J. Mol. Evol. 15, 277–289 (1980)

Journal of Molecular Evolution

© by Springer-Verlag · 1980

DNA Sequence Organization in the *Thysanura Thermobia domestica*

Cynthia K. French and Jerry E. Manning

Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92717, USA

Summary. Cot and chemical analysis show that the haploid genome size of Thermobia domestica is $3-4 x 10^9$ nucleotide pairs. Of this DNA 33% is single copy sequences and 67% is repetitive sequences. The repetitive sequences are predominantly 300 nucleotides in length and are interspersed among the single copy sequences in a short period interspersion pattern similar to that observed in Xenopus and many other higher eucaryotes. The DNA sequence organization observed in Thermobia is compared with that of other more highly evolved insects.

Key words: DNA sequence organization

Introduction

Two general patterns of DNA sequence organization have been observed in eucaryotic chromosomes. In both patterns repetitive DNA sequences have been found to alternate with single copy sequences. The salient structural difference between the two patterns is the length of the repetitive and single copy DNA sequences. The genome pattern which has the widest phylogenetic representation is termed short period interspersion and is characterized by the major fraction of the DNA containing repetitive sequences of average length 300 nucleotides. This type of sequence arrangement is found in all major animal (Davidson, et al., 1973; Britten et al., 1974; Chamberlin et al., 1975; Davidson et al., 1975; Goldberg et al., 1975) and plant phyla (Wolbot and Dove, 1976; Zimmerman and Goldberg, 1977).

A noticably different pattern of DNA sequence organization is the long period interspersion pattern in which long single copy regions, several thousand nucleotides in length, appear to be interspersed by long repeat sequences. This pattern was first

Abbrevations: HAP-hydroxyapatite, Cot = mole nucleotides x liter^{-1} s. N = sodium phosphate, pH 6.8

0022–2844/80/0015/0277/$ 02.60

observed for the genomic DNA of *Drosophila melanogaster* (Manning et al., 1975; Crain et al., 1976A) and has subsequently been observed in the genomes of *Chironomus tentans*, a dipteran (Wells et al., 1976), *Apis mellifera*, a hymenopteran (Crain et al., 1976A), and the water mold, *Achlya* (Hudspeth et al., 1977). Of these, the organism whose genome which has been most extensively characterized is *Drosophila melanogaster*. In *Drosophila* the repetitive sequences average 5,600 nucleotide pairs in length and are interspersed among single copy sequences whose average length exceeds 13,000 nucleotide pairs. Furthermore, the possibility that even a minor fraction of the *Drosophila* genome may exist in a short period interspersion pattern has been clearly excluded (Crain et al., 1976B). Interestingly, an interspersion pattern intermediate between the long and short patterns has been recently described for the chicken genome (Eden and Hendrick, 1978). Here, about 50 % of the genome contains repetitive sequences of length 2,000 nucleotides interspersed among single copy sequences of average length 4,500 nucleotides. The remainder of the genome contains single copy sequences of length 17,000 nucleotides or greater with no observed interruption by repetitive sequences.

Although the long period interspersion pattern has been observed primarily in insects, it is clear that not all insects exhibit this interspersion pattern. Studies on other related insects such as the lepidopteran *Antheraea pernyi* (Efstradiatis et al., 1976), and a dipteran, *Musca domestica* (Crain et al., 1976B), show that the DNA of these two insects is arranged in a short period interspersion pattern qualitatively similar to that of *Xenopus* and most other eucaryotes. Thus, with the appearance of both long and short period interspersion patterns in the class *Insecta*, it is evident that a major structural change in the pattern of DNA organization has occurred during the course of insect evolution.

We dot not know the significance of these two different patterns of genome sequence organization or what the contribution of these two structural arrangements is to normal cellular processes. The evolutionary history and origin of short and long period interspersion is also unclear. In the class *Insecta*, we are fortunate to have a number of organisms displaying both patterns. Thus, by examining the DNA sequence organization of the most primitive living insect, we should be able to determine the possible evolutionary progenitor of the earliest sequence pattern within this class of organisms.

Thermobia domestica, the firebrat, was chosen for our studies because of its well characterized evolutionary relationship to all other insects. Although it has been difficult to assign a clear relationship between the evolutionary progressions of all orders in the class *Insecta*, it is clear that *Thermobia* is the most primitive living insect which is an ancient survivor of the ancestral wingless insects from which winged insects arose (Anderson, 1973). From available comparative embryology and anatomy, it can be concluded that the order, Thysanura, of which *Thermobia* is a member, shares many of the specialized embryological and structural features common to each of the diverged Apterygote (wingless) and Pterygote (winged insects) subclasses in the class *Insecta* (Imms, 1964).

In these studies, we show that the DNA of the most primitive living insect, *Thermobia domestica*, exhibits a short period interspersion pattern which is qualitalively similar to both *Musca domestica* and *Antheraea pernyi*.

Materials and Methods

Culture Conditions

Thermobia domestica were maintained in plastic growth chambers containing tetramin as a food source. The growth chambers were held at 37° C and 80 % humidity. Organisms used for DNA isolation were collected at 90–120 days after hatching and stored at -70° C.

Preparation of DNA

Unlabeled *Thermobia domestica* DNA was extracted from crude nuclear pellets prepared from adults of *Thermobia*. The DNA purification was essentially identical to the method described by Laird and McCarthy, 1969, but included several additional purification steps as previously described (Manning et al., 1975). The purified DNA was stored in 0.05 *M* sodium phosphate buffer, pH 6.8, at - 70° C.

Thermobia DNA containing radioactive label was prepared by spraying 2 mCi of ^{3}H-thymidine (61 Ci/mmole) over the food meal ingested by these organisms. Approximately 200 immature organisms were cultured on the ^{3}H-labeled meal for an eight day period. The insects were then collected, quick-frozen and the nuclear DNA extracted. The specific activity of the purified DNA was 1.0 x 10^3 cpm/μg.

Shearing of DNA

DNA fragments of the required size were prepared by mechanical shearing in the Virtis 60 homogenizer. Conditions used for shearing the DNA were exactly as described by Davidson et al., 1973 and Britten et al., 1974.

Determination of DNA Fragment Lengths

Weight average lengths of single strands of DNA were determined by velocity sedimentation through an alkaline 5–20% sucrose gradient. Eco RI restriction fragments of ^{3}H-labeled DNA from bacteriophage G_4 were sedimented in each gradient and used as an internal molecular weight standard. Sedimentation coefficients were converted to length in nucleotides according to the equation of Studier, 1965. It should be noted that the molecular lengths reported represent an average, not a uniform size, and that each DNA sample contains a distribution of fragment sizes.

Microfluorometry of DNA

Cells from *Thermobia* were prepared by mechanical extraction of the adult tissues. As a standard for quantitation of DNA by microfluorometry, Schneider's cells, grown in Schneider's *Drosophila* medium enriched with 15 % fetal calf serum (Grand Island Biological, Grand Island, N.Y.), were prepared. Cells were fixed on acid-washed glass slides using Carnoy's reagent (3parts ethanol;1 part glacial acetic acid) for ten minutes. Following fixation, slides were washed three times with double-distilled H_2O, incubated ten minutes at 60° C in 1 N HCl, washed several times with H_2O, stained with Shiffs Feulgen reagent for 1–5 h, and air dried overnight and mounted with a coverslip using permount.

In each sample, the fluorescence intensity of 100 nuclei was monitored using a Zeiss RA microscope with epifluorescence attachments. Approximately 100 measurements were taken for each nuclei. The results were then expressed in pg amounts using 0.24 pg as the size of the diploid genome of the Drosophila standard. The instrument was equipped with a Liconix He-cd laser for incident light excitation of the specimen. Emitted light intensities were measured with a Nametric 10s microspectrophotometer, interfaced with a Hewlitt-Packard calculator. The background fluorescence from the microscopic field was shielded by a fixed diaphragma $5\mu-10\mu$ in diameter.

Reassociation and Hydroxyapatite Fractionation of DNA Fragments

Samples of DNA were denatured with 1 *N* NAOH and neutralized with 2.0 *M* NaH_2PO_4, pH 4.3, as previously described by Kram et al., 1972. DNA was reassociated in low salt, 0.12 *M* sodium phosphate buffer, pH 6.8, at 60° C, or in 0.5 *M* sodium phosphate buffer at 65° C. Aliquots of DNA were renatured in sealed capillary pipettes to prevent evaporation. At appropriate Cot values, samples were removed and immediately frozen at -70° C. The extent of DNA reassociation in each sample was determined by separation of single and double stranded DNA fragments by hydroxyapatite chromatography as previously described (Davidson et al., 1973). Quantitation of the amount of DNA in the 0.12 *M* or 0.5 *M* Nap fractions was based on either the absorbance of the fractions at 260 nm (corrected for light scattering at 340 nm) or on the amount of radioactivity present in the fractions. Recovery was quantitatively greater than 95% for each sample. Reassociation kinetics were analyzed by a least squares fitting procedure which has been described in detail (Pearson et al., 1977).

Sl Nuclease Digestion of Single-strand DNA and Measurement of the Lengths of Nuclease-Resistant Fragments

DNA samples were reassociated in 0.12 *M* Nap at 60° C to obtain fast and slow repetitive DNA components. The Cots for *Thermobia* repetitive DNA were Cot < 4.0 for the fast fraction and $4.0 <$ Cot < 250 for the slow fraction. After reassociation to appropriate Cot, the duplex DNA was isolated by binding to HAP, dialyzed into 0.01 *M* NaCl, 0.005 *M* PIPES buffer, pH 6.7, adjusted to 0.15 *M* NaCl, 0.005 *M* PIPES buffer (pH 4.4), .025 *M* sodium acetate, 1 *mM* $ZnSO_4$ and 5 *mM* mercaptoethanol. Sl nuclease (1000 units/ml) were added and the samples were incubated for 45 min at 37° C. DNA concentration was 50 μg/ml. The reaction was terminated by the addition of 2.0 *M* Nap to 0.12 *M* final concentration. The DNA duplexes were collected by HAP chromatography and chromatographed on a 1.25 x 75 cm Sepharose 2B–CL column in 0.12 *M* Nap at 60° C.

Optical Melting of DNA

Thermal denaturation profiles of reassociated DNA samples were obtained by melting the DNA in 0.12 *M* Nap and monitoring the absorbance of the solution at 260 nm in a Gilford model 2400 recording spectrophotometer. The hyperchromicity was calculated according to the formula: $H = \frac{A260(98^\circ C) - A\,260(60^\circ C)}{A260(98^\circ C)}$ where H is the hyper-

chromicity (Davidson et al., 1973; Britten et al., 1974). The Tm was taken as that temperature which produces a 50% increase in hyperchromicity. The fraction of mispaired bases was calculated using the relationship that a 1° C reduction in Tm, after correction for the effect of fragment size, equals 1 % base pair mismatch (Davidson et al., 1973).

Liquid Scintillation Counting

4.0 ml samples from the hydroxyapatite column, in 0.18 *M* Nap were counted in 10 ml of Aquasol (New England Nuclear Boston, Mass.). TCA precipitates were collected on GFA filters and counted in a standard toluene based scintillation solution.

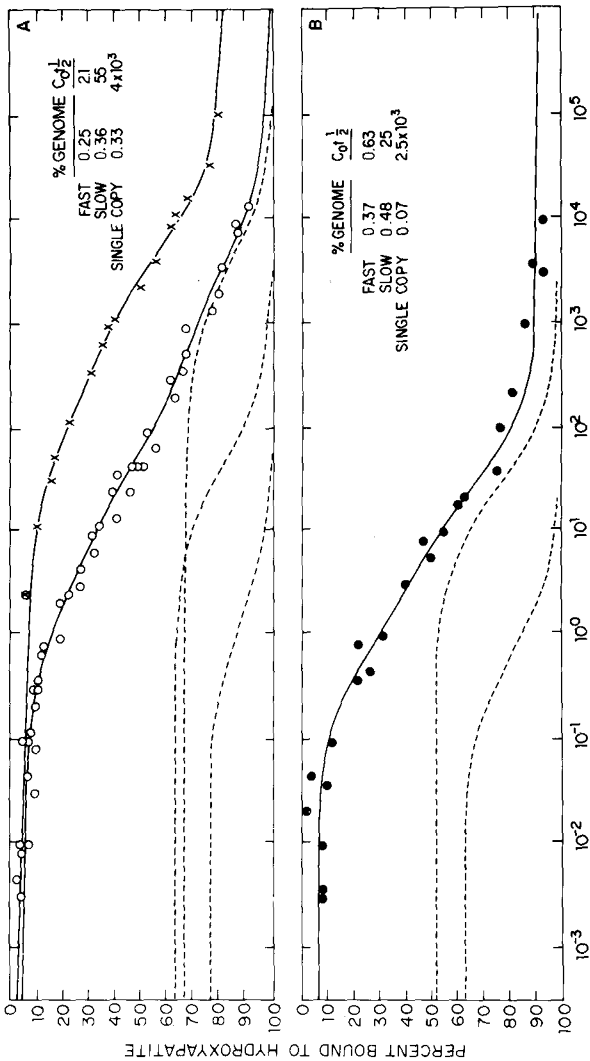

Fig. 1 A, B. Hydroxyapatite reassociation kinetics of *Thermobia* DNA. *Thermobia* DNA of fragment lengths 0.4 kb and 2.1 kb were reassociated as described in Materials and Methods. The fraction of fragments containing duplex regions was measured at the indicated equivalent Cot values by hydroxyapatite chromotography. (A) The curve for the 0.4 kb fragments (○) represents a least squares computer analysis for three second order components. In order to more accurately determine the rate constant for the single copy component, a fraction of ^{3}H-labeled 0.4 kb tracer DNA was mixed with a 33-fold excess of total 0.4 kb unlabeled driver DNA. The curve (x) represents a least squares analysis for two second order components. (B) For the reassociation of 2.1 kb fragments (●), three second order components are assumed in the least squares analysis. The dashed curve represents the components of the computer solution which can be summed to give the total curve. For this solution the rate of each component was fixed at a value obtained from the reassociation of a ^{3}H-labeled kinetic fraction, enriched for the single copy component in the presence of unlabeled unfractionated DNA

Table 1. Kinetic analysis of *Thermobia Domestica* DNA

Fragment length	Component	Fraction of[a] fragments	K pure[b]	K whole[c]	Cot 1/2 whole[c]	Number of copies per in genome	Kinetic[d] complexity
0.4 kb	foldback	0.03					
	fast	0.25	1.7	0.64	2.1	2.1×10^3	3.7×10^5
	slow	0.36	5.28×10^{-2}	1.9×10^{-2}	55	78	1.9×10^7
	single copy	0.33	7.15×10^{-4}	2.38×10^{-4}	4.3×10^3	1	1.4×10^9
2.1 kb	foldback	0.06					
	fast	0.37	4.2	1.6	0.63	4.0×10^3	2.37×10^5
	slow	0.48	8.08×10^{-2}	3.9×10^{-2}	25	98	1.25×10^7
	single copy	0.07		4.0×10^{-4}	2.5×10^3	1	
Single copy	residual repeat	0.26	0.46×10^{-1}	0.119×10^{-1}			0.22×10^8
0.4 kb	single copy	0.47	0.453×10^{-3}	0.211×10^{-3}		1 (e)	0.22×10^{10}

[a] Computed from the reassociation analysis of 0.4 kb or 2.1 kb total DNA fragments, the renaturation kinetics of ^{3}H-labeled single copy fractions. Because of interspersion, these values do not represent actual amounts of each component in the genome

[b] K pure is the parameter which describes the reassociation of a homogeneous kinetic component of which 100 % of the reassociated, DNA fragments bind to hydroxyapatite. The K pure value is computed from the empirical K value obtained from reassociation of DNA fractions enriched in the kinetic components. The relationship between the empirical K and the K pure is: K pure = K/Q. Where Q is the amount of the fraction comprising the kinetic component

[c] K whole and Cot 1/2 whole are the kinetic parameters which describe the reassociation of a kinetic component in the presence of unfractionated DNA. The relationship between K whole and Cot 1/2 whole is Cot 1/2 whole = 1/K whole

[d] The kinetic complexity is expressed in nucleotide pairs. These values were computed relative to the complexity of *E. coli* DNA (4.2×10^6 NPT) and the rate constant of 0.4 kb fragment of *E. coli* ($0.22\ M^{-1}\ S^{-1}$)

[e] Assumed value of 1 for the single copy component

Results

Reassociation Kinetics of Thermobia Domestica Nuclear DNA

The reassociation kinetics of nuclear DNA fragments of two different average lengths were measured by hydroxyapatite chromatography (Fig. 1A–B). Reassociation of the 0.4 kb (kb, kilobases; 1000 nucleotides, or nucleotide pairs for single and double-strand DNA, respectively) fragments occurs over at least six decades of Cot and the curve that best fits the data points as determined by least squares analysis has three kinetic components. As show in Table 1 and Fig. 1A, analysis of the reassociation profile of fragments of average length 0.4 kb identified a fast repetitive component with a reiteration frequency of 2100, present in 25 % of the fragments; a slow repetitive component with a reiteration frequency of 78, present in 36 % of the fragments; and a single copy component present without repetitive DNA elements in 33 % of the genome. Two to three percent of the 0.4 kb fragments bind to hydroxyapatite at a Cot less than 2×10^{-4}, and, thus, represent fold back or zero-time binding DNA. Also, 4–6 % of the DNA sequences failed to reassociate at the highest values of Cot attained.

In order to accurately determine the reassociation rate constant for the putative single copy sequences, trace amounts of ^{3}H-labeled DNA, enriched for single copy sequences, were prepared by removing from the total DNA fragments that bind to hydroxyapatite after reassociation to Cot 1500. The labeled fraction that did not bind to hydroxyapatite (single copy DNA) was then reassociated with excess total

Thermobia DNA. As shown in the upper curve of Fig. 1A, the tracer DNA is enriched for the putative single copy sequences.

The questions of whether this class of slowly reassociating sequences is single copy can be answered by comparing the genome size determined by the reassociation rate constant of the putative single copy DNA with that obtained by direct chemical analysis (Laird, 1971; Wetmur and Davidson, 1968). Under the conditions used in these experiments, the reassociation rate constant for 400 nucleotide fragments of *E. coli* DNA is 0.22 M^{-1} s^{-1}. The genome size of *E. coli* is 4.2 x 10^6 nucleotide pairs (Cairns, 1963). The rate constant for the presumed single copy sequences in *Thermobia* nuclear DNA is 0.21 x 10^{-3} M^{-1}. Based on this kinetic measurement, the haploid genome of *Thermobia* contains 4.4 x 10^9 nucleotide pairs of DNA.

A chemical determination of the amount of DNA per nucleus was performed by measurement of the laser beam induced fluorescence of DNA in Feulgen stained nuclei of whole adult tissue squashes of *Thermobia*. The results indicated a genome size of 6.1 x 10^9 nucleotide pairs (see Materials and Methods). This value is in reasonable agreement with the 8.8 x 10^9 nucleotide pairs/nucleus (4.4 x 10^9 nucleotide pairs/ haploid genome x 2) expected for a diploid genome based on our kinetic studies. Therefore, the most slowly reassociating sequences in *Thermobia* nuclear DNA represent single copy sequences, and the haploid genome size based on both chemical and kinetic analysis is between 3.05 x 10^9 and 4.4 x 10^9 nucleotide pairs.

Interspersion by Repetitive and Single Copy DNA Sequences

DNA sequence interspersion is defined as an alternating arrangement of repetitive and single copy DNA sequences. One method of determining whether these different classes of sequences are interspersed within DNA fragments of a given length is to compare the reassociation profile obtained for long and short DNA fragments, respectively (Davidson et al., 1973; Britten et al., 1974; Chamberlin et al., 1975; Davidson et al., 1975). If a DNA fragment contains both a repetitive sequence and a single copy sequence, the entire fragment will bind to hydroxyapatite after reassociation of only the repetitive portion of the fragment. Thus, if the two classes of sequences were interspersed at the chosen fragment length, one would observe an increase in the fraction of the total DNA reassociating as a repetitive component with a concomitant decrease in the single copy component. We initiated a study of sequence interspersion in the *Thermobia* genome by examining the reassociation profile of DNA fragments 2.1 kb in length (Fig. 1B).

It was observed that both repetitive components increase in amount, with a concomitant decrease in the single copy component. The fast component increased from 25 % to 37 % and the slow component increased from 36 % to 48 % of the DNA. The total increase of the two repetitive components, 24 %, is approximately equal to the total decrease, 26 %, observed for the single copy component. Since 33 % of the genomic DNA is single copy these results indicate that at least 72 % (0.24/0.33=0.72) of the single copy component is interspersed with either or both repetitive sequence classes within fragments 2.1 kb in length. The remaining single copy DNA (7–8 % of the total DNA) does not appear to be interspersed with any observable repetitive sequences at this fragment length.

Hyperchromicity and S_1 Nuclease Studies on Reassociated Fragments of Thermobia Domestica DNA

When short repetitive sequences are interspersed with single copy sequences, the structures formed after reassociation of long DNA fragments are short duplex regions adjacent to single-stranded regions. The hyperchromicity and the S_1 nuclease resistance of the reassociated fragments are, therefore, substantially less than that of native duplex fragments. Such decreases have previously been shown to be typical of genomes in which the short period interspersion pattern is prevalent (Davidson et al., 1973; Davidson et al., 1975; Goldberg et al., 1975). Accordingly, we have examined the arrangement of repetitive and single copy DNA sequences in the genome of *Thermobia* by measuring the optical hyperchromicity and S_1 nuclease resistance of reassociated long fragments of DNA.

First, the two repetitive components in *Thermobia* DNA were isolated. DNA fragments 2.7 kb in length were reassociated to Cot 4.0, and duplex-containing molecules were collected by hydroxyapatite chromatography. An aliquot of the DNA fraction bound to hydroxyapatite at Cot 4.0 was removed, and the hyperchromicity was determined. The DNA fraction remaining single stranded at Cot 4.0 was reassociated to Cot 250, to isolate the slow repetitive component. Duplex-containing molecules were again collected on hydroxyapatite, and a small aliquot removed for hyperchromicity measurements. These two DNA fractions were subjected to a limited S_1 nuclease digestion to cleave single strand tails from the duplexes. The S_1 nuclease digested samples were again passed over hydroxyapatite to separate repetitive duplexes from the digested single-strand tails. These duplex structures were then thermally denatured to determine the Tm and hyperchromicity after S_1 treatment. The results of the experiment are summarized in Table 2.

Table 2. Hyperchromicity and S_1 nuclease studies on reassociated *Thermobia* DNA

		Assay						
Component	Fragment length[a]	Fraction bound to HAP	Hyperchrom. before S_1 digestion[b]	Hyperchrom. following S_1 digestion[b]	Tm(°C) before S_1	Tm(°C) after S_1	Average duplex content c	Average duplex content d
fast	2720	0.39	0.093	0.21	78	78	0.26	0.28
slow	1764	0.37	0.097	0.23	75.0	76.2	0.28	0.30

Native Tm = 85.8° C
Hyperchromicity = 26.6 %

[a] Fragment lengths were measured following the hydroxyapatite fractionation of the reassociated fast and slow repetitive fractions by velocity sedimentation through an alkaline 5–20 % sucrose gradient as described in Materials and Methods

[b] Hyperchromicity is calculated as the fraction of denatured optical density between 60° C and 98° C. The hyperchromicity for native DNA is given

[c] The reassociated sample was fractionated on hydroxyapatite and treated with S_1 nuclease

[d] The average duplex content following hyperchromicity measurements is calculated: $D = (H - 0.026)/(0.266-0.026)$ where D is the average duplex content of the reassociated duplexes. H is the observed hyperchromicity value following S_1 digestion in these experiments. The correction of 0.026 is the hyperchromic change due to the collapse of single strand DNA between 60° C and 98° C

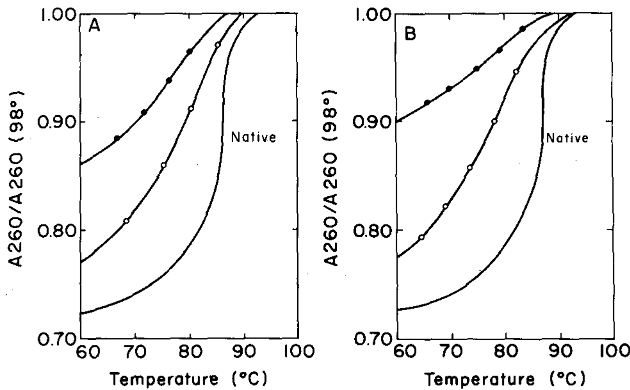

Fig. 2 A and B. Thermal denaturation of reassociated *Thermobia* DNA.
A. Total *Thermobia* DNA of average length 2.7 kb was reassociated to Cot = 4.0, and the duplex fragments were isolated by hydroxyapatite chromotography. Thermal denaturation profiles for the very fast repetitive fraction before S_1 nuclease digestion (•) and following S_1 digestion (○) are indicated. **B.** Represents the thermal denaturation profiles of fast repetitive duplexes isolated at $4.0 < Cot < 250$. A melt of native *Thermobia* DNA is included as a reference. The Tm^0 and hyperchromocity values for the repetitive DNA duplexes are listed in Table 2

Both S_1 nuclease and hyperchromicity measurements indicate that only a small portion (30 %) of the DNA in each repetitive component is double stranded. The observed hyperchromicity for both repetitive components was between 9 and 10 % (Fig. 2). This represents a reduction in hyperchromicity of approximately 16.6 to 17.6 % compared to the hyperchromicity of native duplex DNA. Since the hyperchromicity of purely single stranded DNA is about 2.5 % and that of native DNA is 26.6 %, the fraction of double stranded DNA in the HAP bound fractions may be estimated by $D = (H-0.026)/(0.266-0.026)$ (Davidson et al., 1973). For the fast and slow repetitive components the double stranded fractions are calculated to be 28 % and 30 % respectively.

Under the S_1 digestion conditions employed in these experiments, only single strand DNA is digested while duplex regions containing extensive mismatching remain intact (Chamberlin et al., 1975). Table 2 presents the results of the S_1 digestion experiments. Approximately 26 % of the length of the Cot 4.0 hydroxyapatite-binding fragments are actual duplex structures while 28 % of the Cot 250 fraction is duplex. These values are in reasonable agreement with the hyperchromicity measurements. From the observed decrease in hyperchromicity and the lower percentage of duplex sequences in the reassociated fragments after S_1 digestion, it is concluded that DNA fragments 2.7 kb in length contain both repetitive and single copy sequences, characteristic of a short period interspersion pattern.

Of further interest in this set of experiments is the observed decrease of the Tm^0 as compared to native DNA. Following S_1 treatment, the Tm^0 for the Cot 4.0 and 250 fractions are 78° C, and 76.2° C respectively. The Tm^0 value both before and following S_1 treatment is about 10° C below that of the native Tm^0 of 85.5° C. This indicates that the repetitive duplexes in both fractions contain approximately 10 % base mismatch (Davidson et al., 1973; 1974).

Repetitive Sequence Length of Thermobia

A direct estimate of the size distribution of the repetitive DNA sequences can be made by gel filtration chromatography of S_1 nuclease-resistant duplexes in each repetitive DNA fraction (Davidson et al., 1973; Davidson et al., 1975; Manning et al., 1975). Reassociated DNA fragments of 2.7 kb and 1.7 kb in length, respectively, were digested with S_1 nuclease as described previously. The S–1 resistant duplexes were collected on hydroxyapatite, and the size distribution determined by column chromatography on

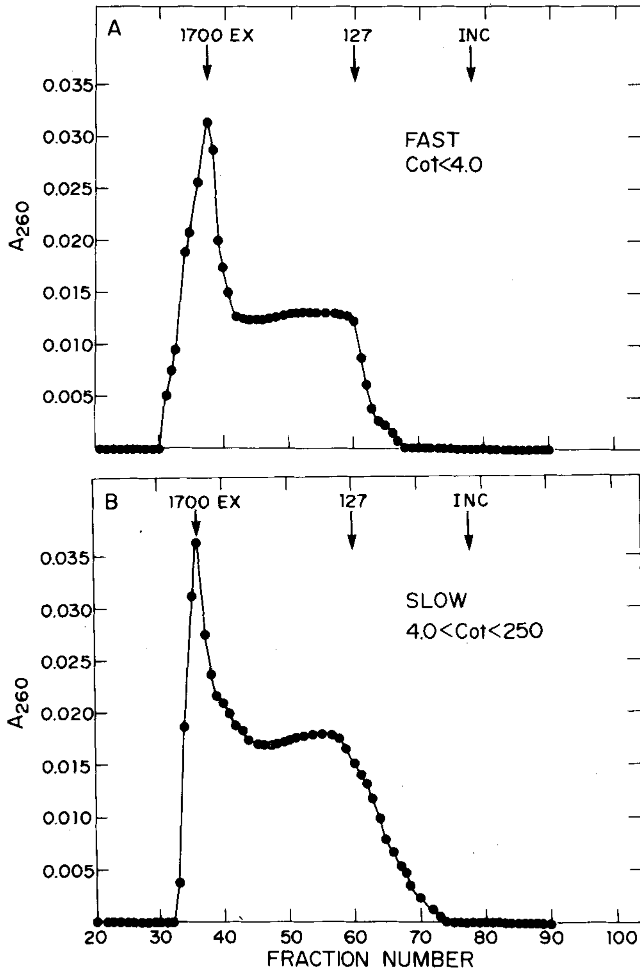

Fig. 3 A and B. Length distribution of *Thermobia* repetitive duplexes.
A. Total *Thermobia* DNA 2.2 kb in length was reassociated to Cot 4.0 and isolated on HAP. The duplexes were then treated with S_1 as described in Materials and Methods. The enzyme digest was then chromatographed on a Sepharose 2B–Cl column at 60° C with long native DNA as an exclusion marker and $^{32}PO_4^3$ – as an inclusion marker. **B.** Isolated duplex 4.0 $<$Cot$<$250 were also chromatographed as previously described.

The duplex length of the size markers was measured by neutral analytical band sedimentation in the Model E ultracentrifuge

Sepharose 2B–CL as shown Fig. 3. The data indicate that for the fast component there is a heterogeneous distribution of fragment lengths. Approximately 58 % of the total mass is excluded from the column indicating a duplex length $\geq$ 1.7 kb. The remaining 42 % of the duplexes chromatograph with an average distribution in the range of 0.3 kb. The elution profile of the slow repetitive components indicates that at least 44 % of the duplexes are excluded at a length $\geq$ 1.7 kb, while 56 % of the duplexes are heterogeneously distributed around a mode of 0.3 kb. These results indicate that the size of the fast and slow repetitive DNA sequences is qualitatively similar to those found in most eukaryotic genomes.

Discussion

There are two principal results of our studies. First, based on the reassociation constant of the single copy fraction of total nuclear DNA as well as chemical analysis of the DNA of whole nuclei, the haploid genome size of *Thermobia domestica* is 3–4 x 10^9 nucleotide pairs. Thirty-three percent is single-copy sequences, and approximately 67 % is repetitive sequences. The repetitive fraction can be divided into two repetition frequency classes with average repetition frequencies of 2.1 x 10^3 copies for the fast fraction, occupying 25 % of the genome, and 78 copies of the slow repetitive fraction, occupying 36 % of the genome. Secondly, the dominant pattern of DNA sequence organization in the *Thermobia* genome is short period interspersion. Between 72 % and 90 % of the single copy sequences are interspersed with repetitive sequences within fragments 2.1 kb or less. At this fragment length, eight percent of the single copy sequences appear not to be interspersed with any detectable repetitive sequences.

Since *Thermobia* represents the most primitive living insect whose genome size and DNA sequence organization has been characterized, some interesting evolutionary implications may be addressed. There are now several lines of evidence which show that a significant decrease of nuclear DNA content accompanies the evolution of species within certain taxonomic groups of both plants and animals (Stebbins, 1966; Hinegardner, 1968; Rees and Hazarika, 1969). For example, in an extensive survey of the teleost fishes, Hinegardner (1968) has demonstrated that within the teleost group the more advanced species of fishes have less DNA than the more primitive species. In the same vein a diminution of nuclear DNA content in several species of Lathyrus parallels their taxonomic order of advancement (Rees and Hazarika, 1969). These studies suggest that within a given taxonomic group those members which are more primitive and closely related to ancestral forms will have a larger nuclear DNA content than the more evolutionally advanced members. Although the number of examples within the class *Insecta* (Table 3) is certainly less complete than that presented for the teleost group, it is apparent that the more highly evolved insects contain significantly less nuclear DNA than their ancient ancestor, *Thermobia*. Although this apparent correlation between nuclear DNA content and trends in insect evolution will remain tentative until a more comprehensive survey of DNA values in advanced and primitive insects has been completed, it is, nevertheless, interesting to consider possible structural changes in the genomic DNA which may accompany diminution of DNA content in the nucleus.

Table 3. Summary of interspersion patterns and genome size in the class insecta

Organism	1C genome size (pg DNA)	Interspersion pattern	Reference
Drosophila melanogaster	0.12	D	Manning et al., 1975; Crain et al., 1976A
Chironomus tentans	0.2	D	Wells et al., 1976
Apis mellifera	0.35	D	Crain et al., 1976B
Sarcophaga bullata	0.60	D	French and Manning
Musca domestica	0.80	X	Crain et al., 1976B
Antheraea pernyi	1.0	X	Efstratiadis et al., 1976
Thermobia domestica	5.6	X	

D = *Drosophila*, long
X = *Xenopus*, short

Of those insects whose DNA sequence pattern is known, *Thermobia* has the highest nuclear DNA content, followed in order by *Antheraea* and *Musca*. Interestingly, the genomic DNA of these three insects is characterized by a short period interspersion of repetitive and single copy DNA sequences. In contrast, the arrangement of the repetitive and single copy DNA sequences in those insects with smaller genome sizes (*Drosophila, Chironomus, Apis* and *Sarcophaga*) is characteristic of the long period interspersion pattern. Although the number of examples is certainly limited, this intriguing correlation between the genome size and the interspersion pattern invites the speculation that significant diminution of genomic size in insects may be accompanied by the appearance of a long period interspersion pattern.

Irrespective of this tentative correlation, it is clear when examining the evolutionary divergence of the class *Insecta*, that at the level of DNA sequence organization the relative arrangement of repetitive and single copy DNA sequences in insects has undergone a major reorganization during the course of evolution of these organisms.

Acknowledgments. We are indebted to Michael Berns and Scott Petersen for use of both equipment and expertise in the microfluorometric measurements of DNA content and to Francine Eden for her valuable suggestions. This research has been supported by grant GM 22207 from the National Institutes of Health.

References

Anderson, D.T. (1973). Embryology and phylogeny of annelids and anthropods. p. 464. New York: Pergamon Press

Bachmann, K. (1972). Chromosoma **37**, 85–93

Britten, R.J., Graham, D.E., Neufeld, B.R. (1974). In: Methods in enzymology, 29E, L. Grossman, K. Moldave, eds., pp. 363–418. New York: Academic Press

Cairns, J. (1963). Cold Spring Harb. Symp. Quant. Biol. **28**, 43–46

Chamberlin, M.E., Britten, R.J., Davidson, E.H. (1975). J. Mol. Biol. **96**, 317–333

Crain, W.R., Davidson, E.H., Britten, R.J. (1967A). Chromosoma **59**, 1–12

Crain, W.R., Eden, F.C., Pearson, W.R., Davidson, E.H., Britten, R.J. (1967B). Chromosoma **56**, 309–326
Davidson, E.H., Galau, G.A., Angerer, R.C., Britten, R.J. (1975). Chromosoma **51**, 253–259
Davidson, E.H., Graham, D.E., Neufeld, B.R., Chamberlin, M.E., Amenson, C.S., Hough, B.R., Britten, R.J. (1974). Cold Spring Harb. Symp. Quant. Biol. **38**, 295–301
Davidson, E.H., Hough, B.R., Chamberlin, M.E., Britten, R.J. (1973). J. Mol. Biol. **77**, 1–23
Eden, F.C., Hendrick, J.D. (1978). Biochemistry **17**, 5838–5844.
Efstratiadis, A., Crain, W.R., Britten, R.J., Davidson, E.H., and Kafatos, F.C. (1976). Proc. Natl. Acad. Sci. USA **73**, 2289–2293
Goldberg, R.B., Crain, W.R., Ruderman, J.V., Moore, G.P., Barnett, T.R., Higgins, R.C., Gelfand, R.A., Galau, G.A., Britten, R.J., Davidson, E.H. (1975). Chromosoma **51**, 225–251
Hinegardner, R. (1968). Am. Nat. **102**, 517–523
Hudspeth, M.E.S., Timberlake, W.D., Goldberg, R.B. (1977). Proc. Natl. Acad. Sci.USA **74**, 4332–4336
Imms, A.D. (1964). Textbook of emotomology. New York: Dutton
Kram, R., Botcham, M., Hearst, J.L. (1972). J. Mol. Biol. **64**, 103–109
Laird, C.C. (1971). Chromosoma **32**, 378–400
Laird, C., McCarthy, B.J. (1969). Genetics **63**, 865–879
Manning, J.E., Schmid, C.W., Davidson, N. (1975). Cell **4**, 141–158
Pearson, W., Davidson, E.H., Britten, R.J. (1977). Nucleic Acids **4**, 1727–1737
Rees, H., Hazarika, M.H. (1969). Chromosomes Today, Proc. 2nd Oxford Chromosome Conf., 1967, 158–171
Stebbins, G.L. (1966). Science **152**, 1463–1469
Studier, F.W. (1965). J. Mol. Biol. **11**, 373–390
Walbot, V., Dove, L.S. (1976). J. Mol. Biol. **101**, 503–536
Wells, R., Rover, H.D., Hollenberg, C.D. (1976). Mol. Gen. Genet. **147**, 45–51
Wetmur, J.G., Davidson, N. (1968). J. Mol. Biol. **31**, 349–361
Zimmerman, J.L., Goldberg, R.B. (1977). Chromosoma **59**, 227–252

Received October 22, 1979; revised March 12, 1980