Mol. Cells, Vol. 14, No. 1, pp. 158-162

Communication

Molecules and Cells

Intragenomic Length Variation of the Ribosomal DNA Intergenic Spacer in a Malaria Vector, *Anopheles sinensis*

In-Jung Whang, Jongwoo Jung, Joong-Ki Park, Gi-Sik Min[1], and Won Kim*

School of Biological Sciences, Seoul National University, Seoul 151-742, Korea;
[1]Department of Biological Sciences, Inha University, Incheon 402-751, Korea.

(Received April 9, 2002; Accepted June 1, 2002)

We determined the complete sequences of six size variants of intergenic spacer (IGS) region from one individual of the malaria vector mosquito species, *Anopheles sinensis*. All six size variants observed in this study show almost the same basic primary structure in which three repeat regions (A, B, and C) are interspersed by highly conserved nonrepeating sections. In contrast to the well-ordered subrepeating patterns found in A and C, the repeat region B displays extremely variable and complicated profiles in the number and arrangement of subrepeat units among different size classes. It is apparent that the prominent level of length difference in the repeat regions B and C is responsible for the intragenomic length variations of the IGS molecule observed in the present study. High level of sequence homology and regularly arranged repeating pattern of 11 to 14 bp motif sequences harbored within the B repeat region allow us to consider that these motif sequences may be associated with their potential role as a recombination site. Compared to those previously published in other mosquito species, the IGS of *A. sinensis* showed a very unique structural format in subrepeat patterns of the IGS region. This result suggests that the structure and sequence profiles of the IGS region would provide useful information for the exploitation of a convenient molecular marker to identify morphologically complicated species complex and to characterize the genetic variation of population. This suggestion is far from being conclusive at present, but a further genetic study will bring more compelling evidences for this pending issue.

Keywords: *Anopheles sinensis*; Intergenic Spacer; Length Variation; Ribosomal DNA.

* To whom correspondence should be addressed.
Tel: 82-2-887-0752; Fax: 82-2-872-1993
E-mail: wonkim@plaza.snu.ac.kr

Introduction

The nuclear ribosomal RNA gene (rDNA) in higher eukaryotes is typically composed of tandem arrays of a basic unit that contain the transcription unit (18S, 5.8S, 28S) and an intervening intergenic spacer (IGS) region. The tandem arrays of the rDNA gene cluster (often called 'multigene family') evolve not independently, but in a concerted fashion (Arnheim, 1983) that confers a high degree of homogeneity among their basic components. Due to its universal presence and abundance (hundreds to several thousands of copies per cell), and different rate of sequence evolution among different regions of the rDNA repeat unit, rDNA has been extensively employed as a powerful genetic marker for accessing phylogenetic relationships of diverse biotic taxa (Crease and Lynch, 1991; Hillis and Dixon, 1991; Hwang and Kim, 1998).

Although the core domain of the rDNA coding regions is highly conserved, IGS region intervening between the 28S and 18S rRNA subunits often displays high degree of sequence divergence and length variation among members of closely related species (Doyle and Beachy, 1985; Saghai-Maroof *et al.*, 1984), or among local populations of the same species (Black *et al.*, 1989; de Merida *et al.*, 1995). Much of the works published to date has consistently confirmed that these length variations observed in IGS region are ascribed to differences in the pattern and number of tandemly arrayed subrepeat units (Williams *et al.*, 1987). Recent molecular surveys for IGS sequence diversity have reported a noticeable length polymorphism among individuals of plant populations (Chou and Tsai, 1999; Rogers and Bendich, 1987; Taira *et al.*, 1988), and particularly, in some cases, have found the existence of different size classes of the IGS within a single individual of freshwater cladoceran (Crease and Lynch, 1991), of salmon (Reed *et al.*, 2000) and of some mosquito species

Abbreviation: IGS, intergenic spacer.

(Beach *et al.*, 1989; McLain *et al.*, 1989). In contrast to these cases, however, little or no length heterogeneity of the IGS regions has been recorded from diverse animal groups [sea urchin (Passananti *et al.*, 1983); insects (Manning *et al.*, 1978; Renkawitz *et al.*, 1979; Schmidt *et al.*, 1982)]. For this reason, species-specific repeat pattern and length variation of IGS region have been widely used in order either to characterize the genetic structure of population, or to identify morphologically indistinguishable cryptic species, including malaria-carrying mosquito species (Black *et al.*, 1989; Collins and Paskewitz, 1996; de Merida *et al.*, 1995; Pepera *et al.*, 1998; Scott *et al.*, 1993).

Re-emerging of malaria is one of the most intensive public health issues in many countries including India, China, and Korea (Chadee *et al.*, 1992; Kho *et al.*, 1999; Sharma, 1996; Sleigh *et al.*, 1998). Due to the fact that *Anopheles sinensis* is one of the major vectors of malaria in Asian countries and considered to be dominant anopheline mosquito species in Korea (Chai, 1999), the availability of a convenient molecular marker which can characterize the genetic structure of vector species would be a prerequisite to more efficient control over the transmission of malaria. In the present study we determined the complete sequences of the ribosomal DNA intergenic spacer (IGS) and firstly report the intragenomic length variation, as a candidate molecular marker, in *Anopheles sinensis.*

Materials and Methods

Mosquito sampling and total genomic DNA isolation Mosquito samples were collected from Jeongeup of Jeollabuk-do Province (35°55′N, 126°87′E) in Korea and identification of specimens was made using taxonomic key published by Ree (2000). Total genomic DNA was prepared using Qiagen DNeasy Tissue kits for the Long PCR amplification.

Long PCR amplification, cloning and sequencing Prior to performing the Long PCR, two universal primers (IJ-28S,5′-GGG AAC GTG AGC TGG GTT TAG ACC GTC-3′; IJ-18S, 5′-CCT GCT GCC TTC CTT GGA TGT GGT AGC C-3′) were designed by referencing the conserved regions of the 18S and 28S rDNA sequences obtained from various eukaryotes. The complete IGS region was amplified by using the Expand Long™ Template PCR System (Roche Co.) following the manufacturer's recommendation: 1 cycle of initial denaturation (2 min at 92°C), 10 cycles [10 s at 92°C, 30 s at 65°C, 8 min (+ 20 s for each cycle) at 68°C], and 1 cycle of prolonged elongation (7 min at 68°C). The amplified PCR products were isolated on 0.7% agarose gel, excised under long-wavelength UV light and extracted using a QIAquick PCR purification kit (Qiagen Co.). The ligates of the purified PCR products were prepared with T4 ligase using pGEM®-T easy vector (Promega Co.) and transformed to DH5α cell line. Plasmid DNA was purified using a QIAprep spin miniprep kit (Qiagen Co.). Subcloning was performed to isolate the size variants of the IGS region. Cyclic sequencing was done for both strands of each size variant with a Big-Dye Terminator sequencing kit (Applied Biosystems, Co.) and reaction products were electrophorased on an ABI 310 automated DNA sequencer. The sequencing primers used in the present study are listed in Table 1.

Table 1. The primers used for sequencing reaction in this study.

ASI-28Sa	5′-GCG TCT CTA CTG TAT TAT TGG-3′
ASI-28Sc	5′-CGC AAG CTC AAA GGT CAT G-3′
ASI-28Se	5′-ACC TAC TGA ATC TAC TTG GC-3′
ASI-18Sa	5′-GTA ACT ATG TTG ATG ATC TCG-3′
ASI-18Sb	5′-CCA TTC GAT ACC GTC AAG CG-3′
ASI-18Sd	5′-CAA GCA CGG ATA CTA CAC GC-3′
ASI-18Se	5′-GTC AAT CAC TTG GAT GTG AAA-3′
ASI-18Sdr	5′-GCG TGT AGT ATC CGT GCT TG-3′
ASI-28Sa	5′-GCG TCT CTA CTG TAT TAT TGG-3′
ASI-28Sc	5′-CGC AAG CTC AAA GGT CAT G-3′
ASI-28Se	5′-ACC TAC TGA ATC TAC TTG GC-3′
ASI-18Sa	5′-GTA ACT ATG TTG ATG ATC TCG-3′
ASI-18Sb	5′-CCA TTC GAT ACC GTC AAG CG-3′
ASI-18Sd	5′-CAA GCA CGG ATA CTA CAC GC-3′
ASI-18Se	5′-GTC AAT CAC TTG GAT GTG AAA-3′
ASI-18Sdr	5′-GCG TGT AGT ATC CGT GCT TG-3′

Data analyses Sequence analyses were performed using the Sequence Navigater (Applied Biosystems, Co.) and Gene Jockey II v. 1.6 (Biosoft, Co.). The subrepeat sequences were aligned by using the Clustal X multiple alignment program (Thompson *et al.*, 1997).

Results and Discussion

In the present study, we firstly report details of structural features of the rDNA IGS region by cloning and directly sequencing 6 size variants detected in malaria vector species, *Anopheles sinensis*. From cloning the long PCR products of the IGS region of a single individual, we obtained fourteen colonies. Using agarose gel electrophoresis, a total of six size variants (clones) are selected (Fig. 1). Direct sequencing of each clone confirmed that all of these represent six different size classes of the IGS, and the complete IGS sequences of each size variant (Clone I-VI; GenBank accession numbers AF498237–AF498242) are ranged from 3,255 to 4,348 bp in length with GC content of 49.9%. Six size variants observed show almost the same basic primary structure in which nonrepetitive and repetitive regions are arranged in an alternative way over the whole length of IGS region (Fig. 1): repetitive region is composed of 3 sections (repeat region A, B, and C, in the 5′ to 3′ direction), accounting 46.9 to 60.1% of the

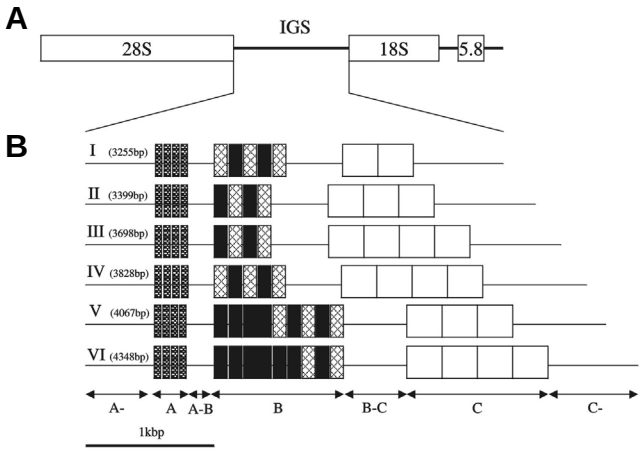

Fig. 1. Comparison of the IGS structural organization of six size variants from one individual. **A.** Schematic diagram of ribosomal DNA unit. **B.** Subrepeat pattern of six size variants of IGS. The repeat region A is indicated by shaded box. Cross-hatching boxes and filled boxes represent b and b′ subrepeat units, respectively. Subrepeats in the repeat region C are marked by blank boxes. Scale bar (1 kbp length) is provided at the bottom.

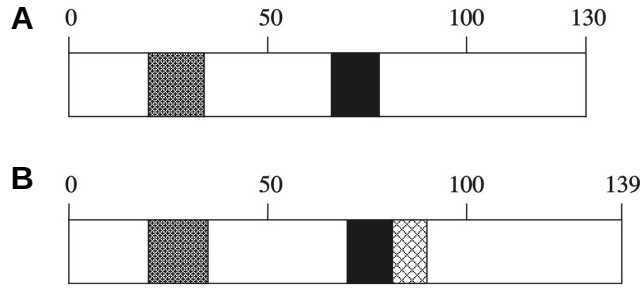

Fig. 2. Schematic representations of the subrepeats b (A) and b′ (B) in the repeat region B. The highly conserved 14, 11, and 10 bp motif sequences are represented by the shaded, filled, and cross-hatched boxes repectively. Numerals mark base position from the 5′ end.

total IGS sequences, that are interspersed by different size of nonrepeating islands designated here as A-, A-B, B-C, and C$^-$ (5′ to 3′ direction), which are approximately 505, 109, 404, and 707 bp, respectively, in length.

In the repeat region A, four different-sized subrepeating units (66, 65, 78, and 79 bp) show a 'head-to-tail' arrangement shared among all six size variants of IGS detected in this study. It is interesting to note that there is 100% sequence homology among the paralleled subunits of six size classes of the IGS clones. In the repeat region C, the 286-bp subrepeat units are tandemly arrayed in a similar fashion of the arrangement to that of the repeat region A, but the number of subrepeat unit varies among the size variants. For example, the clone I and V contain two and three units respectively, whereas each of the other clones (II, III, IV, and VI) is fixed to have four. In contrast to the well-ordered subreating patterns found in the repeat regions A and C, the repeat region B displays extremely variable and complicated format in the number and arrangement of subrepeating units among different size classes: repeat region B is composed of two basic subunits (b, b′) that are about 130 and 139 bp, respectively, in length and the two subunits differ from each other in the number of highly conserved sequence motives. The 14-bp sequence motif (5′-AACGCTTAC-ACTTG-3′; at 21st bp position from the 5′ end) and 11-bp motif (5′-CTTGGTCTGCA-3′; at the middle of the repeat) are shared by b and b′ (Fig. 2), but the latter has an additional 10-bp motif (5′-AGTACACGGT-3′) following the 11-bp motif sequence. Although the b and b′ subrepeat units detected from clones I-IV are alternatively arranged, the clones V and VI comprise a mixture of different sized subunit elements which increased size variation in the repeat region B among six IGS size variants. For example, the third b subrepeat of clone V and VI has a prolonged insertion block (approx. 130 bp) at its 3′ terminal end. Along with the difference in the number of subrepeat units, the size difference resulted from the presence of long-sized insertion block in the repeat region B was substantial (a maximum of 669 bp size difference) between either of clone V or VI and the other clones (Clones I–IV). Considering that the nonrepeating islands intervening three repeating regions show highly conserved in the sequence and length, the prominent level of length difference in repeat regions B and C is responsible for the intragenomic length variation of the IGS molecule observed in the present study.

The IGS region has been well known as a potential 'hot spot' for recombination such as, unequal crossing-over and gene conversion that may accelerate the change of copy number among tandemly repeated rDNA units (Rogers *et al.*, 1986; 1987). Compared to A and C regions, the B repeat region showed substantial variation in the order and number of variably sized subrepeating sequences. Despite the noticeable changes in their structure, the repeat region B contains two highly conserved motif sequences (14 and 11 bp) that were harbored within their basic subunits (b, b′). Note that these motif sequences occur in every B repeat region, irrespective of remarkable size variations among six IGS clones encountered in this study. High level of sequence homology and regularly arranged repeating pattern of these motif sequences allow us to consider that these motif sequences may be associated with their potential role as a recombination site. Our prediction is supported by the existence of traces for variable size of truncated and/or inserted blocks with a relatively low level of sequence homogeneity depending on six size variants. It is assumed that multiple recombination events within the B repeat may play an essential role in accelerating the drastic changes in the copy number

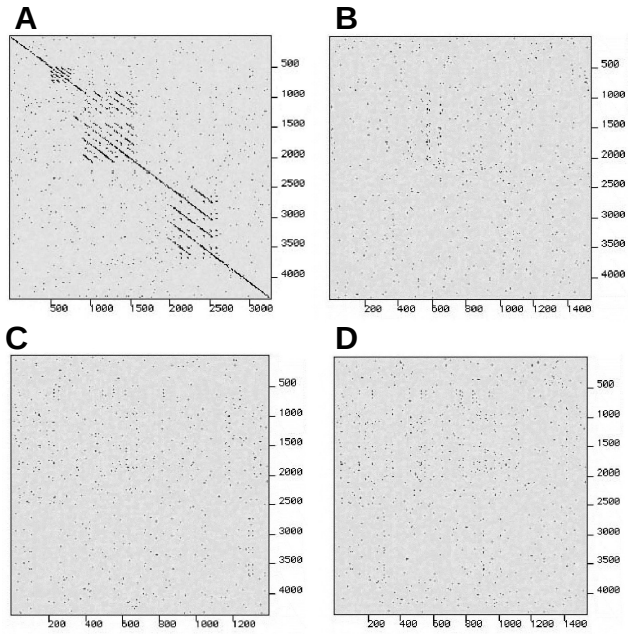

Fig. 3. Dot matrix profiles of IGS sequence of *Anopheles* species. The dot matrix was generated using the GeneJockey II program. Y axis; *A. sinensis* VI. X axis; *A. sinensis* I (A), *A. albimanus* (B), *A. gambiae* (C), *A. quadriannulatus* (D). Different dot colors represent a varying degree of sequence identity among all comparisons of 10 bp sequence unit: match threshold was set by default options (black, 10 matches; blue, 9; red, 8; yellow, 7; sky blue, background).

and subrepeat rearrangement among the six different size classes of the IGS detected in this study.

During the last decade, the IGS region has widely been used, as one of the most powerful molecular marker, to disclose morphologically indistinguishable cryptic species and to characterize the genetic structure of population, including malaria-carrying mosquito species (Black *et al.*, 1989; Collins and Paskewitz, 1996; de Merida *et al.*, 1995; Pepera *et al.*, 1998; Scott *et al.*, 1993). When compared to those previously published in other mosquito species [*A. gambiae*, *A. arabiensis*, *A. merus*, *A. melas* (Scott *et al.*, 1993); *Aedes aegypti* (Wu and Fallon, 1998)] by employing the dot matrix analyses, the IGS of *A. sinensis* presented in this study showed a very unique structural format in subrepeat patterns of the IGS region: An extremely low level of inter-specific sequence homology and species-specific subrepeat pattern were discovered among all pairs of the sequences compared (see Fig. 3). This result suggests that the structure and sequence profiles of the IGS region would provide useful information for the exploitation of a convenient molecular marker to identify morphologically complicated species complex and to characterize the genetic variation of population. This suggestion is far from being conclusive at present, but a further genetic study will bring more compelling evidences for this pending issue.

Acknowledgments We thank two anonymous reviewers for their constructive suggestions on the manuscript. This work was supported by the research grant from the Korea Research Foundation (KRF-2000-015-DP0350) and Brain Korea 21 program (School of Biological Sciences, Seoul National University, Korea).

References

Arnheim, N. (1983) Concerted evolution of multigene families; in *Genes and Evolution of Genes and Proteins*, Nei, M. and Koehn, R. K. (eds.), pp. 36–61, Sinauer, Sunderland.

Beach, R. F., Mills, D., and Collins, F. H. (1989) Structure of ribosomal DNA in *Anopheles albimanus* (Diptera: Culicidae). *Ann. Entomol. Soc. Am*. **81**, 641–648.

Black, W. C. IV, McLain, D. K., and Rai, K. S. (1989) Patterns of variation in the rDNA cistron within and among world populations of a mosquito, *Aedes albopictus* (Skuse). *Genetics* **121**, 539–550.

Chadee, D. D., Maitre, A. L., and Tilluckdharry, C. C. (1992) An outbreak of *Plasmodium vivax* malaria in Trinidad, W. I.. *Ann. Trop. Med. Parasitol.* **86**, 583–590.

Chai, J.-Y. (1999) Re-emerging *Plasmodium vivax* malaria in the Republic of Korea. *Korean J. Parasitol*. **37**, 129–143.

Chou, C. H. and Tsai, C. C. (1999) Genetic variation in the intergenic spacer of ribosomal DNA of *Imperata cylindrical* (L.) Beauv. *Var*. major (Cogongrass) populations in Taiwan. *Bot. Bull. Acad. Sin.* **40**, 319–327.

Collins, F. H. and Paskewitz, S. M. (1996) A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic *Anopheles* species. *Insect Mol. Biol.* **5**, 1–9.

Crease, T. J. and Lynch, M. (1991) Ribosomal DNA variation in *Daphnia pulex*. *Mol. Biol. Evol.* **8**, 620–640.

de Merida, A. M. P., de Mata, M. P., Molina, E., Porter, C. H., and Black, W. C. IV (1995) Variation in ribosomal DNA intergenic spacers among populations of *Anopheles albima-nus* in South and Central America. *Am. J. Trop. Med. Hyg*. **53**, 469–477.

Doyle, J. J. and Beachy, R. N. (1985) Ribosomal gene variation in soybean (*Glycine*) and its relatives. *Theor. Appl. Genet.* **70**, 369–376.

Hillis, D. M. and Dixon, M. T. (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q. Rev. Biol.* **66**, 411–453.

Hwang, S.-K. and Kim, J.-G. (1998) Small-subunit ribosomal DNA of an ectomycorrhizal fungus *Tricholoma matsutake*: sequence, structure, and phylogenetic analysis. *Mol. Cells* **8**, 251–258.

Kho, W. -G., Jang, J.-Y., Hong, S.-T., Lee, H.-W., Lee, W.-J., and Lee, J.-S. (1999) Border malaria characters of reemerging vivax malaria in the Republic of Korea. *Korean J. Parasitol*. **37**, 71–76.

Manning, R. F., Samols, D. R., and Gage, L. P. (1978) The genes for 18S, 5.8S and 28S ribosomal RNA of *Bombyx mori*

are organized into tandem repeats of uniform length. *Gene* **4**, 153–166.

McLain, D. K., Collins, F. H., Brandling-Bennett, A. D., and Were, J. B. O. (1989) Microgeographic variation in rDNA intergenic spacers of *Anopheles gambiae* in Western Kenya. *Heredity* **62**, 257–264.

Passananti, C., Felsani, A., Giordano, R., Metafora, S., and Spadafora, C. (1983) Cloning and characterization of the ribosomal genes of the sea-urchin *Paracentrotus lividus*. *Eur. J. Biochem*. **13**, 233–239.

Pepera, O. P., Cockburn, A. F., Mitchell, S. E., Conn, J., and Seawright, J. A. (1998) Species-specific repeat units in the intergenic spacer of the ribosomal RNA cistron of *Anopheles aquasalis* Curry. *Am. J. Trop. Med. Hyg*. **59**, 673–678.

Ree, H. I. (2000) Manual on medical entomology, pp. 22–39, Institute of Tropical Medicine, Yonsei University, Korea.

Reed, K. M., Hackett, J. D., and Phillips, R. B. (2000) Comparative analysis of intra-individual and inter-species DNA sequence variation in salmonid ribosomal DNA cistrons. *Gene* **249**, 115–125.

Renkawitz, R., Gerbi, S. A., and Glatzer, K. H. (1979) Ribosomal DNA of the fly *Sciara coprophila* has a very small and homogeneous repeat unit. *Mol. Gen. Genet.* **173**, 1–13.

Rogers, S. O. and Bendich, A. J. (1987) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* **9**, 509–520.

Rogers, S. O., Honda, S., and Bendich, A. J. (1986) Variation in the ribosomal RNA genes among individuals of *Vicia faba*. *Plant Mol. Biol*. **6**, 339–345.

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., and Allard, R. W. (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA* **81**, 8014–8018.

Schmidt, E. R., Godwin, E. A., Keyl, H.-G., and Israelewski, N. (1982) Cloning and analysis of ribosomal DNA of *Chironomus thummipiger* and *Chironomus thummi thummi*: the nontranscribed spacer of *Chironomus thummi thummi* contains a highly repetitive DNA sequence. *Chromosoma* **87**, 389–407.

Scott, J. A., Brogdon, W. G., and Collins, F. H. (1993) Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg*. **49**, 520–529.

Sharma, V. P. (1996) Re-emergence of malaria in India. *Indian J. Med. Res.* **103**, 26–45.

Sleigh, A. C., Liu, X. L., Jackson, S., Li, P., and Shang, L. Y. (1998) Resurgence of vivax malaria in Henan province, China. *Bull. W.H.O.* **76**, 265–270.

Taira, T., Kato, A., and Tanifuji, S. (1988) Difference between two major size classes of carrot rDNA repeating unit is due to reiteration of sequences of about 460 bp in the large spacer. *Mol. Gen. Genet*. **213**, 170–174.

Thompson, J. D., Gibson, T. J., Plewniak, D. E., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL X-windows interface: flexible strategies for multiple sequence alignment aided by quality analysis. *Nucleic Acids Res*. **25**, 4876–4882.

Williams, S. M., Furnier, G. R., Fuog, E., and Strobeck, C. (1987) Evolution of the ribosomal DNA spacers of *Drosophila melanogaster*: different patterns of variations on X and Y chromosomes. *Genetics* **116**, 225–232.

Wu, C. C. N. and Fallon, A. M. (1998) Analysis of a ribosomal DNA intergenic spacer region from the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol*. **7**, 19–29.