Genet. Res., Camb (1991), **58**, *pp.* 225–232 *With 5 text-figures Printed in Great Britain*

The *Aedes aegypti* genome: complexity and organization

ANN M. WARREN AND JULIAN M. CRAMPTON*
Wolfson Unit of Molecular Genetics, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

(*Received 17 December 1990 and in revised form 14 June 1991*)

Summary

We describe the use of DNA reassociation kinetics to determine the total genome size and complexity together with the individual complexity and copy number of the single copy, middle repetitive and highly repeated DNA fractions of cell line and larval DNA from the mosquito, *Aedes aegypti*. The genome of *Ae. aegypti* is both large and complex, being one third the size of the human genome, and exhibits a short period interspersed repeat pattern. The implications of patterns of sequence arrangement and genome complexities for experiments aimed at isolating specific classes of DNA sequences, such as mobile genetic elements, are discussed.

1. Introduction

The genomes of all higher eukaryotes contain highly repetitive (approximately 10^5–10^6 copies per haploid genome), moderately repetitive (approximately 10^2–10^4 copies per haploid genome) and non-repetitive (one copy per haploid genome) DNA sequences (Jelinek & Schmid, 1982). Highly repetitive (HR) DNA consists of a few families of relatively short sequences which generally reside at centromeric and telomeric positions. Although they are thought to be important to the structure of chromosomes, their functions are unknown. Middle repetitive (MR) DNA consists of many families of both long (approx. 5000 bp) and short (approx. 300 bp) sequences that are interspersed with unique sequences. It is in this fraction of the genome that most transposable genetic elements (TGEs) reside, particularly the autonomous retrotransposons (RTPs) and some non-viral retroposons such as the *Drosophila* I elements. Some 10–15% of the *Drosophila* genome is made up of MR DNA and about half of this fraction consists of 30–50 families of *copia*-like sequences (Finnegan, 1985). These sequences have reiteration frequencies ranging from 30–60 copies per embryonic cell genome to 40–170 copies per tissue culture cell genome (Potter *et al.* 1979). The functions of both short and long MR sequences remain to be determined although it has been suggested that the short sequences may play a role in gene regulation. Unique or single-copy (SC) DNA contains protein-encoding sequences. Such sequences are not exclusive to this component, some protein coding genes being represented more than once per haploid genome.

Although dispersed repetitive sequences are characteristic of eukaryotic genomes, no single unifying description of their arrangement within the genome can be applied to all organisms (Crain *et al.* 1976; Jelinek & Schmid, 1982). However, the genomes of eukaryotic organisms can be classified into one of two groups; those exhibiting a short period interspersion pattern of repeats (SPI) and those exhibiting a long period interspersion (LPI) pattern of repeats. Examples of LPI patterns are found in the genomes of *Drosophila* (Crain *et al.* 1976) and the mosquito *Anopheles quadrimaculatus* (Black & Rai, 1988). Here, repetitive sequences of approximately 5000 bp are dispersed among single-copy sequences that can be as long as 35000 bp. Human (Jelinek *et al.* 1980), *Xenopus* (Davidson *et al.* 1973) and Ae. *aegypti* (Gale, 1987) genomes exhibit an SPI pattern of repeated interspersion. Here, repetitive sequences of approximately 100–300 bp are interspersed with SC sequences of approximately 1000-2000 bp. Although up to 80% of the DNA of such organisms exhibit this SPI organization, this does not preclude the presence of other types of sequence arrangements within the genome such as the interspersions of repeated sequences significantly longer than 300 bp.

Patterns of sequence arrangement and genome complexities have important implications in experimental design when attempting to isolate DNA sequences. For example, HR sequences within genomic clones containing large DNA inserts from an

* Corresponding author.

organism with an SPI arrangement of repeats may well mask the hybridization characteristics of sequences of interest, e.g. low MR or unique DNA, particularly when probing with complex genomic DNA probes.

In contrast to many systems, the molecular analysis of mosquitoes has so far been very limited. The formal genetics of a number of mosquito strains have been moderately well characterized (Munnsterman & Craig, 1979). The Mendelian basis of insecticide resistance and susceptibility to vector-borne pathogens has been investigated in a number of species together with the phenomena of hybrid sterility, cytoplasmic incompatibility and induced sterility. All these factors have significance in relation to vector control programmes. The ability to clone genes on a functional basis and to manipulate the genome would be of great benefit if applied to insects which are important vectors of disease. An essential prerequisite towards this end is the molecular characterization of the mosquito genome.

Because of its importance as a disease-carrying vector, coupled with its suitability for laboratory research and its potential for studies in genetics, we have directed our research towards the molecular characterization of the mosquito, *Ae. aegypti*, the major urban vector of yellow fever dengue and dengue haemorrhagic fever (DHF). Here, we describe the characterization of the genome complexity of this organism and a cell line derived from it. In addition, sequences from the MR fraction of the *Ae. aegypti* genome have been cloned and their distribution within the genome characterized.

2. Materials and methods

The following strains of mosquito were used in this study and are maintained in the insectaries in the Liverpool School of Tropical Medicine.

Ae. aegypti Cayenne, colonized in 1984, originated in Cayenne, French Guiana.
Ae. aegypti Thai, colonized in 1984, originated in Mahidol, Bangkok, Thailand.
Ae. aegypti London, colonized in 1957, originated in West Africa.
Ae. aegypti Liverpool, colonized in 1936, originated in West Africa
Ae. aegypti Porto Novo, colonized in 1969, originated in Porto Novo, Dahomey, West Africa.
Ae. aegypti American SS, colonized in 1962, a line selected from the Liverpool strain.
Ae. aegypti Bangkok, colonized in 1972, originated in Bangkok, Thailand.
Ae. aegypti REFM, colonized in 1963, originated in West Africa.

Two mosquito cell lines were also used. The Mos20 cell line is thought to have been cultured from first instar larvae of the London strain in 1969. The Warwick cell line is from the Department of Virology, Warwick University.

(i) *Isolation of mosquito DNA*

High-molecular-weight genomic DNA was isolated from mosquitoes essentially as described by Bingham *et al.* (1981). 100–400 fourth-instar larvae, pupae or adults were ground up in liquid nitrogen using a pre-cooled mortar and pestle. The product was suspended in 3 ml of buffer A (10 mM Tris-HCl pH 7·5, 60 mM-NaCl, 10 mM-EDTA, 0·15 mM spermidine, 0·15 mM spermine, 0·5% Triton X-100), homogenized on ice and nuclei collected by centrifugation at 7000 rpm, 7 min, 0 °C. The nuclei were washed once in buffer A and then resuspended in 5 ml buffer A and 10% n-lauryl sarcosine was added to a final concentration of 2%. The suspension was mixed gently to lyse the nuclei and caesium chloride was then added to a final density of 1·68 g/ml. The DNA was then banded by centrifugation at 40000 rpm for 40 h at 18 °C in a Sorvall T865 rotor. The gradient was fractionated and the DNA-containing fractions detected by spotting samples onto a 1% agarose gel containing 0·5 μg/ml ethidium bromide. These fractions were pooled and dialysed extensively against Tris/EDTA (10 mM Tris-HCl pH 7·5, 0·1 mM-EDTA). This method yielded approximately 1–2 μg of DNA per mosquito adult, pupae or fourth-instar larva. Females are larger and yield 20–50% more DNA than males.

(ii) *Preparation of mosquito cell-line DNA*

The *Ae. aegypti*, Mos20 cell line described by Varma & Pudney (1969) was grown in MM medium. The medium was removed and the cells suspended in ice-cold Hayes saline solution (9 g NaCl, 0·26 g $CaCl_2.2H_2O$, 0·2 g KCl, 0·1 g $NaHCO_3$ per litre) using a rubber policeman. The cells were washed in Hayes saline and resuspended evenly in 4·5 ml of lysis solution (0·32 M sucrose, 10 mM Tris-HCl pH 7·5, 5 mM-$MgCl_2$, 1% Triton X-100) and homogenized gently on ice. The nuclei were collected by centrifugation at 10000 rpm, 10 min at 4 °C and resuspended in 2·25 ml of 75 mM-NaCl, 24 mM-EDTA. 125 μl of 10% SDS, 50 μl of 10 mg/ml proteinase K and 75 μl of water were added and the mixture was incubated for 1 h at 37 °C, and then deproteinized by three extractions with phenol/chloroform and two with chloroform. Finally the DNA was recovered by ethanol precipitation.

(iii) *Southern blot analysis*

Restriction endonucleases from Boehringer Mannheim were used as recommended by the supplier.

https://doi.org/10.1017/S0016672300029979 Published online by Cambridge University Press

DNA fragments were separated by horizontal electrophoresis in 1% agarose gels in Tris/acetate/EDTA buffer (40 mM Tris/acetate pH 7·7, 20 mM sodium acetate and 2 mM-EDTA) at 1·5 V/cm. DNA was transferred to nitrocellulose as described by Southern (1975).

DNA was labelled by nick-translation to a specific activity of 2×10^8 cpm/μg using an Amersham kit by the method of Rigby *et al.* (1977). DNA on nitrocellulose filters was pre-hybridized in 5 × SSC (750 mM NaCl, 75 mM sodium citrate), 50% (by vol.) formamide, 5 × Denhardt's solution, containing 50 μg/ml sheared, denatured salmon sperm DNA and 10 μg/ml poly(adenylic acid) for 5 h at 42 °C. Subsequent hybridization was performed in the same buffer containing the denatured probe DNA at 42 °C for 18 h. Filters were washed as indicated in the figure legends and exposed to Fuji X-ray film at −70 °C with intensifying screens.

(iv) C_0t *analysis of* Ae. aegypti *genomic DNA*

Renaturation of *Ae. aegypti* genomic DNA was monitored using S1 nuclease analysis coupled with TCA precipitation (Britten *et al.* 1976). Driver DNA was sheared by sonication to a mean size of 450 bp using a Soniprobe type 7532B (Ultrasonic Instruments Div., Dawes Ltd, U.K.). Following sonication and sizing through a 2% agarose gel, DNA fragments were ethanol precipitated and resuspended in 0·12 M phosphate buffer (PB) pH 7·0 at concentrations of 10 μg DNA/ml for low C_0t values, 100 μg DNA/ml for mid C_0t values and 1 mg DNA/ml for high C_0t values. Tracer DNA was prepared by nick translation of genomic DNA to a specific activity of 10^8 dpm/μg DNA.

Driver and tracer DNAs were mixed and denatured by boiling for 10 min. Renaturation was performed at 63 °C. For each DNA, samples were taken at timed intervals to cover log C_0t values ranging from −2·7 to 4·9. Data was analysed using a nonlinear regression program (Green *et al.* 1982) adapted for DNA reassociation kinetics followed by S1 nuclease analysis (J. Ellis, J. Crampton and H. Townson, unpublished). This adapted program allows analysis of the separate components of genomic DNA, e.g. SC, MR and HR DNAs.

(v) *Hydroxylapatite chromatography*

Hydroxylapatite chromatography was employed for the isolation of the middle repetitive fraction of the genome. Following hybridization at 63 °C for the specified time, the sample was loaded on to a water-jacketed hydroxylapatite chromatography column maintained at 63 °C. Single-stranded (ss) DNA was eluted with several washes of 0·15 M-PB, pH 7·0 at 60 °C. Double-stranded (ds) DNA was then eluted with 0·4 M-PB, pH 7·0.

3. Results

(i) *Complexity analysis*

Renaturation of the *Ae. aegypti* genomic DNA was seen to occur over 7 orders of C_0t (Fig. 1 *a*, *b*) which is consistent with previous data for complex eukaryotic genomes containing SC, MR and HR DNA fractions (Britten & Kohne, 1968). These data were analysed by a nonlinear regression computer program as indicated in the methods and, by comparison with the renaturation of *E. coli* DNA under identical conditions. The $C_0t_{\frac{1}{2}}$ values, complexity and average copy number for each of the three components were calculated and are shown in Table 1. As indicated in the Table, the true $C_0t_{\frac{1}{2}}$ values were derived from the observed $C_0t_{\frac{1}{2}}$ values after correction for the percentage of the total DNA present in each component. The sizes of the *Ae. aegypti* larval and cultured cell haploid genomes were calculated by summation of the number of bases present in each of the three components ($1·5 \times 10^9$ for Mos20; 8×10^8 for Bangkok). These values are in good agreement with previous estimates (see Discussion).

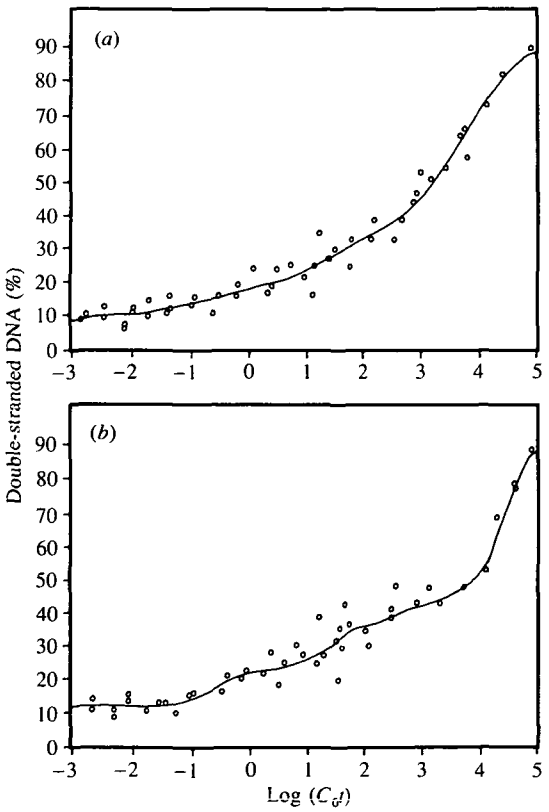

Fig. 1. Reassociation of *Aedes aegypti* DNA from (*a*) the Bangkok strain (larvae) and (*b*) the Mos20 cell line.

https://doi.org/10.1017/S0016672300029979 Published online by Cambridge University Press

Table 1. C_0t *data for* Aedes aegypti *Mos20 cell line and Bangkok strain larval genomic DNA*

Component	$C_0t_{\frac{1}{2}}$ (observed)	Proportion of genome	$C_0t_{\frac{1}{2}}$ (corrected)	Total base pairs	Complexity	Copy number
			(*a*) Mos20 cell line			
Highly repetitive	0·06	0·13	0·007	$1·4 \times 10^8$	$3·7 \times 10^3$	10^5
Middle repetitive	10·0	0·24	2·4	$2·6 \times 10^8$	$1·2 \times 10^6$	10^3
Single copy	3162	0·64	2024	$1·1 \times 10^9$	$1·1 \times 10^9$	1
Haploid genome		1		$1·5 \times 10^9$		
			(*b*) Bangkok strain (larval)			
Highly repetitive	0·1	0·2	0·02	$1·6 \times 10^8$	$1·07 \times 10^4$	10^6
Middle repetitive	56·23	0·2	11·24	$1·6 \times 10^8$	$6·05 \times 10^6$	10^3
Single copy	1485	0·6	891	$4·8 \times 10^8$	$4·8 \times 10^8$	1
Haploid genome		1		8×10^8		

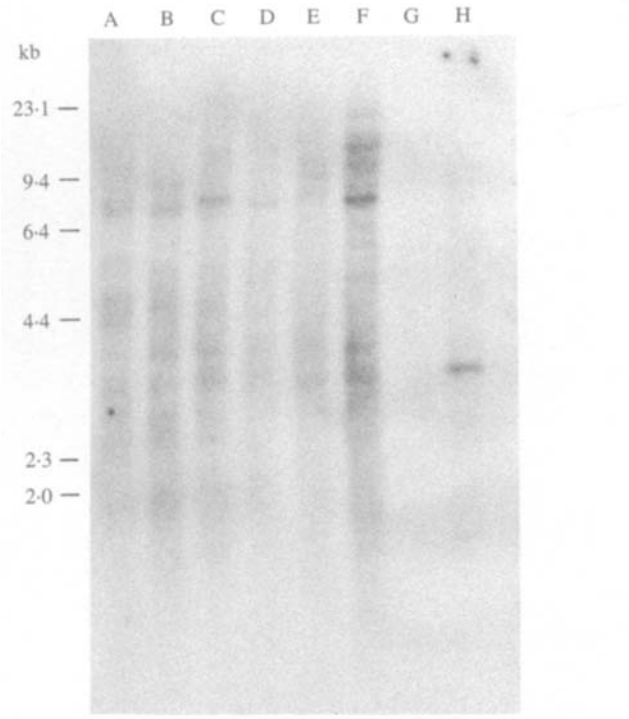

Fig. 2. Southern blot hybridization of digested total genomic *Ae. aegypti* probed pMR1. 5 μg of each DNA were digested with *Eco*R I and run through a 0·8% agarose gel. The filter was hybridized for 18 h at 42 °C, washed in 0·1 × SSC at 65 °C and autoradiographed at −70 °C for 48 h with an intensifying screen. Digested *Ae. aegypti* DNA from the indicated strains and cell lines were loaded on the gel as follows. Lane A, American SS; lane B, Liverpool; lane C, Cayenne; lane D, Porto Nova; lane E, Warwick; lane F, Mos20; lane G, London; lane H, Bangkok. Sizes are in kilobase pairs.

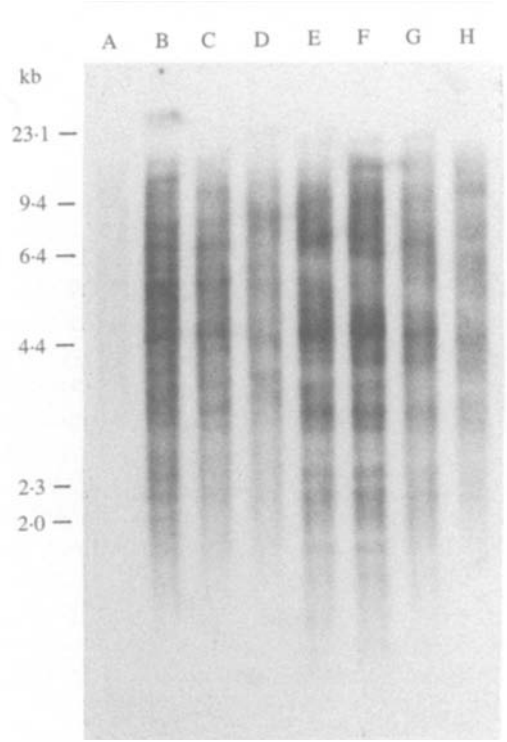

Fig. 3. Southern blot hybridization of digested total genomic *Ae. aegypti* probed with pMR4. The conditions for restriction enzyme digestion, blotting, hybridization and post-hybridization were as given in the legend to Fig. 2. Digested *Ae. aegypti* DNA from the indicated strains and cell lines were loaded on the gel as follows. Lane A, REFM; lane B, Bangkok; lane C, Thai; lane D, Cayenne; lane E, Porto Novo; lane F, American SS; lane G, London; lane H, Mos20. Sizes are in kilobase pairs.

(ii) *Isolation of middle repetitive* Ae. aegypti *DNA*

Using the data from the above analysis (Table 1), renaturation of *Ae. aegypti* Mos20 genomic DNA was carried out to a predetermined C_0t value of 0·014 that would allow renaturation of the HR DNA. The time of incubation, using a concentration of 5 μg DNA/ml, required to achieve this C_0t value was calculated to be 13·9 min. Renaturation conditions were as above. After 14 min at 63 °C, the sample was fractionated into single-stranded and double-stranded components. Addition of labelled tracer DNA to the reaction allowed the proportion of dsDNA (14%) and ssDNA (86%) to be determined.

The ssDNA eluate was dialysed to remove excess sodium phosphate, ethanol precipitated and the pellet taken up in 0.12 M-PB, pH 7·0 to the original concentration of 5 μg/ml. The renaturation reaction was then repeated as above. After eluting ssDNA and dsDNA as before, 3·0% of the genomic DNA was observed in the dsDNA eluate. The ssDNA eluate was again dialysed, ethanol precipitated and the pellet

https://doi.org/10.1017/S0016672300029979 Published online by Cambridge University Press

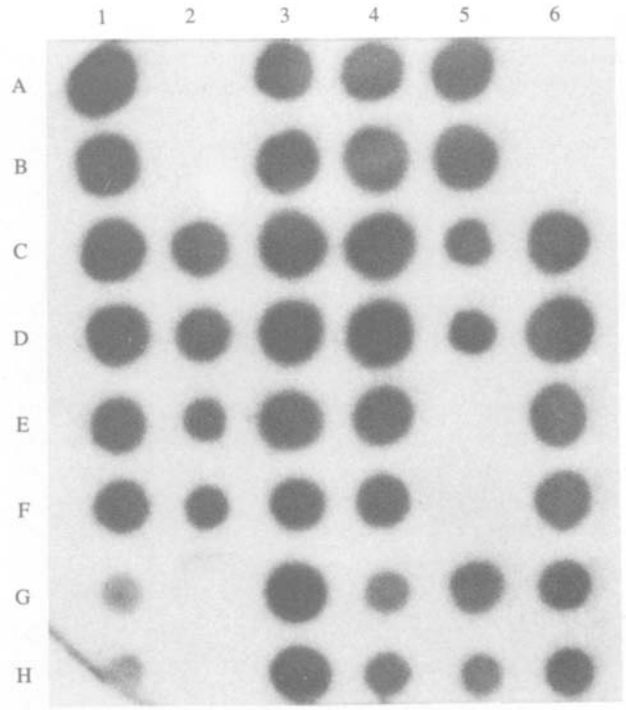

Fig. 4. Ordered array of middle repetitive and extrachromosomal clones probed with the RT type 3 oligonucleotide. Recombinant clones were positioned in the ordered array with the following coordinates. 1A&B–pX1; 2A&B–PS2; 3A&B–pX3; 4A&B–pX4; 5A&B–pX5; 6A&B–pX6; 1C&D–pX7; 2C&D–pX8; 3C&D–pX9; 4C&D–pX10; 5C&D–pX12; 6C&D–pX13; 1E&F–pX15; 2E&F–pX16; 3E&F–pMR1; 4E&F–pMR2; 6E&F–pMR4; 1G&H–pMR5; 2G&H–pMR6; 3G&H–pMR10; 4G&H–RT sub-clone *17.6*; 5G&H–RT sub-clone *copia*; 6G&H–RT sub-clone *412*; 5E&F–blank control. (*17.6, copia* and *412* are retroposons from *Drosophila melanogaster*). After hybridizing for 20 h at 30 °C, filters were washed twice for 5 min in 6 × SSC, 0·1 % SDS at 25 °C and autoradiographed for 2 days at −70 °C with an intensifying screen. Abbreviations: pX, plasmid extrachromosomal; pMR, middle repetitive clone; RT, reverse transcriptase. Sizes are in kilobase pairs.

taken up in 0·12 M-PB to give a sample concentration of 100 μg DNA/ml. This sample was boiled for 10 min and renatured to a predetermined C_0t value of 4·8 allowing renaturation of MR DNA. Using a DNA concentration of 100 μg/ml, the time of incubation required to achieve this C_0t value was calculated to be 4 h.

After renaturation, the sample was ethanol precipitated, taken up in 100 μl S1 nuclease buffer plus 50 units S1 nuclease and incubated for 1 h at 37 °C to remove ssDNA. This was followed by phenol extraction and ethanol precipitation. After vacuum drying, the precipitate was resuspended in TE buffer, and the amount of MR DNA was estimated to be 5 μg as determined by gel electrophoresis, i.e. 5% of the original amount of sheared genomic DNA.

(iii) *Cloning of* Ae. aegypti *Mos20 middle repetitive DNA*

Following isolation of middle repetitive DNA, the fragment ends were repaired using T4 polymerase. This blunt-ended DNA was ligated into the *Sma* I site of the plasmid vector pUC12 and used to transform *E. coli*, strain MC1061. Transformants were selected by their ampicillin resistance. After taking control values into consideration, it was estimated that 800 ng of recombinant DNA yielded 4300 recombinant colonies. Taking $2{\cdot}6 \times 10^8$ bp as the fraction of MR DNA in the *Ae. aegypti* Mos20 genome (Table 1), the probability of having any sequence with an average reiteration frequency of 100, 1000 or 10000 represented in a library of 4300 middle repetitive clones (estimated average insert size of 450 bp) was calculated using the Clarke & Carbon (1976) equation to be 0·75 for a sequence present 100 times in the genome, 0·99 for a sequence present 1000 times and > 0·99 for a sequence present 10000 times. This calculation would suggest that most sequences of interest within this fraction of the genome would have a reasonable probability of being represented in this clone bank.

(iv) *Analysis of the middle repetitive library*

To identify and estimate the percentage of clones containing HR DNA and SC DNA sequences in the MR DNA library, 200 recombinant clones were plated onto a nitrocellulose filter and screened using a radiolabelled Mos20, total genomic DNA probe. The filter was washed at a high stringency, i.e. 0·1 × SSC, 65 °C, two 30-min washes, following hybridization. Autoradiography was performed at −70 °C initially

	▼		▼		▼ ▼▼
AART8	HAFART	<06>	YRHPDTBELER	<52>	LLLYVDICW
Tp 1	NAYLNG	<06>	YMIPSQGWTWD	<68>	ASLHVDEFL
Tal-3	TAFLHG	<06>	YMEQPEGCISE	<58>	LLLYVDDML
copia	TAFLNG	<06>	YMRLPQGISCN	<58>	VLLYVDDVV
Ty912	SAYLYA	<06>	YIRPPPHLGMN	<50>	ICLFVDDMV
Tnt 1-94	TAFLHG	<06>	YMEQPEGFEVA	<57>	LLLYVDDML

Fig. 5. The predicted amino-acid sequence from the *Aedes aegypti* sequence, AART8, aligned with the reverse transcriptase domains of Tp1, Tal-3, *copia*, Ty912 and Tnt1-914 (Rothnie *et al.* 1991). The common amino acids are arrowed.

https://doi.org/10.1017/S0016672300029979 Published online by Cambridge University Press

for 2 days and then 17 days. Under these conditions, one would expect hybridization of a HR sequence to be visualized within 48 h of autoradiographic exposure whereas hybridization to low MR or SC DNA sequences may not be apparent even following a 17-day exposure. Therefore comparison of the two autoradiographs and the master plate should provide an indication of the reiteration value of the insert DNA in each clone.

Following 48 h autoradiographic exposure, approximately 5% of the colonies showed strong hybridization signal intensities (data not shown). These colonies were taken to be recombinant clones containing HR DNA sequences. Approximately 3% of the colonies failed to give a hybridization signal following 17-day autoradiography. It was assumed that these clones contained either SC DNA inserts or MR DNA inserts having a very low copy number. These results suggest that at least 92% of the recombinant clones in the 'MR' DNA library do, in fact, contain MR DNA sequences which cover a wide range of reiteration values.

The plasmid DNA was isolated from several clones containing MR sequences, radiolabelled and used to probe Southern blots of *Eco*R I-digested total genomic DNA from a variety of *Ae. aegypti* strains and cell lines. The hybridization patterns exhibited by two MR clones, pMR1 and pMR4 are illustrated in Figs. 2 and 3, respectively.

Clearly these sequences display intraspecific variation in chromosomal distribution. Furthermore, it is also apparent that the copy numbers of these sequences vary between strains. Particularly noticeable is the apparent absence of pMR1 insert sequences from the American SS strain (Fig. 2) and of pMR4 from the London strain (Fig. 3). Intraspecific variations in chromosomal distribution and copy number are diagnostic features of mobile elements.

As has been shown in *Drosophila melanogaster* (Finnegan, 1985) approximately 6% of the MR fraction of the genome consists of 30–50 families of transposable genetic elements known as retroposons i.e sequences that transpose via an RNA intermediate. Intrinsic to this mode of transposition is reverse transcriptase activity. We therefore decided to synthesize an oligonucleotide with the sequence 5′-TAPyGTXGAPyGAPyATG-3′ (RT oligo, where X is any nucleotide and Py is a pyrimidine) which corresponds to the amino-acid sequence best conserved among retroviruses and retroposons, i.e. the YXDD box (Yuki *et al.* 1986), and used this radiolabelled sequence to screen the MR recombinant clones. As illustrated in Fig. 4 the RT oligo clearly hybridizes to both pMR1 and pMR4. These results together with the observed intraspecific variation displayed by MR1 and 4 suggest that both sequences may be from retroposons. Moreover, partial sequence data for a pMR4 homologous clone, AART8, contains the best conserved sequence areas characteristic of reverse transcriptases. Fig. 5 shows the predicted amino-acid sequence of a portion of AART8 aligned with the reverse transcriptase domains of several other retrotransposons (Rothnie *et al.* 1991). This data supports the idea that these sequences may represent retrotransposons.

4. Discussion

Comparison of the C_0t data for *Ae. aegypti* (Mos20) cell line genomic DNA, with *Ae. aegypti* (Bangkok) larval genomic DNA revealed several differences. Firstly, there is an apparent 2-fold difference in the actual genome size, being $1{\cdot}5 \times 10^9$ bp and $8{\cdot}0 \times 10^8$ bp respectively (Table 2). This variation may reflect changes in the chromosomal content of cultured cells involving such phenomena as aneuploidy (Freshney, 1986). Furthermore, Potter *et al.* (1979) have observed a 3- to 5-fold increase in the copy number of several indigenous TGEs in *D. melanogaster* tissue culture cell genomes as compared with the genomes of embryonic cells. Presumably this increase is 'allowed' because of the 'permissive' environment afforded to tissue culture cells and hence should tolerate more chromosomal changes than embryonic cells.

The hybridization studies of MR clones described here further suggest that the data presented in Table 2 may also be a reflection of elevated levels of replicative transposition of endogenous TGEs within the genomes of *Ae. aegypti* isolated; the Mos20 cell line being derived from a strain other than Bangkok, i.e. London. Black & Rai (1988) suggest that intraspecific variation in genome size in *Ae. albopictus* was primarily due to differences in the amounts of HR DNA. This suggestion is supported somewhat by the data presented in Table 2. An increase in HR DNA could be generated by amplification of HR DNA sequences *per se*. However, they may also be caused indirectly through amplification of MR DNA sequences. This could result in a 'shifting' of sequences from the MR fraction to the HR fraction of the genome. The degree of sequence amplification would no doubt depend on physiological and environmental factors (Young & Schwartz, 1981). One would suspect, however, that high copy number MR sequences would be more prone to a change in 'status' than those with low reiteration frequencies. Amplification of MR DNA could of course produce an apparent increase in the MR DNA fraction, without any obvious increase in the HR DNA fraction. It is perhaps worth emphasizing here that there is no physical distinction between HR and MR sequences in the genome, the demarcation being an arbitrary one for the purpose of simplifying the analysis. There is, therefore, a continuum of levels of repetition of sequences in the genome.

If amplification of MR and/or HR sequences occurs differentially between strains, one might expect

Table 2. *Sequence organization of* Aedes *genomic DNA*

	Total genome Size (bp)	DNA/cell (pgs)	Proportion of genome			Base pairs/component[c]		
			HR	MR	SC	HR	MR	SC
Ae. aegypti (larvae)	8×10^8	0·83	0·2	0·2	0·6	$1·6 \times 10^8$	$1·6 \times 10^8$	$4·8 \times 10^8$
Ae. aegypti (cell line)	$1·5 \times 10^9$	1·5	0·13	0·24	0·64	$1·4 \times 10^8$	$2·6 \times 10^8$	$1·1 \times 10^9$
Ae. albopictus (Calcutta[a])	$8·3 \times 10^8$	0·83	0·27	0·37	0·36	$2·2 \times 10^8$	$3·1 \times 10^8$	$3·0 \times 10^8$
Ae. albopictus (Mauritius[a])	$1·3 \times 10^9$	1·3	0·4	0·27	0·33	$5·2 \times 10^8$	$3·5 \times 10^8$	$4·3 \times 10^8$
Ae. albopictus (Spradling[b])	8×10^8	0·82	0·1	0·1–0·15	0·75–0·8	$8·0 \times 10^7$	$9·6 \times 10^7$	$6·2 \times 10^8$

Abbreviations: HR, highly repetitive; MR, middle repetitive; SC, single copy.
[a] Black & Spradling (1988); [b] Spradling *et al.* (1974); [c] The values for *Ae. albopictus* were calculated from data presented in references [a] and [b].

to detect the quantitative changes in these two components as illustrated in Table 2. There is an obvious increase of MR DNA in the *Ae. aegypti* Mos20 genome ($2·6 \times 10^8$ bp) as compared with Bangkok MR DNA ($1·6 \times 10^8$ bp) with little apparent change in the HR component in the two genomes. This suggests that differential amplification of MR sequences is in part responsible for the increase in the Mos20 genome size. Comparison of these two components in the genomes of *Ae. albopictus* strains (Table 2) reveals significant intraspecific variation particularly with regard to the HR DNA fraction, i.e. Calcutta HR DNA, $2·2 \times 10^8$ bp; Mauritius HR DNA, $5·2 \times 10^8$ bp; Spradling HR DNA, 8×10^7 bp. Furthermore, the genome of the latter strain appears to house three times less MR DNA than either the Calcutta or Mauritius genomes. Again it appears that those genomes with greater amounts of repetitive DNA, be it HR or MR DNA, are in fact larger.

Whilst the values for size and complexities of the total genome and individual components of the *Ae. aegypti* genome may not be absolutely accurate, they can at least be taken as good estimates. Furthermore, the information obtained through these analyses has proved to be most useful in determining the parameters necessary for separation of the different genomic fractions and in particular the preparation of a clone bank which is a reasonable representation of *Ae aegypti* Mos20 MR sequences.

Construction of this short insert size, MR DNA library was carried out in an attempt to circumvent the problems of sequence masking discussed above. In this respect it has served a valuable purpose and has led to the isolation of a number of interesting DNA sequences including pMRI and 4. The intraspecific variation observed for both the clones pMR1 and pMR4, is in marked contrast to the remarkable degree of homogeneity observed for the rDNA of *Ae. aegypti* (Gale & Crampton, 1989). The rDNA copy number in *Ae. aegypti* has been estimated at 500 copies per haploid genome. As in all other higher eukaryotes so far studied, these genes are organized in a tandem, head to tail, repeated fashion. Mapping data for rDNA clones have indicated homogeneity of the 9 kb repeat length. However, restriction enzyme site polymorphisms and variation in unit length have been observed in several clones. This variation appears to be due to integration of nomadic non-rDNA within the non-transcribed spacer regions of the rDNA repeat unit. This integrated DNA may well be similar to site-specific rDNA insertion elements observed in the mosquitoes *Anopheles gambiae* and *A. arabiensis* (Paskewitz & Collins 1989) and/or the RI, RII elements in other insect rDNA (Weiner *et al.* 1986).

The intraspecific variation in distribution of the sequences present in the clones pMR1 and 4, together with their homology to the conserved portion of the reverse transcriptase sequence would suggest that they may represent retroposons. However, a good deal more characterization, such as DNA sequence determination and characterization of terminal regions, is required before any firm conclusions can be drawn. These sequences show similarities to well characterized TGEs from other organisms such as the *copia*-like elements in *Drosophila*. Endogenous TGEs have provided a practical means for both the molecular characterization and manipulation of *Drosophila*. In a similar fashion, endogenous TGEs may prove useful tools, as markers and possibly mutagens, for achieving these objectives in *Ae. aegypti* and other medically important insect vectors of disease.

This study, together with filter hybridization experiments using an *Ae. aegypti* genomic library probed with total genomic and rDNA probes (Gale, 1987) has shown that the genome of *Ae. aegypti* is both large (one third the size of the human genome) and complex and exhibits and SPI pattern of repeat sequences. As such, molecular characterization of this genome will prove more difficult than the *Drosophila* genome, but is essential if molecular biology is to make a significant

contribution to our understanding of insects which are important disease carriers.

We thank the Wellcome Trust for financial support during the course of this work. J. M. C. is a Wellcome Trust Senior Research Fellow in Basic Biomedical Sciences.

References

Black, W. C. & Rai, K. S. (1988). Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genetical Research* **51**, 185.

Bingham, P. M., Levis, R. & Rubin, G. M. (1981). Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. *Cell* **25**, 693–704.

Britten, R. J. & Kohne, D. E. (1968). Repeated sequences in DNA. *Science* **161**, 529–540.

Britten, R. J., Graham, D. E., Eden, F. C., Painchaud, D. M. & Davidson, E. N. (1976). Analysis of repeating DNA sequences by reassociation. *Journal of Molecular Evolution* **9**, 1–23.

Clarke, L. & Carbon, J. (1976). A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9**, 91–96.

Coen, E. S., Today, J. M. & Dover, G. (1982). Rate of turnover of variants in the rDNA gene of *D. melanogaster*. *Nature* **295**, 564–568.

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

Davidson, E. H., Hough, B. R., Amenson, C. S. & Britten, R. J. (1973). General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *Journal of Molecular Biology* **77**, 1–23.

Finnegan, D. J. (1985). Transposable elements in eukaryotes. *International Review of Cytology* **93**, 281–326.

Finnegan, D. J. 91988). I-factors in *Drosophila melanogaster* and similar elements in other eukaryotes. *SGM Symposium 43* (ed. A. J. Kingsman, K. J. Chatev and S. M. Kingsman), 'Transposition', pp. 271–285. Cambridge University Press.

Freshney, R. I. (ed.) (1986). Introduction: principles of sterile technique and cell propagation. In *Animal Cell Culture, A Practical Approach*, pp. 1–11. Oxford and Washington, DC, IRL Press.

Gale, K. R. (1987). Characterization of the mosquito, *Aedes aegypti* at the molecular level of genetics. Ph.D. thesis, University of Liverpool.

Gale, K. R. & Crampton, J. M. (1989). The ribosomal genes of the mosquito *Aedes aegypti*. *European Journal of Biochemistry* **185**, 311–317.

Green, S., Field, J. K., Green, C. D. & Beynon, R. J. (1982). A microcomputer program for analysis of nucleic acid hybridization data. *Nucleic Acids Research* **10**, 1411–1421.

Jelinek, W. R. & Schmid, C. W. (1982). Repetitive sequences in eukaryotic DNA and their expression. *Annual Review of Biochemistry* **51**, 813–844.

Munstermann, L. E. & Craig, G. B., Jr. (1979). Genetics of *Aedes aegypti*; updating the linkage map. *Journal of Heredity* **70**, 291–296.

Paskewitz, S. M. & Collins, F. H. (1989). Site-specific ribosomal DNA insertion elements in *Anopheles gambiae* and *A. arabiensis*: nucleotide sequence of gene-element boundaries. *Nucleic Acids Research* **17**, 8125–8133.

Potter, S. S., Brorien, W. J., Dunsmuir, P. & Rubin, G. M. (1979). Transposition of elements of the *412, copia* and *297* dispersed repeated families in *Drosophila*. *Cell* **17**, 429–439.

Rigby, P. W. J., Deickmann, M., Rhodes, C. & Berg, P. (1977). Ubiquitous, interspersed repeated sequences in mammalian genomes. *Journal of Molecular Biology* **113**, 237–251.

Rothnie, H. M., McCurrach, K. J., Glover, L. A. & Hardman, N. (1991). Retrotransposon-like nature of Tp1 elements: implications for the organization of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum*. *Nucleic Acids Research* **19**, 279–286.

Southern, E. M. (1975). Detection of specific sequences in among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Spradling, A., Penman, S., Campo, M. S. & Bishop, J. O. (1974). Repetitious and unique sequences in the heterogenous nuclear and cytoplasmic messenger RNA of mammalian and insect cells. *Cell* **3**, 32–30.

Varma, M. G. R. & Pudney, M. (1969). The growth and serial passage of cell lines from *A. aegypti* (L) larvae in different media. *Journal of Medical Entomology* **6**, 432–439.

Weiner, A. M., Deininger, 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.

Young, M. W. & Schwartz, H. E. (1981). Nomadic gene families in *Drosophila*. *Cold Spring Harbour Symposia on Quantitation Biology* **45**, 629–640.

Yuki, S., Ishimaru, S., Iuouye, S. & Saigo, K. (1986). Identification of genes for reverse transcriptase-like enzymes in two *Drosophila* retrotransposons, *412* and *gypsy*; a rapid detection method of reverse transcriptase genes using YXDD box probes. *Nucleic Acids Research* **14**, 3017–3029.