multiple amplifications of relatively small segments of DNA compared to the complete set of related sequences. A single amplification would have produced the observed distribution of Hind III fragments only if two octamers were present for every tetramer in a contiguous piece of DNA containing no trimers, pentamers, hexamers, or heptamers. This is quite unlikely since the preamplified number of octamers for 4% divergence would be less than five copies per genome while the number of tetramers would be greater than 500. Amplifications of small segments of DNA might be fairly frequent; multiple amplifications within highly repetitive regions would have the same effect as conserving particular sequences.

ACKNOWLEDGMENTS

We thank John Valois for materials. This research was supported by the Office of Health and Environmental Research, U.S. Department of Energy, under contract W-7405-eng-26 with the Union Carbide Corporation. N.T.C. was supported as a predoctoral fellow by Grant GM 1974 from the National Institute of General Medical Sciences. *To whom correspondence should be sent.

REFERENCES

- 1 Davidson, E.H. and Britten, R.J. (1979) *Science* 204, 1052-1059
- 2 Vaughn, J.C. (1975) *Chromosoma* 50, 243-257
- 3 Holland, C.A. and Skinner, D.M. (1977) *Chromosoma* 63, 223-240

-
- 4 Christie, N.T. and Skinner, D.M. (1979) *Nucleic Acids Res.* 6, 781-796
 - 5 Beattie, W.G. and Skinner, D.M. (1972) *Biochim. Biophys. Acta* 281, 169-178
 - 6 Hinkle, N.F., Chambers, C.A. and Skinner, D.M. (1979) *J. Cell Biol.* 83, 194a.
 - 7 Manuelidis, L. (1976) *Nucleic Acids Res.* 3, 3063-3076
 - 8 Chambers, C.A., Schell, M.P. and Skinner, D.M. (1978) *Cell* 13, 97-110
 - 9 Smith, G. (1973) *Cold Spring Harbor Symp. Quant. Biol.* 38, 507-513
 - 10 Skinner, D.M. (1977) *BioScience* 27, 790-796
 - 11 Cordeiro-Stone, M. and Lee, C.S. (1976) *J. Mol. Biol.* 104, 1-24
 - 12 Southern, E.M. (1975) *J. Mol. Biol.* 94, 51-69
 - 13 Altenberger, W., Hörz, W. and Zachau, H.G. (1977) *Eur. J. Biochem.* 73, 393-400
 - 14 Fittler, F. (1977) *Eur. J. Biochem.* 74, 343-352
 - 15 van deSande, J.H., Kleppe, K. and Khorana, H.G. (1973) *Biochemistry* 12, 5050-5055
 - 16 Lillehaug, J.R. and Kleppe, K. (1975) *Biochemistry* 14, 1225-1229
 - 17 Roberts, R.J., Myers, P.A., Morrison, A. and Murray, K. (1976) *J. Mol. Biol.* 102, 157-165
 - 18 Maniatis, T., Jeffrey, A. and van deSande, H. (1975) *Biochemistry* 14, 3787-3794
 - 19 Prunell, A., Strauss, F. and Leblanc, B. (1977) *Anal. Biochem.* 78, 57-65
 - 20 Vereijken, J.M., van Mansfield, A.D.M., Bass, P.D. and Jansz, H.S. (1975) *Virology* 68, 221-233
 - 21 Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchison, C.A. III, Slocombe, P.M. and Smith, M. (1977) *Nature* 265, 687-695
 - 22 Schroeder, J.L. and Blattner, F.R. (1978) *Gene* 4, 167-174
 - 23 Slack, J. (1974) *Biopolymers* 13, 2241-2264
 - 24 Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) in *Methods in Enzymology*, Grossman, L. and Moldave, K., Eds., Vol. 29E, pp. 363-418, Academic Press, New York
 - 25 Musich, P.R. and Skinner, D.M. (1972) *J. Cell Biol.* 55, 184a
 - 26 Cooke, H.J. (1975) *J. Mol. Biol.* 95, 87-99
 - 27 Manteuil, S., Hamer, D.H. and Thomas, C.A., Jr. (1975) *Cell* 5, 413-422
 - 28 Maxam, A. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci.* 74, 560-564
 - 29 Christie, N.T. and Skinner, D.M. Submitted for publication.
 - 30 Musich, P.R., Maio, J.J. and Brown, F.L. (1977) *J. Mol. Biol.* 117, 657-677
 - 31 Chao, M.V., Gralla, J. and Martinson, H.G. (1979) *Biochemistry* 18, 1068-1074
 - 32 Tatchell, K. and Van Holde, K.E. (1979) *Biochemistry* 18, 2871-2880
 - 33 Stratling, W.H., Müller, U. and Zentgraf, H. (1978) *Exp. Cell Res.* 117, 301-311
 - 34 Maio, J.J., Brown, F.L. and Musich, P.R. (1977) *J. Mol. Biol.* 117, 637-655
 - 35 Manuelidis, L. and Wu, J.C. (1978) *Nature* 276, 92-94
 - 36 Macaya, G., Cortadas, J. and Bernardi, G. (1978) *Eur. J. Biochem.* 84, 179-188
 - 37 Kopecka, H., Macaya, G., Cortadas, J., Thiéry, J. and Bernardi, G. (1978) *Eur. J. Biochem.* 84, 189-195
 - 38 Sherman, F. and Stewart, J.W. (1974) *Genetics* 78, 97-113
 - 39 Lawrence, C.W., Stewart, J.W., Sherman, F. and Christensen, R. (1974) *J. Mol. Biol.* 85, 137-162
 - 40 Koch, R.E. (1971) *Proc. Natl. Acad. Sci.* 68, 773-776
 - 41 Benzer, S. and Freese, E. (1958) *Proc. Natl. Acad. Sci.* 44, 112-119
 - 42 Benzer, S. (1961) *Proc. Natl. Acad. Sci.* 47, 403-415
-

- 43 Carlson, M. and Brutlag, D. (1977) Cell 11, 371-381
- 44 Rosenberg, H., Singer, M. and Rosenberg, M. (1978) Science 200, 394-402
- 45 Britten, R.J. and Davidson, E.H. (1971) Q. Rev. Biol. 46, 111-138
- 46 Southern, E.M. (1970) Nature 227, 794-798
- 47 Fry, K. and Salser, W. (1977) Cell 12, 1069-1084
- 48 White, M.J.D. (1973) Animal Cytology and Evolution, 3rd edn., Cambridge University Press, London and New York
- 49 Lejeune, J., Dutrillaux, B., Rothoré, M. and Prieur, M. (1973) Chromosoma 43, 423-444
- 50 Bush, G.L., Case, S.M., Wilson, A.C. and Patton, J.L. (1977) Proc. Natl. Acad. Sci. 74, 3942-3946
- 51 Barnes, S.R., Webb, D.A. and Dover, G. (1978) Chromosoma 67, 341-363