Genet. Res., Camb. (1988), **51**, *pp.* 185–196 *With* 10 *text-figures Printed in Great Britain*

Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization

WILLIAM C. BLACK, IV AND KARAMJIT S. RAI

Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

(*Received 7 January 1987 and in revised form 5 October 1987*)

Summary

DNA reassociation kinetics were used to determine the amounts and organization of repetitive and unique DNA in four mosquito species: *Anopheles quadrimaculatus* (Say), *Culex pipiens* (L.), *Aedes albopictus* (Skuse) and *Ae. triseriatus* (Say). Intraspecific variation in repetitive DNA amounts was examined in two geographic strains of *Ae. albopictus* fom Calcutta, India and the island of Mauritius. Repetitive and unique sequences in *An. quadrimaculatus* were distributed in a pattern of long period interspersion. Repetitive DNA in all other mosquito species exhibited a pattern of short period interspersion. The amounts of fold-back, middle repetitive, and highly repetitive sequences increased with genome size. The amount of foldback DNA increased at a much slower rate than the middle and highly repetitive sequences. Intraspecific variation in genome size in *Ae. albopictus* was due primarily to the amounts of highly repetitive DNA. S1 nuclease digestion of repetitive DNA in all species revealed a positive correlation between genome size and the proportion of the repetitive DNA consisting of short repeats. The amounts of long and short repeats increased with genome size but short repeats increased at a higher rate. The repetitive DNA of the Mauritius strain contained approximately 15% more short repeats than the Calcutta strain. These findings suggest that genome evolution in mosquitoes has resulted from changes in both the amounts and organization of repetitive elements.

1. Introduction

Repetitive DNA exists in a bewildering array of sequences and sizes in the eukaryotic genome. The amounts and organization of repetitive elements vary widely and in a manner independent of taxonomic affinities (Bouchard, 1982). Yet the commonality of two basic organization patterns throughout all eukaryotic genomes suggests that certain principles may govern the way that repetitive elements become established and spread (Davidson *et al.* 1975 for review). The first type of genome organization has been termed 'short period interspersion' and indicates a pattern of single-copy sequences, 1000–2000 base pairs (bp) in length, alternating regularly with short (200–600 bp) and moderately long (1000–4000 bp) repetitive sequences. This organization is characteristic of the DNA in the majority of animal species. The second pattern of organization is termed 'long period interspersion' and describes a pattern of long ($\geqslant$ 5600 bp) repeats alternating with very long ($\geqslant$ 13 kbp) uninterrupted stretches of unique sequences. The long repeats consist of a mixture of long and short elements (Crain *et al.* 1976*a*; Epplen *et al.* 1978, 1979). Although long period interspersion is characteristic of most species with small genomes [0·10–0·50 picograms (pg)/haploid genome], the genomes of some species within this range exhibit short period interspersion (Samols & Swift, 1978). Alternatively most avian genomes are relatively large (1–2 pg) and yet exhibit patterns of long interspersion (Epplen *et al.* 1978, 1979).

The function of repetitive DNA in genome evolution has been the subject of much contention at the molecular (Dover & Flavell, 1982) and organismic (Cavalier-Smith, 1985) levels. Towards a better understanding of its role much research has been devoted to determining abundances and sequence homologies of specific repeats across large taxa (e.g. the *Alu* sequence in Mammalia (Weiner *et al.* 1986)). However this has led to no better understanding of how these sequences arise and spread in taxonomic families, genera or, most importantly, at the intraspecific level in populations. Intraspecific variation in DNA amounts is well documented in a number of species and yet to our knowledge no studies have been made of intraspecific variation in repetitive DNA amounts or organization. Such knowledge might well abate some of the confusion and mystique (Dawkins, 1976; Doolittle & Sapienza, 1980; Cavalier-Smith, 1980; Orgel & Crick,

1980; Orgel *et al.* 1980) surrounding this class of DNA.

We suggest that much can be learned by examining the abundance, organization and size of repetitive elements in a single taxonomic family. Towards this end our laboratory has been studying genome evolution in mosquitoes (Family Culicidae) (Rao, 1985; McLain *et al.* 1986, 1987). Mosquitoes are a family well suited for such a study. The haploid genome size of mosquitoes varies from 0·25 pg in *An. quadrimaculatus* to 1·92 pg in *Aedes zoosophus* (Dyar and Knab) (Rao, 1985). The genome of a single species, *Ae. albopictus*, varies from 0·87–1·32 pg among geographic strains. The purpose of the present study was to describe intraspecific and interspecific variation in the amounts and organization of repetitive and unique DNA. This variation was examined using DNA reassociation kinetics in four species with widely different genome sizes. In addition, intraspecific variation in the abundance of repetitive sequences was examined in two geographic strains of *Ae. albopictus* possessing the largest and smallest genomes. By relating overall genome size to the abundance and organization of repetitive elements we hoped to identify patterns in the way that repetitive DNA arises and spreads in a single taxonomic family.

2. Materials and methods

(i) *Origin and age of mosquito strains*

All stocks used in this study are currently maintained in the Vector Biology Laboratory at the University of Notre Dame. *An. quadrimaculatus* was received from H. Schoof, Technical Developments Laboratory, U.S. Public Health Service, Savannah, Georgia in 1971. *Cx. pipiens* was received from R. Lowe, USDA, Gainesville, FL, in 1969. *Ae. albopictus* from Calcutta, India was sent by M. Bhattacharya in 1973. The strain from Candos, Mauritius was provided by C. Courtois in 1970. *Ae. triseriatus* was collected from Izaac Walton Preserve, St Joseph, Co., Indiana by R. Beach in 1969.

(ii) *DNA isolation and purification*

Pupae and fourth instar larvae (15–20 g) were homogenized in 40 ml of cold (−20 °C) homogenization buffer (6·6 M guanidine hydrochloride, 100 mM potassium acetate, 142 mM 2-mercaptoethanol) with a Dounce homogenizer. The homogenate was centrifuged at 2000 *g* for 30 min to remove exoskeletons and crude cellular debris. DNA in the supernatant was precipitated with cold 95% ethanol (−20 °C), placed in the −20 °C freezer for 1 h and then centrifuged. The resulting pellet was resuspended in 8 ml of TE (10 mM Tris, 1 mM EDTA pH 8·0), 0·8 ml of 1 M-KCl, and 0·8 ml of 10% SDS and heated in a 65 °C water bath for 1 h to reduce nuclease activity. The temperature was dropped to 55 °C and 10 mg of Proteinase K (Sigma) was added. The mixture was incubated overnight (12–16 h). The solution was then extracted twice with buffer saturated phenol (Maniatis *et al.* 1982), twice with a 1:1 mix of phenol and chloroform/isoamyl alcohol (24:1) and twice with chloroform/isoamyl alcohol. At this stage the DNA solution had a murky grey appearance. To remove these contaminants (probably glycogen) the solution was diluted up to 29 ml in TE and 29·45 g of CsCl were added with 2 ml of ethidium bromide (10 mg/ml). The solution was sealed in a 39 ml Beckman polyallomer centrifuge tube and centrifuged in a Beckman VTi 50 rotor at 36 K for 40 h. The resulting band was removed to a sterile 50 ml tube and the ethidium bromide was extracted with isoamyl alcohol. DNA was ethanol precipitated after diluting the solution in 2 vol of TE to prevent precipitation of the CsCl. This procedure yielded approximately 1 mg of high-molecular-weight DNA (> 30 kb) for 4 g of pupae and larvae.

To determine the accuracy of our methods, measurements were made on sea urchin (*Strongylocentrus purpuratus* L.) and *E. coli* DNA. Sea urchin DNA was isolated from sperm according to the method of Britten *et al.* (1974). DNA was isolated from *E. coli* following the method of Rodriguez & Tait (1983).

(iii) *Preparation and sizing of DNA fragments*

DNA (1–2 mg) to be fragmented was dissolved in 4 ml of sonication buffer (Maniatis *et al.* 1982) in a 15 ml polypropylene tube. A mean fragment length of 2000 bp was obtained by cooling the sonication mixture to 5 °C and sonicating (Branson sonifier, Model S110) for 5 s at the lowest setting. A mean fragment length of 400 bp was obtained by cooling the sonication mixture to 0 °C and sonicating for twenty 10 s bursts at the highest setting. Average fragment length was determined using 0·75% agarose gel and *Hin* dIII restricted lambda fragments as standards.

(iv) *DNA reassociation*

Following sonication, DNA fragments were ethanol precipitated and resuspended in 0·12 M phosphate buffer pH 7·0 (PB) at an approximate concentration of 100 μg DNA/ml for low C_0t values and 500 μg/ml for mid C_0t values. DNA was dissolved at a concentration of 1 mg/ml in 0·4 M-PB for high C_0t values. When reassociating in 0·4 M-PB, equivalent C_0t values were calculated by multiplying by 4·9 (Britten *et al.* 1974). 40 μg of DNA were used for each point on the C_0t curve and approximately 10 points were collected for each of the three reassociation conditions. Reassociation therefore took place in a total volume of 4 ml for low C_0t, 0·8 ml for mid C_0t and 0·4 ml for high C_0t values. Following the method of Britten *et al.* (1974), DNA was disassociated in a

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 1. *SAS program used to calculate DNA reassociation kinetics*

```
DATA COT;
TITLE AEDES TRISERIATUS;
INPUT COT REASS;
REASS=REASS/100;
CARDS;
.
data
.
PROC NLIN METHOD=MARQUARDT;
PARM PFAST=0.30 PMID=0.30 PSLOW=0.30
     KFAST=10.0 KMID=1.00 KSLOW=0.001;
DENA=1+(KFAST*PFAST*COT);
DENB=1+(KMID*PMID*COT);
DENC=1+(KSLOW*PSLOW*COT);
FAST=PFAST/DENA;
MID=PMID/DENB;
SLOW=PSLOW/DENC;
MODEL REASS=FAST+MID+SLOW;
DER.PFAST=1/(DENA**2.0);
DER.PMID=1/(DENB**2.0);
DER.PSLOW=1/(DENC**2.0);
DER.KFAST=-COT*PFAST*PFAST/(DENA**2.0);
DER.KMID=-COT*PMID*PMID/(DENB**2.0);
DER.KSLOW=-COT*PSLOW*PSLOW/(DENC**2.0);
```

Abbreviations: COT, the C_0t value at which a reassociation determination was made. REASS, the proportion of DNA which reassociated at COT. PFAST, proportion of the genome comprised of highly repetitive sequences. PMID, proportion of the genome comprised of middle repetitive sequences. PSLOW, proportion of the genome comprised of unique sequences. KFAST, reassociation rate of highly repetitive sequences. KMID, reassociation rate of middle repetitive sequences. KSLOW, reassociation rate of unique sequences.

100 °C water bath for 10 min and then reassociated at 60 °C for low and mid C_0t or at 66 °C for high C_0t values. At various times over the course of reassociation, samples were removed and washed with 0·03 M-PB onto a water-jacketed, hydroxyapatite (Bio-rad) column maintained at 60 °C. Single-stranded DNA was then eluted with 60 °C 0·12 M-PB and double-stranded DNA with 60 °C 0·4 M-PB. The concentrations of single- and double-stranded DNA were measured spectrophotometrically at 260 and 320 nm to correct for background contamination and light scattering.

(v) *Determination of reassociation kinetics*

The proportion of the genome comprised of highly repetitive, middle repetitive and unique sequences as well as the reassociation rates of these three components were estimated with the nonlinear procedure (PROC NLIN) in the now widely available SAS language (SAS, 1988) (Table 1).

Genome size was estimated with the reassociation rate of the unique component using the relationship (Lewin, 1980):

$$\text{Genome size (pg)} = C_0t_{\frac{1}{2}}\text{ (unique)} \times (0{\cdot}0044\text{ pg}/2{\cdot}68\text{ mol s}), \quad (1)$$

where 0·0044 pg = the genome size of *E. coli* (Cairns, 1963) and 2·68 mol s = the $C_0t_{\frac{1}{2}}$ of *E. coli* DNA (Table 3).

The rate of reassociation of long fragments of unique DNA was predicted from the reassociation rate of short unique fragments with the relationship (Wetmur & Davidson, 1968):

$$\text{Rate long fragments} = \text{rate short fragments} \times \text{SQRT}(L2/L1), \quad (2)$$

where L1 = short fragment length and L2 = long fragment length.

The equation predicts the reassociation rate expected for long DNA fragments on the assumption that they consist entirely of unique sequences.

(vi) *Repeat length determination*

DNA was sonicated into 2000 bp fragments, dissociated as above and reassociated to a C_0t of 10 mol s in 0·3 M-NaCl, 0·01 M-PIPES pH 6·7 at 64 °C. An equivalent C_0t was calculated by multiplying by 2·3 (Britten *et al.* 1974). Once reassociated, an equal volume of 0·05 M sodium acetate, 0·2 mM zinc sulfate (pH 4·4), and 0·01 M 2-mercaptoethanol were added. S1 nuclease (Bethesda Research Labs) was added at a rate of 0·5 Units/μg of DNA and the mixture incubated at 37 °C for 45 min. The reaction was stopped by adding $\frac{1}{2}$ vol of 0·4 M-PB. Products

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 2. *Genome size and chemical properties of DNA in four mosquito species*

Species	Haploid genome size (pg)	Hyper-chromicity (% ± S.E.)	T_m (°C) (± S.E.)	Guanosine cytosine content (%)
An. quadrimaculatus	0·24	29·4 ± 1·75	88·10 ± 0·30	46·3
Cx. pipiens	0·54	26·7 ± 0·64	82·80 ± 0·21	33·3
Ae. albopictus				
Calcutta	0·86	25·7 ± 1·00	83·50 ± 0·10	35·0
Mauritius	1·32	25·5 ± 0·28	84·20 ± 0·25	36·7
Ae. triseriatus	1·52	24·9 ± 0·60	82·90 ± 0·40	33·6
Sea urchin	0·89	28·5 ± 0·45	83·10 ± 0·40	34·1

Groups	D.F.	Sum squares	Mean square	*F*
Analysis of variance of hyperchromicities among species				
Species	4	24·94	6·24	3·32
Anopheles vs. other spp.	(1)	21·83	21·83	11·62*
Remainder	(3)	3·11	1·04	0·55
Error	6	11·27	1·88	
Total	10	36·21		
Analysis of variance of melting temperatures (T_m) among species				
Species	4	40·64	10·16	67·36***
Anopheles vs. other sp.	(1)	36·91	36·91	244·71***
Ae. albo. vs. Ae. tris. and *Cx.*	(1)	2·07	2·07	13·75**
Calcutta *vs.* Mauritius	(1)	0·42	0·42	2·80
Remainder	(1)	1·23	1·23	8·18*
Error	6	0·91	0·15	
Total	10	41·55		

* $P \leqslant 0{\cdot}05$. ** $P \leqslant 0{\cdot}01$. *** $P \leqslant 0{\cdot}001$.

of digestion were separated from the S1-resistant repetitive duplex DNA with hydroxyapatite chromatography.

Agarose A-50 columns (1·2 × 115 cm) were poured around a support of 6 mm siliconized glass beads according to the method of Britten *et al.* (1974). Resistant duplexes were chromatographed in 0·12 M-PB. Consistent results were obtained by pouring a fresh column for each run. The size of eluted fragments was determined by concentrating and running them on a 0·75% agarose gel against *Hin* dIII-restricted lambda DNA.

(vii) *Optical monitoring of DNA melting profiles*

To estimate hyperchromicities and melting temperatures, DNA samples were thermally denatured in 0·12 M-PB in electrically heated cuvettes. Samples were sealed against evaporation with mineral oil. The hyperchromic shift was monitored in a Bausch and Lomb 601 spectrophotometer. Hyperchromicity was calculated using the formula:

$$H = (A(260\text{–}320)\ \max - A(260\text{–}320)\ 60\ °C)/A(260\text{–}320)\ \max. \quad (3)$$

The melting temperature, T_m, was the point at which 50% of the hyperchromic shift had occurred. The guanosine-cytosine content of the DNA was estimated from T_m, with the equation (Mandel & Marmur, 1968):

$$\%\,(G\text{-}C) = 2{\cdot}44(T_m - 69{\cdot}14). \quad (4)$$

3. Results

Haploid genome size estimated from Feulgen cytophotometry (Rao, 1985), hyperchromicity, melting temperature and base composition of DNA in each

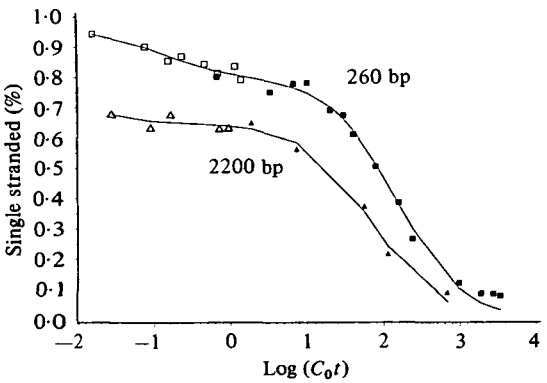

Fig. 1. Reassociation of *Anopheles quadrimaculatus* DNA. Both curves represent least squares solutions with two components. The upper curve represents reassociation of 260 bp fragments, the lower curve resulted from reassociation of 2200 bp fragments. Solid squares and triangles indicate reassociations carried out at 66 °C.

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

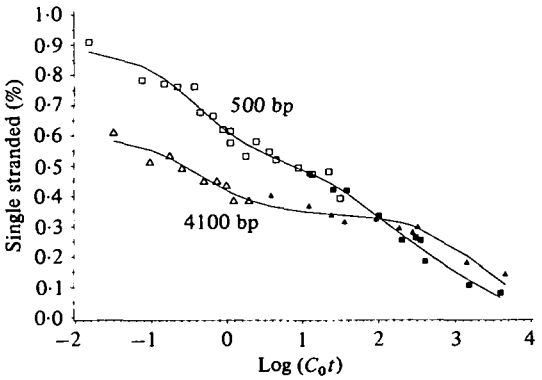

Fig. 2. Reassociation of *Culex pipiens* DNA. The upper curve represents a least squares solution with three components, the lower curve was fitted to a two component model. The upper curve represents reassociation of 500 bp fragments, the lower curve resulted from reassociation of 4100 bp fragments. Solid squares and triangles indicate reassociations carried out at 66 °C.

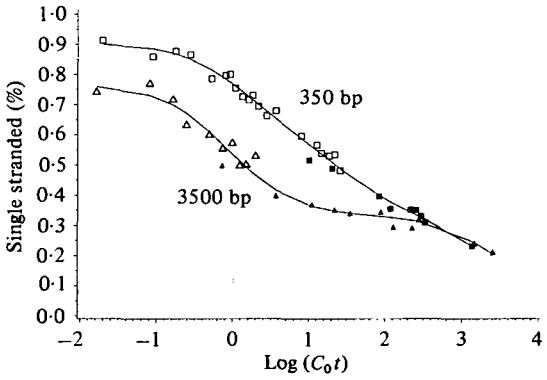

Fig. 3. Reassociation of *Aedes albopictus* (Mauritius strain) DNA. The upper curve represents a least squares solution with three components, the lower curve was fitted to a two component model. The upper curve represents reassociation of 350 bp fragments, the lower curve resulted from reassociation of 3500 bp fragments. Solid squares and triangles indicate reassociations carried out at 66 °C.

of the four mosquito species are listed in Table 2. Standard errors were calculated from at least two replicate measurements. Determinations were made on sea urchin DNA to determine the accuracy of our method. The T_m of sea urchin DNA is close to the published value (83·5 °C) (Graham *et al.* 1974) and the hyperchromicity is only slightly greater than the 27·8% reported. The analysis of variance indicates that the magnitudes of the hyperchromic shifts were homogeneous among all mosquito species except *Anopheles*. The significantly greater hyperchromicity and T_m in this species is probably a consequence of its higher G-C content. The T_m of the two *Ae. albopictus* strains were statistically homogeneous but significantly greater than those of *Cx. pipiens* and *Ae. triseriatus*.

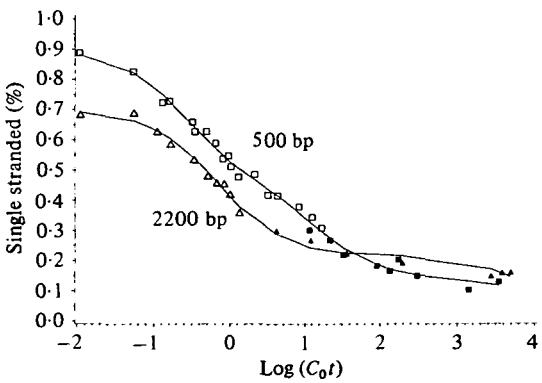

Fig. 4. Reassociation of *Aedes triseriatus* DNA. The upper curve represents a least-squares solution with three components, the lower curve was fitted to a two-component model. The upper curve represents reassociation of 500 bp fragments, the lower curve resulted from reassociation of 2200 bp fragments. Solid squares and triangles indicate reassociations carried out at 66 °C.

(i) *Reassociation kinetics of mosquito DNA*

Reassociation data and the least squares fit for short and long DNA fragments are plotted in Figs 1–4. Reassociation kinetics appear in Table 3. The reassociation rate of *E. coli* DNA was determined to test the accuracy of our reassociation method. The rate agrees well with published estimates (Lewin, 1980). Initially an attempt was made to obtain replicate reassociation curves. However, the sonication procedure yielded fragments which varied widely around a mean of 400 bp, making replication of entire curves impossible. As an alternative, C_0t values from each set of reassociations (low, middle and high) were overlapped to determine the amount of variation for replicate C_0t values for each batch of sonicated DNA. Points collected under different reassociation conditions were congruous (Figs 1–4).

In all species but *Anopheles* the best fit was obtained with a 3-component model with highly repetitive, middle repetitive and unique sequences. The *Anopheles* model was best fitted with a two-component model with highly repetitive and unique sequences. Approximately 20% of the *Anopheles* genome consisted of repetitive DNA with a small proportion (app. 4%) of the DNA reassociating before a C_0t of 0·01 mol s. This DNA consists largely of fragments which reassociate quickly by folding back on themselves (Lewin, 1980). As expected for a species with a small genome, a large proportion (approx. 80%) of the *Anopheles* genome consisted of unique sequences. Surprisingly, the *Culex* genome was found to consist of approximately 20% unique sequences even though it contained only twice as much DNA as the *Anopheles* genome.

In the Calcutta strain of *Ae. albopictus* 36% of the DNA was found to be unique as compared with 34% in the Mauritius strain. Converted to absolute

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 3. *DNA renaturation kinetic analysis of 4 mosquito species*

Species	Proportion of fragments (%)	Reassociation rate (1/mol s)	$C_0t_{\frac{1}{2}}$ mix (mol s)	$C_0t_{\frac{1}{2}}$ pure (mol s)
E. coli				
Fragments unreassociated = 8·03%				
Fragment size = 500 bp [Range = 270–900 bp]				
Unique	100·00	0·37306	2·681	2·681
An. quadrimaculatus				
Short fragments				
Fragments unreassociated = 7·85%				
Fragment size = 260 bp [Range = 170–520 bp]				
Foldback	3·83	—	—	—
Highly repetitive	16·40	49·53	0·020	0·003
Unique	79·77	0·00885	113·043	90·177
Long fragments				
Fragments unreassociated = 9·01%				
Fragment size = 2200 bp [Range = 1090–4680 bp]				
Immediate	32·97	—	—	—
Fast	6·82	16·02	0·062	0·004
Slow	60·21	0·02114	47·315	28·488
Cx. pipiens				
Short fragments				
Fragments unreassociated = 8·03%				
Fragment size = 500 bp [Range = 400–700 bp]				
Foldback	10·96	—	—	—
Highly repetitive	38·28	6·41	0·156	0·060
Middle repetitive	29·05	0·046	21·525	6·253
Unique	21·72	0·00312	320·737	69·659
Long fragments				
Fragments unreassociated = 14·45%				
Fragment size = 4100 bp [Range = 1900–7550 bp]				
Immediate	40·02	—	—	—
Fast	26·13	8·75	0·114	0·030
Slow	33·85	0·00140	714·984	242·030
Ae. albopictus – Calcutta				
Short fragments				
Fragments unreassociated = 11·68%				
Fragment size = 290 bp [Range = 220–420 bp]				
Foldback	9·96	—	—	—
Highly repetitive	17·09	91·16	0·011	0·002
Middle repetitive	36·85	0·270	3·697	1·362
Unique	36·09	0·00183	547·421	197·577
Long fragments				
Fragments unreassociated = 17·88%				
Fragment size = 920 bp [Range = 500–1700 bp]				
Immediate	27·68	—	—	—
Fast	43·61	0·879	1·137	0·496
Slow	28·71	0·00057	1754·386	503·707

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 3. (*cont.*)

Species	Proportion of fragments (%)	Reassociation rate (1/mol s)	$C_0t_{\frac{1}{2}}$ mix (mol s)	$C_0t_{\frac{1}{2}}$ pure (mol s)
Ae. albopictus – Mauritius				
Short fragments				
Fragments unreassociated = 23·12%				
Fragment size = 350 bp [Range = 170–580 bp]				
Foldback	9·55	—	—	—
Highly repetitive	30·56	2·33	0·430	0·131
Middle repetitive	26·91	0·122	8·191	2·205
Unique	32·98	0·00101	990·122	326·495
Long fragments – 3500 bp				
Fragments unreassociated = 21·20%				
Fragment size = 3500 bp [Range = 1780–5790 bp]				
Immediate	23·25	—	—	—
Fast	43·64	2·62	0·382	0·167
Slow	33·11	0·00073	1378·981	456·523
Ae. triseriatus				
Short fragments				
Fragments unreassociated = 11·54%				
Fragment size = 500 bp [Range = 380–800 bp]				
Foldback	9·88	—	—	—
Highly repetitive	43·76	9·12	0·110	0·048
Middle repetitive	30·66	0·258	3·877	1·189
Unique	15·70	0·00077	1291·803	202·781
Long fragments				
Fragments unreassociated = 15·57%				
Fragment size = 2200 bp [Range = 960–5300 bp]				
Immediate	30·07	—	—	—
Fast	48·26	3·06	0·327	0·158
Slow	21·67	0·00046	2173·818	471·003

amounts (Table 4) there was approximately 1·4 times more unique DNA in the Mauritius strain. Differences in the amounts of highly repetitive DNA accounted for most of the intraspecific variation in genome size. There were approximately 1·5 times more foldback sequences and 2·7 times more highly repetitive sequences in the Mauritius strain. There were approximately equal amounts of middle repetitive DNA in the two strains. The genome of the congeneric species, *Ae. triseriatus*, contained a large proportion (85%) of repetitive DNA and most of this was highly repetitive.

In most cases the number of copies of repetitive elements in mosquitoes (Table 4) followed patterns consistent with other animal species (Lewin, 1980). Middle repetitive elements existed as 10–300 copies in the genome whereas highly repetitive elements occurred in copy numbers of 5000–15000. However in the Calcutta strain of *Ae. albopictus* highly repetitive DNA consisted of sequences copied approximately 50000 times. This suggests that this strain has a small amount of highly repetitive DNA which consists of very few types of repeats. Intraspecific variation thus exists not only in amounts of repetitive DNA but sequence complexity as well.

In *Anopheles*, *Culex*, and the Calcutta strain of *Ae. albopictus* the kinetically determined genome sizes [equation (1)] were 0·186, 0·527 and 0·899 pg respectively. These were all close to the estimates obtained by the Feulgen spectrophotometric method (Rao, 1985). The kinetic estimate in the Mauritius strain was 1·63 pg which was 1·2 times as large as the value (1·32 pg) estimated by the Feulgen method. The kinetic estimate for *Ae. triseriatus* (2·12 pg) was 1·4 times as large as the Feulgen estimate (1·52 pg). These discrepancies are consistent with differences found using the two methods in other species (Lewin, 1980).

The amounts of foldback, highly repetitive and middle repetitive DNA (Table 4) are plotted by species in Fig. 5. In general, the amounts of these three repetitive classes increase with genome size. The

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 4. *Amounts and numbers of repeats of repetitive and unique sequences in each of four mosquito species*

Species	Amount (pg)	Complexity (bp × 10^{-6})[a]	Repetitive frequency
An. quadrimaculatus			
Foldback	0·009	—	—
Highly repetitive	0·040	0·032	5599
Unique	0·195	177·158 (= 0·186 pg)	1
Cx. pipiens			
Foldback	0·059	—	—
Highly repetitive	0·207	0·244	2057
Middle repetitive	0·157	33·733	15
Unique	0·117	502·468 (= 0·527 pg)	1
Ae. albopictus – Calcutta			
Foldback	0·086	—	—
Highly repetitive	0·148	0·017	49902
Middle repetitive	0·319	5·794	148
Unique	0·312	857·898 (= 0·899 pg)	1
Ae. albopictus – Mauritius			
Foldback	0·126	—	—
Highly repetitive	0·404	0·674	2303
Middle repetitive	0·356	12·837	121
Unique	0·436	1551·683 (= 1·63 pg)	1
Ae. triseriatus			
Foldback	0·150	—	—
Highly repetitive	0·665	0·172	11782
Middle repetitive	0·466	6·075	333
Unique	0·239	2024·467 (= 2·12 pg)	1

[a] The complexity in base pairs of each sequence class is calculated based on a genome size in *E. coli* of 4·2 × 10^6 bp which reassociates with a $C_0t_{\frac{1}{2}}$ of 2·68 mol s (Table 3).

amounts of foldback sequence generally increased at a much slower rate than the middle repetitive and highly repetitive sequences.

(ii) *Genome organization in mosquitoes*

The reassociation rates of long and short DNA fragments were compared to ascertain the pattern of interspersion of repetitive and unique sequences (Table 3). The predicted long-fragment reassociation rate [equation (2)] in *Anopheles* was 0·026/mol s which was very close to the observed rate of 0·021/mol s. This suggests that long fragments consisted entirely of single copy sequences uninterrupted by blocks of repetitive DNA. This further suggests that the repetitive and unique fragments were at least 2200 bp long. Such a pattern is consistent with the hypothesis that DNA sequences in the *Anopheles* genome follow a pattern of long period interspersion.

The predicted long fragment reassociation rate [equation (2)] in *Culex* was 0·0089/mol s corresponding with a $C_0t_{\frac{1}{2}}$ of 112 mol s. No reassociation (Fig. 2) occurred in this region of the curve and 70% of reassociation had occurred prior to it. This suggests that most long fragments contained one or more blocks of repetitive DNA such that they reassociated at a faster than predicted rate. This trend is consistent with the hypothesis that the majority of DNA sequences in *Culex* followed a short period interspersion pattern.

A similar pattern was observed in *Aedes* spp. In *Ae. albopictus* the predicted long-fragment reassociation

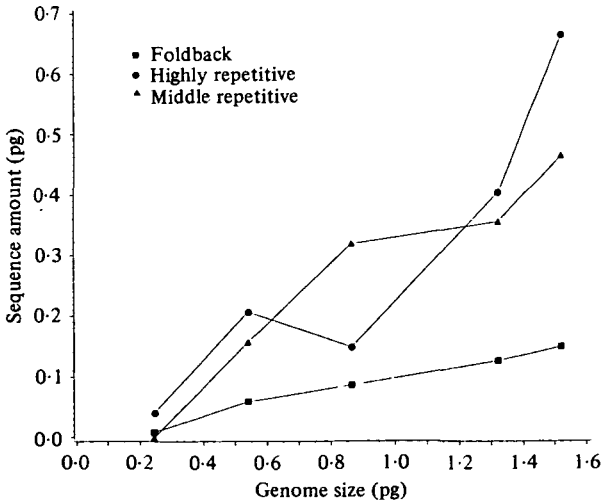

Fig. 5. The amounts of foldback, middle repetitive and highly repetitive DNA detected in each species plotted against genome size in that species.

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

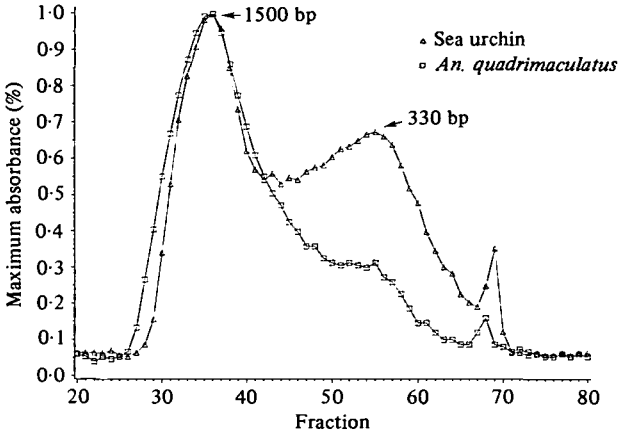

Fig. 6. Agarose A-50 fractionation of 2000 bp long sea urchin and *An. quadrimaculatus* DNA fragments reassociated to C_0t 10 mol s. Each fraction corresponds to 1 ml. The exclusion peak begins at fraction 28. The inclusion peak occurs at fraction 68.

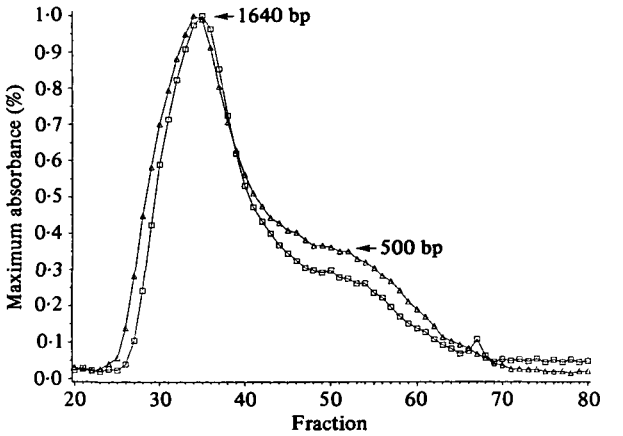

Fig. 7. Agarose A-50 fractionation of 2000 bp *Cx. pipiens* DNA fragments reassociated to C_0t 10 mol s. Each fraction corresponds to 1 ml. The exclusion peak begins at fraction 25. The inclusion peak occurs at fraction 67.

rate was 0·0033/mol s for both strains, which corresponded with a $C_0t_{\frac{1}{2}}$ of 300 mol s. Almost all reassociation occurred prior to the predicted $C_0t_{\frac{1}{2}}$ (Fig. 3). All of long-fragment reassociation with *Ae. triseriatus* DNA took place prior to the predicted $C_0t_{\frac{1}{2}}$ of 617 mol s (predicted rate = 0·001 62/mol s) (Fig. 4). These observations are consistent with the hypothesis that DNA sequences in culicine and aedine mosquitoes exhibit a pattern of short period interspersion.

In *Culex* and *Aedes* spp. the reassociation rate of the unique component of the short fragments was faster than the rate of reassociation of slow components among the long fragments. Yet long unique sequences should in theory [equation (2)] reassociate faster than short unique sequences. However an assumption of equation (2) is that short and long fragments are narrowly distributed around lengths L1 and L2, respectively. In practice the range of fragment sizes obtained when sonicating native DNA into large fragments is larger than the range observed when sonicating DNA into small fragments (Table 3). The slowly reassociating components of long-fragment curves in Figs 2, 3 and 4 parallel the reassociation rate of the unique component of the short fragment curve. Therefore, the hybridization seen late in the reassociation of long fragments probably represented the reannealing of short fragments of unique sequences.

(iii) *S1-nuclease digestion to determine the lengths of repetitive elements in mosquitoes*

To characterize repetitive elements further, short and long repetitive elements were generated and isolated by reassociating long fragments to an intermediate C_0t of 10 mol s, digesting them with S1-nuclease to remove single stranded (unreassociated) sections and chromatographing them on an agarose column. We initially ran this procedure on sea urchin DNA to ensure the accuracy of our methods. The size profile of fragments obtained for sea urchin and *Anopheles* DNAs appear in Fig. 6. The profile obtained for sea urchin is very similiar to that obtained by Britten *et al.* (1976) and contrasts sharply with that obtained for *Anopheles*. Approximately half of the repetitive DNA in sea urchins is 1000 bp or longer, while the remainder digests into shorter fragments with a modal size of 330 bp. In contrast, 80% of repetitive DNA in *Anopheles* is 1000 bp or longer and only 20% is shorter than 1000 bp. The profile is very similar to that obtained with *Drosophila melanogaster* (Crain *et al.* 1976*a*) and further supports the hypothesis that DNA in *An. quadrimaculatus* follows a pattern of long period interspersion.

Two independent digestion profiles were made with *Culex* DNA to determine the reproducibility of our method (Fig. 7). The two digestion profiles were identical for large fragments and only a 3% difference exists among short fragments in replicate runs. Approximately 70% of repetitive DNA was $\geqslant$ 1000 bp in length. The small repeats generated by this procedure were broadly distributed in size and no modal size was detected.

Long fragments were identically distributed in both strains of *Ae. albopictus* (Fig. 8). But the two strains exhibit striking differences in the abundance and distribution of short repeats. Approximately half of the repetitive DNA in the Mauritius strain consisted of small fragments as contrasted with 34% in the Calcutta strain. A modal distribution of short fragments was detected in the Mauritius strain with a mean of approximately 330 bp fragments and a range of 250–600 bp. No mode was observed among short repeats in the Calcutta strain. These observations are consistent with a hypothesis that differences in repetitive DNA amounts in the two geographic strains were generated by the loss or gain of short repetitive elements. The S1 digestion profile of *Ae. triseriatus* repetitive DNA (Fig. 9) indicated that approximately

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

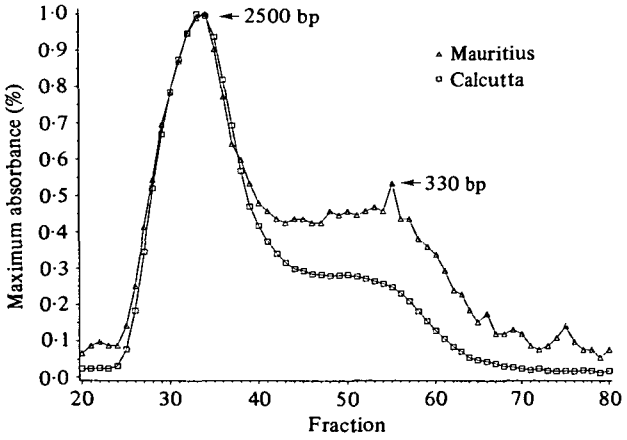

Fig. 8. Agarose A-50 fractionation of 2000 bp *Ae. albopictus* DNA fragments reassociated to C_0t 10 mol s. Each fraction corresponds to 1 ml. The exclusion peak begins at fraction 26. The inclusion peak occurs at fraction 66.

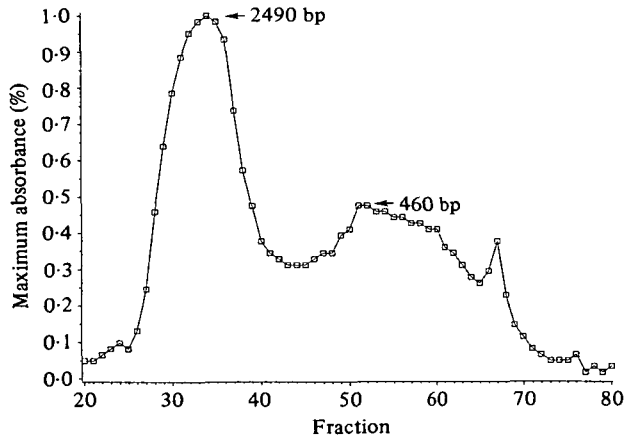

Fig. 9. Agarose A-50 fractionation of 2000 bp *Ae. triseriatus* DNA fragments reassociated to C_0t 10 mol s. Each fraction corresponds to 1 ml. The exclusion peak begins at fraction 26. The inclusion peak occurs at fraction 68.

half was ⩾ 1000 bp in size while the remainder was broadly distributed from 200–600 bp with a mode of approximately 460 bp.

The proportion of DNA consisting of long (⩾ 1000 bp) and short (< 1000 bp) repeats is plotted by species in Fig. 10. The proportion of repetitive DNA consisting of short repeats grew with increasing genome size. This suggests that short repeats increased at a more rapid rate than the long repeats.

(iv) *Melting profiles of long and short repeats*

The thermal characteristics of long, intermediate and short repeats are reported by species in Table 5. In *Anopheles* the hyperchromicities of repetitive fragments decreased as fragments became smaller. The melting temperatures of long and short repeats were significantly lower than native DNA. These reductions probably represent base pair mismatching in reassociated fragments and suggest that sequence homology may be greater in longer repeat families.

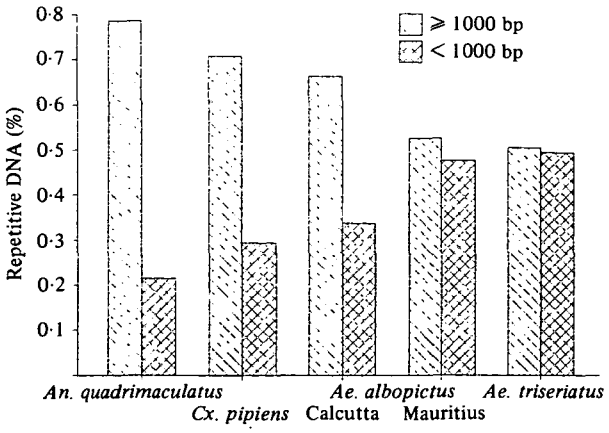

Fig. 10. The proportion of repetitive DNA consisting of long (⩾ 1000 bp) and short (< 1000 bp) repeats plotted by species.

In *Culex* the hyperchromicities of repetitive fragments in the different fractions were approximately equal to one another and were equal or only slightly less than the hyperchromicity of native DNA (Table 1). The melting temperatures decreased significantly from the largest to the smallest fractions, suggesting that the amount of base pair mismatch increased with decreasing fragment size. What is surprising is that melting temperatures of the long repetitive fragments were significantly higher than the native DNA. This suggests that the long repetitive DNA in *Culex* may be richer in G-C content than the native DNA.

In both strains of *Ae. albopictus*, melting temperatures significantly decreased with fragment size, suggesting that the amount of base pair mismatch was greater in smaller fragments. As in *Culex* the T_m of the longest repetitive fragments was greater than or equal to those in the native DNA. The T_m of Mauritius long repeats was equal to the melting temperature of the native DNA while long repeats from the Calcutta strain melted at a significantly higher temperature than native DNA. In contrast with *Culex*, a T_m higher than native DNA was only observed in the largest fragments (⩾ 2650 bp) and the temperature differential was less. This suggests that the long repeats in *Ae. albopictus* may be richer in G-C content than the native DNA. In *Ae. triseriatus* DNA, the T_m of all repetitive fractions was significantly less than the native DNA and decreased with fragment size, suggesting as with the other species that the amount of base pair mismatch increased in smaller fragments.

4. Discussion

Our results indicate that the genomic organization of *An. quadrimaculatus* is of the 'long period interspersion' type and is in most respects similar to that of *D. melanogaster* and its congeneric species (Laird & McCarthy, 1969). Genome size in anopheline mosqui-

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

Table 5. *Thermal stability of S1-resistant fragments in four mosquito species*

Species	Size (bp)	Hyper-chromicity	T_m (°C)	Corrected T_m[a]	95% C.I.	% bpm[b]
An. quadrimaculatus						
Excluded peak	1620	26·7	84·8	85·2	(84·8, 85·6)	2·9
Valley	740	23·9	84·6	85·5	(84·6, 86·3)	2·6
Included peak	360	21·9	85·3	87·1	(85·8, 88·4)	1·0
Cx. pipiens						
Excluded peak	1640	26·7	85·3	85·7	(85·2, 86·2)	−2·9
Valley	942	27·8	83·8	84·5	(83·7, 85·2)	−1·7
Included peak	580	25·6	78·7	79·8	(79·1, 80·6)	3·0
Ae. albopictus – Calcutta						
Excluded peak	2650	27·9	84·9	85·1	(84·7, 85·6)	−1·6
Valley	1020	26·4	82·7	83·3	(82·7, 84·0)	0·2
Included peak	500	24·9	77·4	78·7	(78·1, 79·3)	4·8
Ae. albopictus – Mauritius						
Excluded peak	2510	27·3	84·1	84·4	(83·9, 84·8)	−0·2
Valley	940	27·3	82·3	83·0	(83·0, 83·0)	1·2
Included peak	460	23·7	76·7	78·1	(76·8, 79·4)	6·1
Ae. triseriatus						
Excluded peak	2490	25·4	81·2	81·5	(80·5, 82·4)	1·4
Valley	940	33·3	78·8	79·5	(77·9, 81·1)	3·4
Included peak	460	16·7	73·8	75·2	(73·1, 77·3)	7·7

[a] The melting temperature was corrected for the effect of fragment size by adding 650/fragment size to the measured temperature (Wetmur & Davidson, 1968).
[b] Britten *et al.* (1974) estimated 1% base pair mismatch (% bpm) per 1 °C reduction in T_m. % bpm = T_m (native) − T_m (reassociated DNA). C.I. = confidence interval.

toes varies from 0·25 to 0·34 pg (Jost & Mameli, 1972). It will be interesting to determine whether the same pattern of genomic organization exists in other species of *Anopheles*. Other insects found to have this type of organization are the honeybee (*Apis mellifera* (L.)) Crain *et al.* (1976*b*), a flesh fly (*Sarcophaga bullata* (Parker)) Samols & Swift (1979), and a midge (*Chironomus tentans* (Fabricios)) Wells *et al.* (1976). *Culex pipiens* (0·545 pg) and *Sarcophaga bullata* (0·590 pg) have very nearly the same genome size but different genome organizations. This indicates that, as with other animal taxa, the presence of long period interspersion is not necessarily associated with small genome size in insects.

Other members of the family Culicidae examined in this study exhibited 'short period interspersion' genomic organization. Other insects found to have this type of organization are the house fly (*Musca domestica* (L.)) Crain *et al.* (1976*b*), and two lepidopteran species *Bombyx mori* (L.) (Gage, 1974) and *Antherea pernyi* (Guérin-Méneville) (Efstratiadis *et al.* 1976). To our knowledge the family Culicidae is the only one so far examined which contains species exhibiting both types of sequence organization. This is exciting because it presents the possibility that further investigation of the repetitive elements in the various genera may indicate how the transition from long to short interspersion (or *vice versa*) is made.

The amounts of fold-back, highly repetitive and middle repetitive DNA increased linearly with genome size. The major exception was the genome of *Cx. pipiens* which consisted of approximately the same proportion of repetitive sequences (78%) as that of *Ae. triseriatus* (84%) even though the *Culex* genome is one-third the size. The amounts of foldback sequence generally increased at a much slower rate than the middle repetitive and highly repetitive sequences possibly suggesting different multiplicative mechanisms.

Intraspecific variation in genome size in *Ae. albopictus* was due primarily to the amounts of highly repetitive DNA. The S1-digestion further revealed that repetitive DNA in the Mauritius strain contained approximately 15% more short repeats. This finding is important in light of the recent observation that many if not all short repeat families in mammals are retropseudogenes derived from known RNA polymerase transcripts (Weiner *et al.* 1986). This suggests that intraspecific variation in genome size may be generated by retrotransposition events.

Among all species there was a positive relationship between genome size and the proportion of the repetitive DNA consisting of short repeats (Fig. 10). This is consistent with the hypothesis that short repeats increased at a more rapid rate than long repeats. This also supports the mechanism of retrotransposition as a means for generating and multiplying repetitive elements.

https://doi.org/10.1017/S0016672300024289 Published online by Cambridge University Press

The most important conclusion of our findings is that enormous variation exists in the amounts and organization of repetitive DNA at the family level. This highlights the need for more research into how repetitive elements arise and proliferate within families. We suggest that future investigations also aim at determining the amount of intraspecific variation in repetitive element amounts, size, sequence and organization.

We thank Drs J. A. Ferrari, D. K. McLain and P. N. Rao for helpful comments and discussion. Thanks to Dr N. M. DuTeau for advice and assistance. Thanks to Dr M. Fennewald for use of equipment. We thank Brian Turco for raising mosquitoes. This work was supported by NIH training grant 5T32 AI 07030 and NIH grant 5R01 AI 21443 to K.S.R.

References

Britten, R. J., Graham, D. E. & Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation. In *Methods in Enzymology*, vol. 29, part E (ed. L. Grossman and K. Moldave), pp. 363–418. New York: Academic Press.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. & Davidson, E. H. (1976). Evolutionary divergence and length of repetitive sequences in sea urchin DNA. *Journal of Molecular Evolution* **9**, 1–23.

Bouchard, R. A. (1982). Moderately repetitive DNA in evolution. In *International Review of Cytology*, vol. 76 (ed. G. H. Bourne and J. F. Danielli), pp. 113–193. New York: Academic Press.

Cairns, J. (1963). The chromosome of *E. coli*. *Cold Spring Harbor Symposium on Quantitative Biology* **28**, 43–46.

Cavalier-Smith, T. (1980). How selfish is DNA? *Nature* (*Lond.*) **285**, 617–618.

Cavalier-Smith, T. (1985). Eukaryotic gene numbers, non-coding DNA, and genome size. In *The Evolution of Genome Size* (ed. T. Cavalier-Smith), pp. 69–103. Chichester: John Wiley & Sons.

Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H. & Britten, R. J. (1976*a*). Absence of short period interspersion of repetitive and non-repetitive sequences in the DNA of *Drosophila melanogaster*. *Chromosoma* (*Berl.*) **56**, 309–326.

Crain, W. R., Davidson, E. H. & Britten, R. J. (1976*b*). Contrasting patterns of DNA sequence arrangement in *Apis mellifera* (Honeybee) and *Musca domestica* (Housefly). *Chromosoma* (*Berl.*) **59**, 1–12.

Dawkins, R. (1976). *The Selfish Gene*. London: Oxford University Press.

Davidson, E. H., Galua, G. A., Angerer, R. C. & Britten, R. J. (1975). Comparative aspects of DNA organization in metazoa. *Chromosoma* (*Berl.*) **51**, 253–259.

Doolittle, W. F. & Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* (*Lond.*) **284**, 617–618.

Dover, G. A. & Flavell, R. B. (1982). *Genome Evolution*. London: Academic Press.

Efstratiadis, A., Crain, W. R., Britten, R. J., Davidson, E. H. & Kafatos, F. C. (1976). DNA sequence organization in the lepidopteran *Antheraea pernyi*. *Proceedings National Academy Science* (*USA*) **73**, 2289–2293.

Epplen, J. T., Diedrich, U., Wagenmann, M., Schmidtke, J. & Engel, W. (1979). Contrasting DNA sequence organization patterns in Sauropsidian genomes. *Chromosoma* (*Berl.*) **75**, 199–214.

Epplen, J. T., Leipoldt, M., Engel, W. & Schmidtke, J. (1978). DNA sequence organisation in avian genomes. *Chromosoma* (*Berl.*) **69**, 307–321.

Gage, L. P. (1974). The *Bombyx mori* genome: analysis by DNA reassociation kinetics. *Chromosoma* (*Berl.*) **45**, 27–42.

Graham, D. E., Neufeld, B. E., Davidson, E. H. & Britten, R. J. (1974). Interspersion of repetitive and nonrepetitive DNA sequences in the sea urchin genome. *Cell* **1**, 127–137.

Jost, E. & Mameli, M. (1972). DNA content in nine species of Nematocera with special reference to the sibling species of the *Anopheles maculipennis* group and the *Culex pipiens* group. *Chromosoma* (*Berl.*) **37**, 201–208.

Laird, C. D. & McCarthy, B. J. (1969). Molecular characterization of the *Drosophila* genome. *Genetics* **63**, 865–882.

Lewin, B. (1980). *Gene Expression, vol. 2: Eukaryotic Chromosomes*, 2nd edn. New York: John Wiley & Sons.

Mandel, M. & Marmur, J. (1968). Use of ultraviolet absorbance-temperature profile for determining the guanine plus cytosine content of DNA. In *Methods in Enzymology*, vol. 12, part B (ed. L. Grossman and K. Moldave), pp. 195–206. New York: Academic Press.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory. 545 pp.

McLain, D. K., Rai, K. S. & Fraser, M. J. (1986). Interspecific variation in the abundance of highly repeated DNA sequences in the *Aedes scutellaris* (Diptera: Culicidae) subgroup. *Annals of the Entomological Society of America* **79**, 784–791.

McLain, D. K., Rai, K. S. & Fraser, M. J. (1987). Intraspecific and interspecific variation in the sequence and abundance of highly repeated DNA among mosquitoes of the *Aedes albopictus* subgroup. *Heredity* **58**, 373–381.

Orgel, L. E. & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* (*Lond.*) **284**, 604–607.

Orgel, L. E., Crick, F. H. C. & Sapienza, C. (1980). Selfish DNA. *Nature* (*Lond.*) **288**, 645–646.

Rao, P. N. (1985). Nuclear DNA and chromosomal evolution in mosquitoes. Ph.D. dissertation, University of Notre Dame, Notre Dame, Indiana.

Rodriguez, R. L. & Tait, R. C. (1983). *Recombinant DNA Techniques: An Introduction*. London: Addison-Wesley Publ. Co.

Samols, D. & Swift, H. (1979). Genomic organization in the flesh fly *Sarcophaga bullata*. *Chromosoma* (*Berl.*) **75**, 129–143.

SAS (1988). Statistical Analysis System User's Guide: Statistics. SAS Institute Inc., Cary, North Carolina.

Weiner, A. M., Deiniger, P. L. & Efstratiadis, A. (1986). Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annual Review of Biochemistry* **55**, 631–661.

Wells, R., Royer, H. & Hollenberger, C. P. (1976). Non *Xenopus*-like DNA organization in the *Chironomus tentans* genome. *Molecular and General Genetics* **147**, 45–51.

Wetmur, J. G. & Davidson, N. (1968). Kinetics of renaturation of DNA. *Journal Molecular Biology* **31**, 349–370.