Insect Molecular Biology (2003) **12**(3), 211–216

Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex

J. Biedler,* Y. Qi,* D. Holligan,† A. della Torre,‡ S. Wessler† and Z. Tu*

**Department of Biochemistry, Virginia Polytechnic Institute and State University,* †*Department of Plant Sciences, University of Georgia,* ‡*Dipartimento di Scienze di Sanita Pubblica, University of Rome 'La Sapienza', Italy*

Abstract

Transposable element (TE) display was shown to be a highly specific and reproducible method of detecting the insertion sites of TEs in individuals of the African malaria mosquito, *Anopheles gambiae*, and its sibling species, *A. arabiensis*. Relatively high levels of insertion polymorphism were observed during the TE display of several families of miniature inverted-repeat TEs (MITEs) that have variable copy numbers. The genomic locations of selected insertion sites were identified by matching the sequences of their corresponding bands in a TE display gel to specific regions of the draft *A. gambiae* genome assembly. We discuss different scenarios in which TE display will provide powerful dominant and co-dominant genetic markers to study the behaviour of TEs in *A. gambiae* populations and to illustrate the complex population genetics of this intriguing disease vector. We suggest that TE display can also provide tools for a phylogenetic analysis of the *A. gambiae* complex.

Keywords: MITEs, interspersed repeats, genome, transposon, disease vector.

Introduction

The *Anopheles gambiae* species complex consists of seven cryptic species (Coluzzi *et al*., 2002; Krzywinski & Besansky, 2003) that include *Anopheles gambiae sensu stricto* (referred to as *A. gambiae* herein) and *Anopheles arabiensis*, two of the most important vectors of malaria, a disease that is responsible for more than a million deaths every year. Vector control, a major component of malaria control strategies, is hampered by increasing insecticide resistance and the genetic heterogeneity of the vector complex. One novel approach which is being actively pursued aims to replace vector mosquitoes in wild populations with genetically modified mosquitoes that are incompetent disease vectors. We are interested in molecular and genetic analyses of mosquito transposable elements (TEs), which may contribute to a genetic strategy for controlling mosquito-borne diseases by providing transformation tools, gene-driving mechanisms and genetic markers for mapping and population studies. TEs are mobile genetic elements that have the ability to replicate and spread in the genome, through either DNA-mediated or RNA-mediated transposition (Finnegan, 1992). We have previously described eight families of miniature inverted-repeat TEs (MITEs) in *A. gambiae* (Tu, 2001). These MITEs share many common structural characteristics with other MITEs, including terminal inverted-repeats (TIRs), small size, no coding potential, and AT richness. They are believed to have used the transposition machinery of autonomous DNA-mediated TEs, taking advantage of shared TIRs (Tu, 2001). MITEs are interspersed in the *A. gambiae* genome, ranging from 40 to 1340 copies per family. Here we report the development of TE display, a genome-wide detection method for TE insertions (e.g. Van den Broeck *et al*., 1998; Casa *et al*., 2000), for several families of MITEs in *A. gambiae* and its sibling species *A. arabiensis*. More importantly, TE display revealed high levels of MITE insertion polymorphism in *A. gambiae*, which suggests that TE display can be used as powerful tools for studying the population genetics of this important vector. We have also discussed the potential use for TE display in phylogenetic analyses of the *A. gambiae* complex and in studies of TE behaviour in mosquito populations.

Received 26 September 2002; accepted after revision 17 December 2002.
Correspondence: Dr Zhijian Tu, Department of Biochemistry, Virginia Tech, Blacksburg, VA 24061, USA. Tel.: +1 540 231 8062; Fax: +1 540 231 9070; e-mail: jaketu@vt.edu

Results and discussion

Pegasus *TE display in* A. gambiae*: specificity and reproducibility*

Pegasus was chosen for the initial development of TE display in *A. gambiae* because it is relatively low in copy

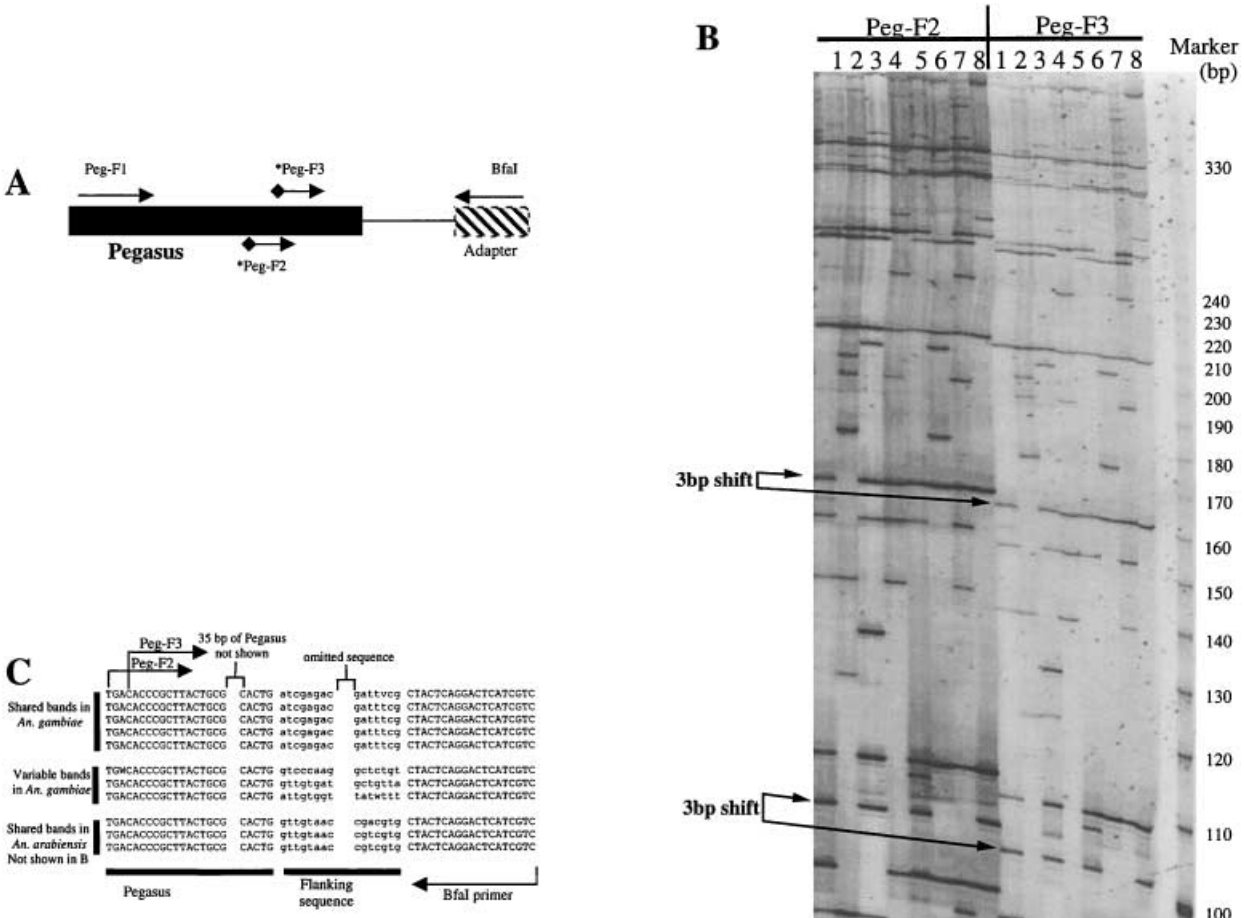

Figure 1. TE display of *Pegasus* elements in *Anopheles gambiae*. (A) A diagram showing adapter-ligated genomic DNA restriction products with TE-specific and adapter specific primers. Peg-F1 (Pegasus-F1) was used for the first round PCR. Either Peg-F2 (Pegasus-F2) or Peg-F3 (Pegasus-F3) was used as the radio-labelled primer for the second round nested PCR. (B) Partial image of a TE display using primers for the *Pegasus* element with eight female individuals from an *A. gambiae* colony (GAMCAM) originally collected from Cameroon. The eight samples on the left were amplified with Peg-F2. The eight samples on the right were the same as those on the left, except they were amplified with primer Peg-F3, which was designed to amplify a product smaller by three bases. The three base shift is clearly observable. A size marker is shown on the right. (C) Sequence analysis of bands from a TE display gel, which were re-amplified and sequenced. Five co-migrating bands from different *A. gambiae* individuals represented a common *Pegasus* insertion site as indicated by the same flanking sequence, which was mapped to a unique fragment in the *A. gambiae* genome (AAAB01008815, positions 68954–69092 on the minus strand). Three variable bands represented insertions at different loci, as indicated by variable flanking genomic sequences. The flanking sequences 'gtccaag …' and 'attgtggt …' were mapped to two repetitive regions while 'gttgtgat …' was mapped to a short unique fragment in the *A. gambiae* genome (AAAB01008956, positions 193489–193514 on the minus strand). The specificity of the *Pegasus* TE display was also demonstrated in *A. arabiensis* on the basis of sequencing of TE display bands, although the image is not shown here. The sequenced insertion site in *A. arabiensis* can also be mapped to a unique fragment in the *A. gambiae* genome with 98% identity (AAAB01008888, positions 1044407–10444515 on the minus strand).

number (34–90 copies; Besansky *et al.*, 1996; Tu, 2001). In addition, *Pegasus* insertion site polymorphism has been previously demonstrated using *in situ* hybridization in a highly inbred *A. gambiae* strain named PEST (Mukabayire & Besansky, 1996). Here we used the GAMCAM strain of *A. gambiae* to develop TE display. Briefly, ≈ 50 ng of DNA was digested using *Bfa*I. The digested DNA fragments were ligated to a *Bfa*I adapter. A pre-amplification and selective amplification were used to amplify the fragments between specific *Pegasus* sequences and the *Bfa*I adapter sequence (see Experimental procedures and Fig. 1A). A radio-labelled primer was used in the selective amplification. The amplified fragments were separated on a sequencing gel and the banding patterns of each individual were analysed. Two methods were used to demonstrate the specificity of the TE display. In selective amplification, the use of the primer Peg-F3 resulted in a 3 bp shift in all bands that were produced with primer Peg-F2 (Fig. 1B). Because the labelled primers (Peg-F2 and Peg-F3) were 3 bp apart in the *Pegasus* sequence, the faithful 3 bp shift in all of the TE display bands indicated a high specificity for the technique. Furthermore, bands from a TE display gel were

re-amplified and sequenced, showing that they contained *Pegasus* sequences as well as flanking genomic and adapter sequences in the expected order (Fig. 1C). Co-migrating bands among different individuals had the same flanking genomic sequence, indicating that they were from the same genomic locus. Bands that were of different sizes had different flanking genomic sequences (Fig. 1C). All of these flanking sequences were mapped to the *A. gambiae* genome assembly and their specific locations are described in the legend of Fig. 1. It should be noted that none of the four mapped loci had a *Pegasus* insertion in the genome assembly, which was derived from the highly inbred PEST strain. It will be interesting to determine whether there is a significant difference in the insertion patterns of *Pegasus* between the PEST strain and the GAMCAM colony that was used in the current study. Finally, to test the reproducibility of this procedure, the TE display was repeated using the same samples which produced the same banding pattern (data not shown).

Display of TEs of variable copy numbers in A. gambiae

TE display was performed with additional TE families that have different copy numbers, including *TAA-II-Ag* (320 copies), *TA-III-Ag* (970 copies), and *TA-Iα-Ag* (1340 copies), all of which are recently described MITEs (Tu, 2001). To reduce the number of bands on the TE display gel (Casa *et al*., 2000), it was necessary to arbitrarily add a selective base (T) to the 3′ end of the adapter primer (Fig. 1A, *Bfa*I primer) in the cases of *TA-III-Ag* (Fig. 2) and *TA-Iα-Ag* (data not shown). For *TAA-II-Ag,* no selective base was necessary (Fig. 3). We conclude that the TE display technique can be adapted for different TE families in *A. gambiae*.

TE display revealed relatively high levels of MITE insertion polymorphism in A. gambiae

Relatively high levels of insertion polymorphism were seen for three *A. gambiae* MITEs (Figs 1–3). On the basis of the TE display of eight females in the GAMCAM colony that originated in the Cameroon, the relative percentages of polymorphic sites were estimated to be 90%, 72% and 59% for *Pegasus*, *TA-III-Ag* and *TAA-II-Ag*, respectively. A polymorphic site is defined here as a band that is not shared in all individuals. Heterozygous and homozygous TE insertions are not distinguished. The TE display shown here is designed to maximize the detection of TE insertion polymorphism by using primers close to the termini and by using touch-down PCR, both of which reduce the amplification of degenerate copies, which may be fixed in the population. This is consistent with the fact that all the sequenced bands contained the intact 3′ terminus (Fig. 1C). The above calculations are therefore not absolute measurements of the level of MITE insertion polymorphism. However, the percentages are relatively high, indicating that at least significant fractions of the insertion sites of the full-length MITEs are polymorphic.

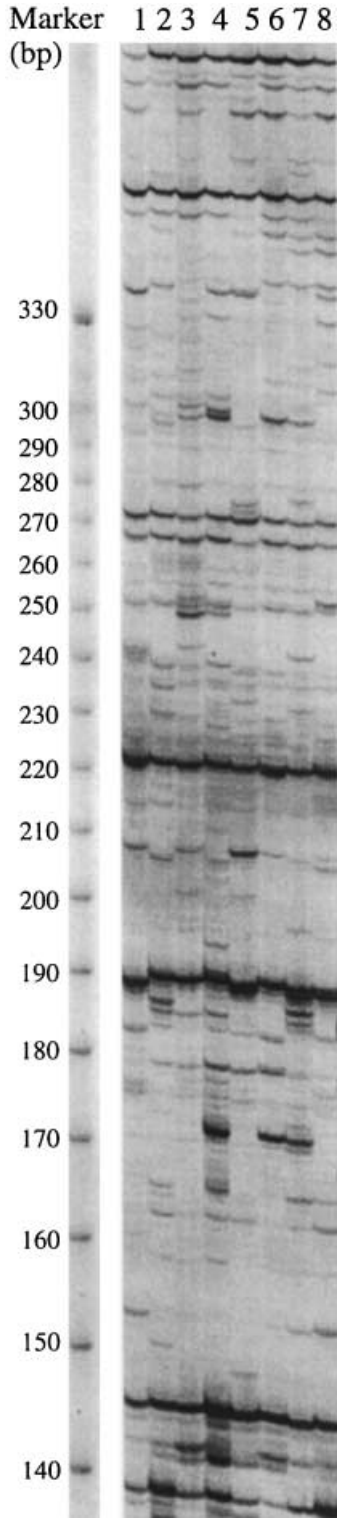

Figure 2. TE display of the *TA-III-Ag* elements in *Anopheles gambiae*. Shown here is a partial image of a TE display with eight female individuals from an *A. gambiae* colony GAMCAM. The primers for the first round PCR are TA-III-Ag-F1 and *Bfa*I + 0. The primers for the second round PCR are TA-III-Ag-F2 and *Bfa*I + T. The extra base added to the adaptor primer is designed to reduce the number of potential PCR bands because *TA-III-Ag* is highly repetitive. A size marker is shown on the left.

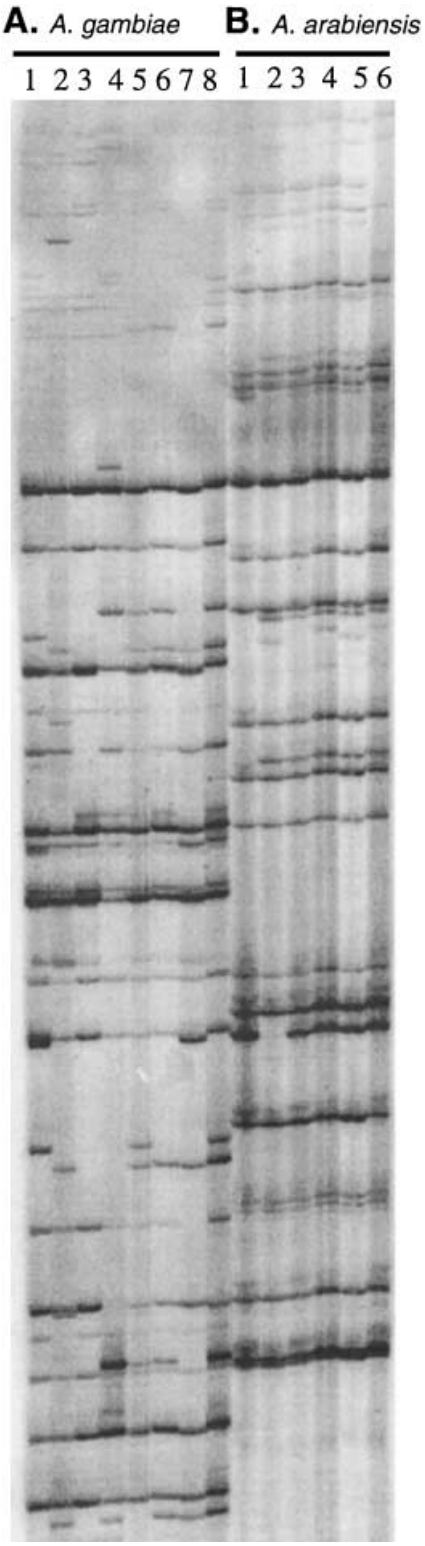

Figure 3. TE display of the *TAA-II-Ag* elements in *Anopheles gambiae* and *A. arabiensis*. Shown here is a partial image of a TE display with eight female individuals from an *A. gambiae* colony GAMCAM, and six female individuals from an *A. arabiensis* colony ARMOR which was originally collected from Mali. The primers for the first round PCR are TAA-II-Ag-F1 and *Bfa*l + 0. The primers for the second round PCR are TAA-II-Ag-F2 and *Bfa*l + 0.

Similar levels of polymorphism and banding patterns were observed in males and in a second colony GASPIN that originated in Mali, although the relative frequencies at different loci were different between the GAMCAM and GASPIN colonies (data not shown).

TE display and comparison of insertion patterns between A. gambiae *and* A. arabiensis

When female individuals of *A. gambiae* and *A. arabiensis* were compared using TE displays for *Pegasus*, *TA-Iα-Ag*, *TA-III-Ag* and *TAA-II-Ag*, obvious differences were observed. As shown in Fig. 3, although some bands were shared between the two sibling species, many bands were unique. It is not yet clear whether this difference in TE display banding patterns was mainly due to insertion site variation or to a general sequence difference between the two species that resulted in changes in the *Bfa*I sites near TEs. The specificity of *A. arabiensis* TE display was confirmed by cloning and sequencing several bands from a *Pegasus* TE display gel (Fig. 1C).

Potential applications

The development of a specific and reproducible TE display method and the demonstration of relatively high levels of MITE insertion polymorphism in *A. gambiae* provide exciting opportunities for developing TE-based population genomic tools and to study the behaviour of TEs in populations of the *A. gambiae* complex. One major challenge in establishing sophisticated vector control programmes and meaningful epidemiological studies has been the genetic complexities of *A. gambiae* populations (Powell *et al.*, 1999). Several recent studies using a number of genetic markers have made significant progress towards illustrating this complexity, while pointing to the need for more extensive research (e.g. Black & Lanzaro, 2001; della Torre *et al.*, 2001). The development of new population genomic approaches is needed because a conflict exists between results obtained using different types of genetic markers and markers at different genomic locations (Besansky, 1999). Because many MITE families are interspersed throughout the genome, and because relatively high levels of MITE insertion polymorphisms have been demonstrated, MITEs are great sources of polymorphic markers across different regions of the *A. gambiae* genome. For example, TE display may be used to rapidly generate a large number of highly specific and reproducible dominant markers for population studies and for genotyping. In addition, TE display can be used to rapidly isolate and locate polymorphic insertion sites by sequencing the recovered bands (Fig. 1C), and comparing the sequences to the genome assembly (Holt *et al.*, 2002). These polymorphic insertion sites provide co-dominant markers when sequences flanking a TE at the specific locus are used as primers to PCR amplify genomic DNA isolated from an individual sample (e.g. Batzer *et al.*, 1994;

Stoneking *et al*., 1997; Roy-Engel *et al*., 2001). One potential limitation of this method is the possible precise excision of a DNA-mediated TE, that may be confused with the non-insertion state. However, this is not likely to be a major issue because no active DNA transposons have been reported in *A. gambiae*, although there is evidence of excision of an internally deleted DNA-mediated TE named *Ikirara* in an *A. gambiae* cell line, at a rate of 0.038% (Leung & Romans, 1998). In addition to the genetic heterogeneity at the population level, the *A. gambiae* species complex includes seven reproductively isolated cryptic species that were previously identified according to their chromosomal inversion patterns (e.g. Coluzzi *et al*., 1979; Krzywinsky & Besansky, 2003). Apparent differences in banding patterns were observed between *A. gambiae* and *A. arabiensis* in the TE display of a few MITE families including *TAA-II-Ag* (Fig. 3). Comparative TE displays of the seven sibling species may help to resolve their phylogenetic relationship, which has thus far been a very difficult problem, perhaps mainly because of introgression at some loci but not at others (Krzywinsky & Besansky, 2003). In this regard, TE display may be especially helpful because it presumably scans a large number of TE insertions interspersed in the genome. Finally, TE display and the locus-specific PCR methods will also provide useful tools for the study of how TEs are maintained and spread in field populations. Such information will help to evaluate the possibility of using transposons to drive transgenes into wild populations.

Experimental procedures

Mosquitoes

Two laboratory colonies of *A. gambiae*, derived from original collections from Cameroon (colony GAMCAM) and Mali (colony GASPIN), were maintained as previously described (della Torre *et al*., 2001). These two colonies belonged to the M and S molecular forms, respectively (della Torre *et al*., 2001). An *A. arabiensis* colony (ARMOR), originally collected from Mali, was also used.

TE display

Genomic DNA for TE display was isolated from individual mosquitoes using the DNAzol isolation reagent from Molecular Research Center Inc. (Cincinnati, OH). The general procedures for TE display were modified from Casa *et al*. (2000). Digestion of genomic DNA and the ligation of adapters was done in a single step. Approximately 50 ng of DNA was digested/ligated in a 50 μl reaction containing 1× restriction/ligation buffer (10 mM Tris-acetate, 10 mM Mg-Acetate, 50 mM K-Acetate, pH 7.5), 5 units of *Bfa*I, 50 pmol of Bfa-I adapter (5′-GACGATGAGTCCTGAG-3′ and 5′-TACTCAGGACTCAT-3′), 50 ng/μl BSA, 5 mM DTT, 1.2 mM ATP and 1 unit of T4 DNA-ligase. The quality of the digestion/ligation was monitored by running 20 μl of each sample on a 1% agarose gel stained with ethidium bromide. A smear of DNA fragments ranging in size from ≈ 100–1000 bp verified successful digestion. The remainder of the reaction was diluted fourfold with 0.1× TE, and was then subject to pre-selective amplification in which PCR was performed using primers specific for the adapter sequence (*Bfa*I + 0, 5′-GACGATGAGTCCTGAGTAG-3′) and TE of interest (Pegasus-F1 5′-AGTCGTGTCATGGTAACC-3′, TAA-II-Ag-F1 5′-GGATTTCAGCCATACAACC-3′, TA-III-Ag-F1 5′-TGGATATCAATACAATTGGACCAA-3′). Pre-amplification was carried out in 30 μl reactions containing 3 μl of the diluted restriction/ligation reaction, 2.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.3 μM Bfa-I + 0 primer, 0.3 μM TE-specific primer, 2.5 units of Taq-Polymerase and Taq-Polymerase buffer. PCR conditions were as follows: 1 × 72 °C for 2 min, 1 × 94 °C for 3 min, 24 × (94 °C for 30 s, 58 °C for 30 s, 72 °C for 60 s), and 1 × 72 °C for 5 min. PCR amplification was monitored by running 10 μl of each reaction on an agarose gel as described above. The remaining volume was diluted 10-fold with 0.1× TE. A second round selective amplification was then performed in 10 μl reactions. In this case, radio-labelled (γ-^{33}P), TE-specific primers (Pegasus-F2 5′-TGACACCCGCTTACTGCG-3′, Pegasus-F3 5′-CACCCGCTTACTGCGAAT-3′, TAA-II-Ag-F2 5′-ACGCTCTCCCATACAAATC-3′, TA-III-Ag-F2 5′-AATATGGGAACCCACCAAAAATTGA-3′) were used in addition to primers specific for the adapter sequence (*Bfa*I + 0 5′-GACGATGAGTCCTGAGTAG-3′, or *Bfa*I + T 5′-GACGATGAGTCCTGAGTAGT-3′). Labelling of the primer and the 30–330 bp DNA ladder (Gibco/BRL, Rockville, MD) was performed using T4 polynuleotide kinase (Gibco/BRL; Casa *et al.* 2000). A touchdown PCR protocol was used, which consisted of 1 cycle at 94 °C for 5 s, 9 cycles at 94 °C for 30 s, 67–59 °C (decreasing by 1 °C per cycle) for 30 s, and 1 cycle at 72 °C for 60 s. The samples were further amplified for 27 cycles at 94 °C for 30 s, 58 °C for 30 s, 72 °C for 60 s, followed by a 5 s extension at 72 °C. The PCR products were run on a sequencing gel. The gel was lifted to filter paper, dried, and exposed to an X-ray film for 24 h (Casa *et al*., 2000).

Recovery, cloning and sequencing of bands in TE display gels

To recover the bands from a TE display gel, a modified version of a previously described protocol (Stumm *et al*., 1997) was used. A window was cut through the X-ray film to expose the band of interest. The TE display gel was scratched with a fine needle or pipette tip through the window on the film that was aligned with the gel. The scratched gel material was placed in a PCR tube containing 20 μl reaction mix for approximately 1 min. PCR amplification then proceeded as in selective amplification. PCR products were separated on 1–1.5% agarose gels, excised and purified using a Sephaglas Bandprep Kit from Amersham Pharmacia Biotech (Arlington Heights, IL). Purified PCR products were cloned in a pCR 2.1 vector using an Original TA Cloning Kit from Invitrogen (Carlsbad, CA). Cloned PCR products were sequenced with an IRD800 dye-labelled T7 primer using a 4200S Gene ReadIR sequencing instrument from Li-Cor (Lincoln, NE). Searches for matches of nucleotide sequences in the database were done using BLAST (Altschul *et al*., 1997). Multiple sequences were aligned using PILEUP (GCG, Genetics Computer Group, Madison, WI, version 10, 1999).

Acknowledgements

This work was supported by NIH grant no. AI42421 to Z. Tu and by the Agricultural Experimental Station at Virginia Tech.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* **25**: 3389–3402.

Batzer, M.A., Stoneking, M., Alegria-Hartman, M. *et al.* (1994) African origin of human-specific polymorphic *Alu* insertions. *Proc Natl Acad Sci USA* **91**: 12288–12292.

Besansky, N.J. (1999) Complexities in the analysis of cryptic taxa within the genus *Anopheles*. *Parassitologia* **41**: 97–100.

Besansky, N.J., Mukabayire, O., Bedell, J.A. and Lusz, H. (1996) *Pegasus*, a small terminal inverted repeat transposable element found in the white gene of *Anopheles gambiae*. *Genetica* **98**: 119–129.

Black, W.C. and Lanzaro, G.C. (2001) Distribution of genetic variation among chromosomal forms of *Anopheles gambiae.* s.s.: introgressive hybridization, adaptive inversions, or recent reproductive isolation? *Insect Mol Biol* **10**: 3–7.

Casa, A.M., Brouwer, C., Nagel, A., Wang, L., Zhang, Q., Kresovich, S. and Wessler, S. (2000) The MITE family heartbreaker (Hbr): molecular markers in maize. *Proc Natl Acad Sci USA* **97**: 10083–10089.

Coluzzi, M., Sabatini, A., Petrarca, V. and Di Deco, M.A. (1979) Chromosomal differentiation and adaptation to human environment in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* **73**: 483–497.

Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M.A. and Petrarra, V. (2002) A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**: 1415–1418.

Finnegan, D.J. (1992) Transposable elements. *Curr Opin Genet Dev* **2**: 861–867.

Holt, R.A., Subramanian, G.M., Halpern, A. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Krzywinski, J. and Besansky, N.J. (2003) Molecular systematics of Anopheles: from subgenera to subpopulations. *Annu Rev Entomol* **48**: 111–139.

Leung, S.S. and Romans, P. (1998) Excisions of the *Ikirara1* transposon in an *Anopheles gambiae* cell line. *Insect Mol Biol* **7**: 241–248.

Mukabayire, O. and Besansky, N.J. (1996) Distribution of *T1, Q, Pegasus* and *mariner* transposable elements on the polytene chromosomes of PEST, a standard strain of *Anopheles gambiae*. *Chromosoma* **104**: 585–595.

Powell, J.R., Petrarca, V., della Torre, A., Caccone, A. and Coluzzi, M. (1999) Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia* **41**: 101–113.

Roy-Engel, A.M., Carroll, M.L., Vogel, E., Garber, R.K., Nguyen, S.V., Salem, A.H., Batzer, M.A. and Deininger, P.L. (2001) *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics* **159**: 279–290.

Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., Jenkins, T., Tahir, M.A., Deininger, P.L. and Batzer, M.A. (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* **7**: 1061–1071.

Stumm, G.B., Vedder, H. and Schlegel, J. (1997) A simple method for isolation of PCR fragments from silver-stained polyacrylamide gels by scratching with a fine needle. *Technol Tips Online* **1**: 17: T01115, http://research.bmn.com/tto/

della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V. and Coluzzi, M. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol* **10**: 9–18.

Tu, Z. (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* **98**: 1699–1704.

Van den Broeck, D., Maes, T., Sauer, M., Zethof, J., de Keukeleire, P., D'Hauw, M., Van Motagu, M. and Gerets, T. (1998) Transposon Desplay identifies individual transposable elements in high copy number lines. *Plant J* **13**: 121–129.