Genet. Res., Camb. (2000), **75**, *pp.* 129–135. *With* 2 *figures.* Printed in the United Kingdom © 2000 Cambridge University Press

Genome complexity and organization in the red imported fire ant *Solenopsis invicta* Buren

JIANBO LI* AND KEVIN M. HEINZ
Department of Entomology, Texas A & M University, College Station, TX 77843-2475, USA

(*Received 16 June 1999 and in revised form 12 August 1999*)

Summary

The red imported fire ant *Solenopsis invicta* is the most destructive invading arthropod in the southern United States, yet little is known about its genome complexity and organization. Here we report the size, organization and GC content of *S. invicta* genome. DNA reassociation kinetics using S1 nuclease assay and a modified second-order kinetics model indicated that the *S. invicta* genome is approximately 0·62 picograms or $5{\cdot}91 \times 10^8$ base pairs, composed of 36% unique, 41% moderately repetitive and 23% highly repetitive/foldback sequences. Comparison of the reassociation kinetics of short and long DNA fragments revealed that the sequence arrangement follows a pattern of short period interspersion, as in most organisms with relatively large genomes. Melting-temperature analysis showed that the GC content of the fire ant genomic DNA is 34·8%, similar to that of most eukaryotic organisms. The results reveal that the fire ant genome is much larger and more complex than those of a number of hymenopteran insects studied to date. Our study provides a foundation for further analysis and genetic manipulation of the *S. invicta* genome.

1. Introduction

Eukaryotic genomes are composed of repetitive and unique or single-copy sequences with genes often located in the latter (Britten & Kohne, 1968; Jelinek & Schmid, 1982; Lewin, 1997). Highly repetitive sequences (< 200 base pairs (bp) in length) are usually located in centromeric or telomeric positions; however, moderately repetitive sequences are interspersed with unique sequences throughout the genome, commonly in one of two patterns (Davidson *et al.*, 1975). Genomes of most organisms follow a pattern of short period interspersion in which short (200–600 bp) to moderately long (1000–4000 bp) repetitive sequences alternate with unique sequences of 1000–2000 bp. Species with small genomes (0·1–0·5 picograms (pg) per haploid genome) often exhibit a long period interspersion pattern in which long (> 5000 bp) repetitive sequences are interspersed with very long (> 10000 bp) unique sequences (Black & Rai, 1988).

Patterns of sequence arrangement, complexity of each sequence component, and guanine–cytosine (GC) contents of genomes are of significant importance in molecular studies (Warren & Crampton, 1991). Highly repetitive sequences are often present as tandem repeats with 10^5–10^6 copies per haploid genome. An analysis of these sequences has significant value in intraspecific taxonomic studies. Moderately repetitive sequences present in 2–10^4 copies are often transposable elements (Doolittle, 1985). Isolation and analysis of these sequences may reveal past or present activities and the role of transposable elements in the evolutionary history of the species. Unique or single-copy DNA generally contains protein-encoding sequences. Isolation of this sequence component can increase the chance of locating genes of interest. Genome size plays a significant role in the design of some molecular studies; for example, it determines the number of clones needed in the construction of genomic libraries. The pattern of sequence arrangement and GC content also influence cloning strategies and DNA hybridization conditions in such experiments (Warren & Crampton, 1991).

The red imported fire ant *Solenopsis invicta* Buren has been studied extensively since it first invaded the US seven decades ago and became a serious pest,

* Corresponding author. Tel: +1(409)458 2767. Fax: +1(409)845 3144. e-mail: jianbo-li@tamu.edu

inflicting damage in the hundreds of million dollars annually to the US economy and ecosystem. Little, however, is known about the genome of *S. invicta* other than that it consists of $n = 16$ chromosomes (Glancey *et al.*, 1976) and that it contains numerous stretches of dinucleotide repeats (Krieger & Keller, 1997). The objective of this study was to analyse the fire ant genome to provide a foundation for further molecular genetic studies of *S. invicta*. Specifically, we used melting-temperature analysis and reassociation kinetics to characterize GC content, complexity and organization of the fire ant genome. Such information will aid in the construction and screening of genomic libraries, gene discovery and strategies of molecular genetic manipulation in an effort to develop a novel genetically based control programme for *S. invicta*.

2. Materials and methods

(i) *DNA preparation and fragmentation*

Solenopsis invicta pupae that would become queens or males were collected in the vicinity of College Station, Texas. Genomic DNA was extracted from fresh samples using a Qiagen DNA Extraction Kit, Genomic-tip 500/G (Qiagen, Valencia, CA) following the manufacturer's protocol with the following minor modifications: RNase A (4 mg/ml) from Gentra System (Minneapolis, MN) and proteinase K (20 mg/ml) from Promega (Madison, WI) were used in place of the Qiagen counterparts. Freshly prepared proteinase K was used to ensure maximum yield. This protocol produced high quantities of high-molecular-weight DNA (> 50 kb). Extracted DNA was eluted in Tris buffer (10 mM, pH 8·0), with the DNA concentration adjusted to 1 mg/ml and quantified using a TKO-100 Mini Fluorometer (Hoefer Scientific Instruments, San Francisco, CA). Hoechst dye solution and calf thymus DNA standards for DNA quantification were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). *Escherichia coli* DNA was purchased from Sigma (St Louis, MO) and resuspended in Tris buffer at 1 mg/ml before use.

Short and long DNA fragments of *S. invicta* and *E. coli* were generated using a sonicator (High Intensity Ultrasonic Processor Model VCX600, Sonics and Materials, Danbury, CT). Sonication DNA solution (0·25 mg/ml) was prepared by mixing 3 volumes of sonication buffer (20 mM-Tris-HCl, pH 8·0, 1 mM-EDTA, pH 8·0, 5 mM-β-mercaptoethanol) (Sambrook *et al.*, 1989) and 1 volume of the aforementioned DNA solution. Eight millilitres of sonication DNA solution was placed in a 50 ml centrifuge tube and all tubes were placed in an ice bath (0 °C) before use. The amplitude of the sonicator was set at 25% and the standard horn with a $\frac{1}{2}$ inch tip (13 mm) was used for generating both short and long DNA fragments. Based on a preliminary analysis with the sonicator, short DNA fragments (600 bp, centre of a smear of 200–1200 bp) were produced with a 10 s burst, repeated 15 times. Long DNA fragments (2600 bp, centre of a smear of 600–4000 bp) were generated with two 5 s bursts followed by an additional 15 s burst. Sonication DNA solutions were placed back in the ice bath for 30 s between bursts. Fragmented DNA was precipitated with 0·2 M-NaCl and an equal volume of 100% isopropanol and centrifugation. Precipitated DNA was washed twice with 70% ethanol, air dried, and resuspended in H_2O. DNA fragment sizes were determined by electrophoresis using a 1·5% agarose–TBE gel. DNA concentration was estimated using a fluorometer.

(ii) *Estimation of melting temperature and GC content*

Short DNA fragments of both *S. invicta* and *E. coli* were diluted in 0·2 × SSC–1 mM-EDTA (1 × SSC = 0·15 M-NaCl, 0·015 M sodium citrate) at 0·1 mg/ml and dispensed in 25 μl aliquots into 0·3 ml polymerase chain reaction (PCR) tubes. All samples were incubated in a 50 °C water bath for 3 h followed by a 5 min incubation at 1 of 26 temperatures ranging from 50 °C to 100 °C at 2 °C intervals. Two samples or replicates were used at each temperature and 52 samples were used in a test for each species. The incubation temperatures were maintained using a thermocycler, the Perkin-Elmer GeneAmp PCR System 9600 (Norwalk, CT). The highest possible temperature on the thermocycler, 99·9 °C, was set as 100 °C. At the end of the incubation, 200 μl of prewarmed (at 37 °C) 1 × S1 buffer containing 50 U S1 nuclease (Promega, Madison, WI) was added and the samples were further incubated in a 37 °C water bath for 30 min to digest single-stranded DNA. The digestion reaction was terminated by adding 20 μl of stop buffer (1 M Tris, pH 9·0, 0·1 M-EDTA, pH 8·0). Samples were stored at 4 °C before the undigested double-stranded DNA was quantified using a TKO-100 fluorometer. The experiment was repeated twice with two replicates in each test.

A generalized logistic model,

$$y = y_0 + a\{1 + \exp[-(x - x_0)/b]\}^{-c},$$

was used for a least squares solution of the melting-temperature data based on the overall mean, two experiment means, or four individual replicates at each temperature, where y is percentage single-stranded DNA, y_0 is the minimum percentage, a is the difference between the minimum and maximum, x is the temperature, x_0 is the temperature at the inflection point, b is the slope of the tangent at the inflection point, and c is the degree of symmetry of the sigmoidal curve. Parameters of the model were estimated using PROC NLIN in SAS by providing a first partial

https://doi.org/10.1017/S0016672399004401 Published online by Cambridge University Press

derivative and a wide range of initial values for each parameter (SAS Institute, 1996). The graphing software SigmaPlot (Jandel Scientific, 1997) was used to generate a best-fit sigmoidal curve. The melting temperature (T_m), defined as the temperature at which 50% of the DNA denatured, was estimated by solving the above equation at $y = y_0 + \frac{1}{2}a$, yielding

$$x = T_m = x_0 - b\ln(2^{1/c} - 1),$$

which can be calculated using a simple SAS DATA statement. The GC content of *S. invicta* was determined using

$$T_m = K + (15{\cdot}7 \pm 0{\cdot}5)\ (\log_{10} X) + 0{\cdot}41(\%\ \mathrm{GC})$$

(Gillis *et al.*, 1970), where K is a rate constant and was estimated by substituting T_m for *E. coli* obtained from this experiment and the known GC content for *E. coli* (= 50%), and X is the concentration of SSC buffer (0·2×, in this experiment).

(iii) *Kinetics of DNA reassociation*

DNA reassociation kinetic analysis is commonly performed using either hydroxyapatite chromatography or S1 nuclease assay. S1 nuclease assay was used in this study because of its simplicity and reproducibility (Gillis *et al.*, 1970; Randhir & Hanau, 1997). DNA reassociation tests were performed at 62 °C for both short and long DNA fragments. *Solenopsis invicta* and *E. coli* DNA were prepared in 0·5 × SSC–1 mM-EDTA at 0·5 mg/ml, dispensed in 10 μl aliquots into 0·3 ml PCR tubes and stored at 4 °C before use. For both short (600 bp) and long (2600 bp) DNA fragment sizes, 17 C_0t values were used. DNA samples were denatured at 100 °C for 8 min using the Perkin-Elmer thermocycler as described in the previous section. The samples were then removed from the thermocycler and immediately placed in a water bath maintained at 62 °C for DNA renaturation. Two samples of *S. invicta* and two of *E. coli* were removed from the water bath at each predetermined C_0t value, and 200 μl of 1 × S1 buffer containing 50 U S1 nuclease (prewarmed at 37 °C) were immediately added to the samples. The samples were further incubated at 37 °C for 30 min to digest the unrenatured, single-stranded DNA. The digestion reaction was terminated by adding 20 μl of stop buffer. All samples were stored at 4 °C before quantification of the remaining double-stranded DNA using a TKO-100 fluorometer.

Data of DNA reassociation kinetics were analysed by fitting a reassociation kinetics model with fractions of single-stranded DNA as the dependent variable and C_0t values as the independent variable. Reassociation kinetics using hydroxyapatite chromatography can be accurately described by a second-order equation, $C_{ss}/C_0 = (1 + kC_0t)^{-1}$ (Pearson *et al.*, 1977; Lewin, 1997), while that assayed with S1 nuclease follows a modified second-order form, $C_{ss}/C_0 = (1 + kC_0t)^{-n}$ (Smith *et al.*, 1975; Pearson *et al.*, 1977). C_0t is measurement of DNA reassociation in units of moles of nucleotide phosphate per litre (C_0) × time, t, in seconds, at which reassociation of a DNA sample is terminated; C_{ss} is the amount of DNA remaining single-stranded at time, t; C_0 is the initial concentration of DNA; k is the reassociation rate constant; and n is the second-order variable which can be estimated using *E. coli* data obtained from this experiment (Smith *et al.*, 1975). When n reduces to 1, the modified equation reduces to the second-order form. A SAS program using PROC NCLIN (SAS Institute, 1996) was written specifically for fitting DNA reassociation kinetics data that follow a modified second-order model (this program can be obtained on request from J. L.). The S1 nuclease assay data were analysed using this SAS program. Fractions of three components (unique, moderately repetitive and highly repetitive sequences) and their corresponding reassociation rate constants, ks, which reflect the presence of other sequence components (if more than one exist), were initially estimated using the SAS program. The C_0t value at which 50% of a specific DNA sequence component renatured, $C_0t_{1/2}$, can be estimated by $C_0t_{1/2} = (1/k)(2^{1/n} - 1)$ for the modified second-order model. When n reduces to 1 as in the second-order model, $C_0t_{1/2} = 1/k$. Both the number of components and the initial value(s) of each parameter can be varied to achieve a best fit of the model determined by the minimum residual mean square (RME). A similar SAS program for estimating parameters of a second-order model was previously reported by Black & Rai (1988).

DNA sequence arrangement of the fire ant genome was analysed by comparing the reassociation kinetics of short and long fragments. If short and long fragments consist of similar DNA sequences as in a pattern of long period interspersion, the rate of reassociation of long fragments (k_L) can be predicted by the rate of reassociation of short fragments (k_S): $k_L = k_S \times \sqrt{L/S}$, where L and S are the length (bp) of short and long fragments, respectively (Wetmur & Davidson, 1968). A faster rate of reassociation of long fragments indicates the presence of repetitive sequences, which suggests that the genome is of a pattern of short period interspersion. Haploid genome size can be estimated by addition of the size of repetitive and unique fractions of the genome (bp) or by the following equation: genome size (pg) = $C_0t_{\frac{1}{2}(\mathrm{unique})}(0{\cdot}0044)/C_0t_{\frac{1}{2}E.coli}$, where 0·0044 (pg) is the size of the *E. coli* genome (Lewin, 1997).

3. Results

(i) *Melting temperature and GC content*

The GC content of the *S. invicta* genome was estimated

https://doi.org/10.1017/S0016672399004401 Published online by Cambridge University Press

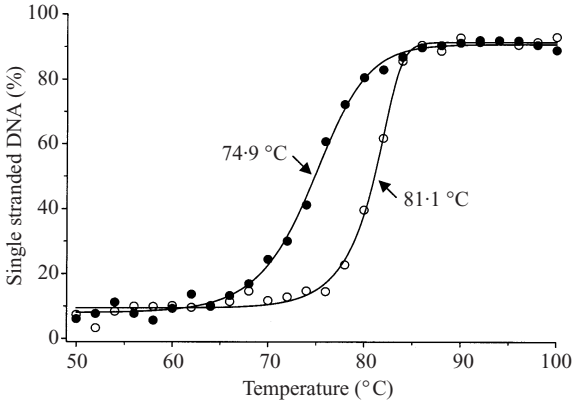

Fig. 1. Melting curves of *Solenopsis invicta* and *Escherichia coli* DNA. Open circles represent *E. coli* and filled circles *S. invicta*. Each circle represents a mean of two experiments each with two replicates. Arrows indicate melting temperatures.

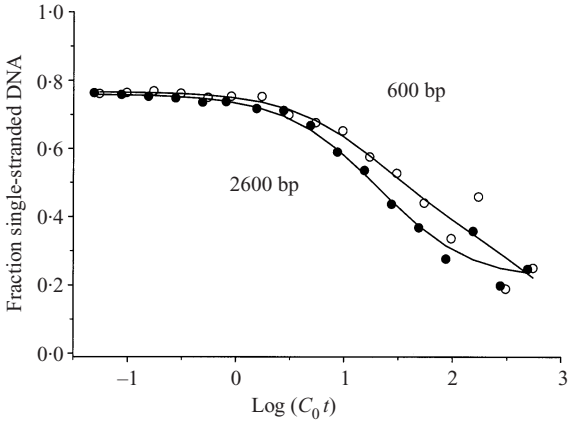

Fig. 2. $C_0 t$ curves for *Solenopsis invicta* DNA sheared to short and long fragment sizes. Both curves represent the least squares solutions with two components (moderately repetitive and unique sequences). The upper curve (open circles) represents renaturation of short DNA fragments (600 bp) and the lower curve (filled circles) of long DNA fragments (2600 bp). Each circle represents a mean of two experiments each with two replicates.

using melting-temperature analysis with *E. coli* genomic DNA as a standard. The thermal denaturation profiles for both species are presented in Fig. 1. Under our experimental conditions the melting temperature (T_m) was estimated to be 81·1 °C for *E. coli* and 74·9 °C for *S. invicta*. Thermal profile analysis based on the overall mean, two experiment means, or four individual replicates at each temperature yielded essentially identical T_m values for either species with at least 99 % of variability accounted for by the best-fit least squares model. The reaction constant, K, was determined to be $71{\cdot}6 \pm 0{\cdot}3$ ($\bar{x} \pm$ SE) based on the T_m estimated in this study and the known GC content for *E. coli* (= 50 %). The GC content of *S. invicta* was then calculated as $34{\cdot}8\% \pm 1{\cdot}7\%$ based on the K and T_m for the fire ant.

(ii) *Reassociation kinetics*

Solenopsis invicta DNA of both short (600 bp) and long (2600 bp) fragments were assayed using S1 nuclease. The results of the kinetic analysis are presented in Fig. 2. Fractions of single-stranded DNA were plotted against $\log_{10} C_0 t$ values. The best-fit curve for short or long DNA fragments represents a least squares solution with two components: moderately repetitive and unique sequences (Table 1). The curves were generated using the SAS program we developed. Reassociation kinetic analysis of short DNA fragments (600 bp) revealed that the *S. invicta* genome contains 36 % ($\pm 0{\cdot}139$ % SE) unique sequences and 41 % moderately repetitive DNA sequences with 34 copies per haploid genome. The remaining 23 % of the genome renatured before a $C_0 t$ value of 0·03 mol s, indicating that the sequences are highly repetitive and may consist of a large portion that fold back on themselves quickly (Lewin, 1997).

Long DNA fragments were assayed and analysed in comparison with short fragments to determine the pattern of sequence arrangement and genome organization in *S. invicta*. When the DNA fragment sizes increased from 600 bp to 2600 bp, the fraction of unique sequences of the genome reduced from 36 % to 21 %, suggesting that repetitive sequences are located within 1000 bp of 42 % of the 600 bp unique sequences. If the long DNA fragments consist of unique sequences only, the rate of reassociation would be 0·008 mol^{-1} s^{-1} as predicted by the rate of reassociation of the short fragments. However, a significantly faster rate was observed (Table 1), confirming the presence of repetitive sequences in the long fragments and indicating that the *S. invicta* genome follows a pattern of short period interspersion. This pattern of genome organization is common in organisms with genome larger than 0·5 pg (Black & Rai, 1988).

The haploid genome of *S. invicta* was estimated to be $5{\cdot}91 \times 10^8$ bp or 0·62 pg (ranging from 0·45 to 1·01 pg within 1 SE of unique sequences) (Table 1). The moderately repetitive sequences were of moderate complexity ($7{\cdot}2 \times 10^6$) with 34 copies, making up $2{\cdot}13 \times 10^8$ bp or 41 % of the genome. The remaining 23 % of the genome, or $1{\cdot}38 \times 10^8$ bp, probably consists of foldback/highly repetitive sequences.

4. Discussion

Our DNA reassociation kinetic analysis represents a significant step towards a better understanding of the *S. invicta* genome and provides a foundation for further study of fire ant molecular genetics. The information revealed in this study may facilitate the development and deployment of genetically based control strategies for this pest. The haploid genome of

https://doi.org/10.1017/S0016672399004401 Published online by Cambridge University Press

Table 1. *Reassociation kinetic analysis of* Solenopsis invicta *genomic DNA*

Species	DNA components	Fraction of fragments[a]	k_{mix} (mol^{-1} s^{-1})[a]	k_{pure} (mol^{-1} s^{-1})[b]	$C_0 t_{\frac{1}{2}mix}$ (mol s)[c]	$C_0 t_{\frac{1}{2}pure}$ (mol s)[d]	Copy numbers[e]	Complexity (bp)[f]	Size (bp)[g]
E. coli									
Short and long DNA fragments	Unique (short)	1·000	0·195	0·195	5·198	5·198	—	—	—
	(long)	—	0·380	0·380	2·668	2·668	—	—	—
S. invicta									
Short fragments	Foldback/highly repetitive	0·233	—	—	—	—	—	—	$1{\cdot}375 \times 10^8$
	Moderately repetitive	0·407	0·047	0·114	21·782	8·860	34	$7{\cdot}159 \times 10^6$	$2{\cdot}404 \times 10^8$
	Unique	0·360	0·001	0·004	731·220	263·551	1	$2{\cdot}127 \times 10^8$	$2{\cdot}127 \times 10^8$
	Haploid genome								$5{\cdot}906 \times 10^8$
Long fragments	Immediate	0·240							
	Fast	0·555	0·048	0·087	20·980	11·645			
	Slow	0·205	0·0003	0·001	3754·329	740·097			

[a] Fraction of each DNA component and corresponding k_{mix} were derived from the SAS program we developed for best-fitting a modified second-order model.
[b] $k_{pure} = k_{mix}$/fraction of corresponding DNA sequences.
[c] $C_0 t_{\frac{1}{2}mix} = (1/k_{mix})(2^{1/n} - 1)$, where n is the exponent variable in the modified second-order model.
[d] $C_0 t_{\frac{1}{2}pure} = (1/k_{pure})(2^{1/n} - 1)$ or $C_0 t_{\frac{1}{2}mix} \times$ fraction of corresponding component.
[e] DNA sequence copy number was estimated by $C_0 t_{\frac{1}{2}\,mix\,unique\,sequences} / C_0 t_{\frac{1}{2}mix\,repetitive\,sequences}$.
[f] Complexity of each fraction was estimated by $(k_{pure\ (Escherichia\ coli}/k_{pure\ (unique)}) \times$ genome $size_{(E.coli)(=4{\cdot}2\times10^6)}$.
[g] Size of each component was the product of copy number and complexity, except that of foldback/highly repetitive sequences, which was calculated as fraction of that component × genome $size_{(S.\ invicta)}$; and the size of haploid genome was the sum of the size of each sequence component.

S. invicta was estimated to be $5{\cdot}91 \times 10^8$ bp or 0·62 pg (Table 1). The fire ant genome is significantly larger than the average size of five other hymenopteran insect genomes, 0·26 pg, estimated to date (Regev *et al*., 1998). Our estimate of the fire ant genome based on S1 nuclease assay is in close agreement with an estimate (0·69–0·80 pg) done by flow cytometric analysis (Ellison & Johnston, 2000). The difference in genome size between the fire ant and previously studied hymenopteran insects may reflect differences in genome organization. Our study revealed that the fire ant genome is composed of three components: 36% unique, 41% moderately repetitive and 23% foldback/highly repetitive sequences (Table 1). The unique sequences represent only slightly over one-third of the genome. In the honeybee, *Apis mellifera* L., over 90% of its 0·35 pg genome is composed of unique sequences (Crain *et al*., 1976). A small fraction of unique sequences ($< 40\%$) is characteristic of species with large genomes ($> 0{\cdot}5$ pg per haploid genome) (Black & Rai, 1988; Ma *et al*., 1992; Palmer *et al*., 1994).

Genome size has significant implications in certain molecular studies. For example, a large fire ant genome requires a large number of clones and more screening experiments in the construction and analysis of genomic libraries. However, a significant amount of unique sequences in the fire ant genome (36%) provides easy access to potential gene-containing sequences that can be isolated using hydroxyapatite chromatography or S1 nuclease assay. Isolation of unique DNA sequences will facilitate identification of potential genes of interest.

Our study indicates that sequence organization of the fire ant genome follows a pattern of short period interspersion in which unique sequences are interspersed with short to moderately long repetitive sequences (Table 1). The results suggest that sequence arrangement of the fire ant genome is similar to that of most species with large genomes. The large amount of moderately repetitive sequences (41%) indicates that there might be transposable elements in the fire ant genome, because many moderately repetitive sequences are derived from transposable elements (Doolittle, 1985). The possible existence of transposable elements, active or inactive, in the fire ant genome, along with other information revealed in this study will facilitate our effort to genetically transform the fire ant for basic ecological studies or for the possible development of novel genetic control strategies (Ashburner *et al*., 1998). Highly repetitive and foldback sequences comprise 23% of the fire ant genome. Although these type of sequences are often located at centromeric and telomeric positions (Jelinek & Schmid, 1982), they can also be distributed throughout the genome as well. The pattern of short period interspersion of DNA sequences in the fire ant genome has significant implications in molecular studies. For example, in genomic library screening with complex genomic DNA probes, identification of sequences of interest may be ineffective because either the hybridization conditions are difficult to optimize or the repetitive sequences mask the true identities of the sequences (Warren & Crampton, 1991).

The melting-temperature analysis in this study reveals that the GC content of the fire ant genome is 34·8%. Our GC estimate is similar to the 33–43% observed for most insect species previously studied (Shapiro, 1970; Black & Rai, 1988). This information will aid in the determination of hybridization conditions for future studies of the fire ant genome.

We thank R. Hanau and U. Melcher for helpful discussions, F. Robinson and S. Johnston for sharing equipment, and C. Coates, J. Mason and S. Hutchison for helpful comments on an early draft of the manuscript. This research was funded by a grant from the Texas Imported Fire Ant Research and Management Plan to the authors.

References

Ashburner, M., Hoy, M. A. & Peloquin, J. J. (1998). Prospects for the genetic transformation of arthropods. *Insect Molecular Biology* **7**, 201–214.

Black, W. C. IV & Rai, K. S. (1988). Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genetical Research* **51**, 185–196.

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

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

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

Doolittle, W. F. (1985). The evolutionary significance of middle repetitive DNAs. In *The Evolution of Genome Size* (ed. T. Cavalier-Smith), pp. 443–487. New York: Wiley.

Ellison, J. R. & Johnston, J. S. (2000). A rapid flow cytometric method for determination of DNA content from insect nuclei. *cytometry* (in press).

Gillis, M., Ley, J. D. & Cleene, M. D. (1970). The determination of the molecular weight of bacterial genome DNA from renaturation rates. *European Journal of Biochemistry* **12**, 143–153.

Glancey, B. M., Romain, M. S. St & Crozier, R. H. (1976). Chromosome numbers of the red and the black imported fire ants, *Solenopsis invicta* and *S. richteri*. *Annals of the Entomological Society of America* **69**, 469–470.

Jandel Scientific (1997). *SigmaPlot Scientific Graphing Software*, *Windows version 4*. San Rafael, CA: Jandel Scientific.

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

Krieger, M. J. B. & Keller, L. (1997). Polymorphism at dinucleotide microsatellite loci in fire ant *Solenopsis invicta* populations. *Molecular Ecology* **6**, 997–999.

Lewin, B. (1997). Genomes. In *Genes VI* (ed. B. Lewin), pp. 645–662. New York: Oxford University Press.

Ma, R. Z., Black, W. C. IV & Reese, J. C. (1992). Genome size and organization in an aphid (*Schizaphis graminum*). *Journal of Insect Physiology* **38**, 161–165.

Palmer, M. J., Bantle, J. A., Guo, X. & Fargo, W. S. (1994). Genome size and organization in the ixodid tick. *Amblyomma americanum* (L.). *Insect Molecular Biology* **3**, 57–62.

Pearson, W. R., Davidson, E. H. & Britten, R. J. (1977). A program for least square analysis of association and hybridization data. *Nucleic Acids Research* **4**, 1727–1737.

Randhir, R. J. & Hanau, R. M. (1997). Size and complexity of the nuclear genome of the *Colletrichum graminicola*. *Applied Environmental Microbiology* **63**, 4001–4004.

Regev, A. R., Lamb, M. J. & Jablonka, E. (1998). The role of DNA methylation in invertebrates: developmental regulation or genome defence? *Molecular Biology and Evolution* **15**, 880–891.

SAS Institute (1996). *The SAS System for Windows*, release 6.12. Cary, NC: SAS Institute.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Press.

Shapiro, H. S. (1970). Distribution of purines and pyrimidines in deoxyribonucleic acids. In *Handbook of biochemistry*, 2nd edn (ed. H. A. Sober), pp. H80–H103. Cleveland, OH: The Chemical Rubber Co.

Smith, M. J., Britten, R. J. & Davidson, E. H. (1975). Studies on nucleic acid reassociation kinetics: reactivity of single-stranded tails in DNA–DNA renaturation. *Proceedings of the National Academy of Sciences of the USA* **72**, 4805–4809.

Warren, A. M. & Crampton, J. M. (1991). The *Aedes aegypti* genome: complexity and organization. *Genetical Research* **58**, 225–232.

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