Published in final edited form as:
Gene. 2008 October 15; 423(1): 63–71. doi:10.1016/j.gene.2008.06.023.

Topi, an *IS630/Tc1/mariner* – type transposable element in the African malaria mosquito, *Anopheles gambiae*

Ramanand A. Subramanian[a], **Olabiyi O. Akala**[a], **Johnson O. Adejinmi**[b], and **David A. O'Brochta**[a,*]

a*Center for Biosystems Research; University of Maryland Biotechnology Institute; 9600 Gudelsky Drive, Rockville, Maryland, 20850*

b*Department of Veterinary Microbiology and Parasitology; University of Ibadan, Ibadan, Nigeria*

Abstract

IS630/Tc1/mariner elements are diverse and widespread within insects. The African malaria mosquito, *Anopheles gambiae*, contains over 30 families of *IS630/Tc1/mariner* elements although few have been studied in any detail. To examine the history of *Topi* elements in *Anopheles gambiae* populations, *Topi* elements (n = 73) were sampled from five distinct populations of *Anopheles gambiae* from eastern and western Africa and evaluated with respect to copy number, nucleotide diversity and insertion site-occupancy frequency. *Topi 1* and *2* elements were abundant (10–34 per diploid genome) and highly diverse ($\pi = 0.051$). Elements from mosquitoes collected in Nigeria were *Topi 2* elements and those from mosquitoes collected in Mozambique were *Topi 1* elements. Of the 49 *Topi* transposase open reading frames sequenced none were found to be identical. Intact elements with complete transposase open reading frames were common, although based on insertion site -occupancy frequency data it appeared that genetic drift was the major force acting on these *IS630/Tc1/mariner* -type elements. *Topi* 3 elements were not recovered from any of the populations sampled in this study and appear to be rare elements in *Anopheles gambiae*, possibly due to a recent introduction.

Keywords

Topi; transposable elements; *Tc1*; *mariner*; *Anopheles gambiae*; malaria

1. Introduction

IS630/Tc1/mariner elements are a diverse and widespread megafamily of transposable elements within insects (Robertson, 2002). The African malaria mosquito, *Anopheles gambiae*, contains over 30 families of *IS630/Tc1/mariner* elements although few have been studied in any detail (Shao and Tu, 2001; Coy and Tu, 2005). There is a general interest in Class II transposable elements in mosquitoes for a number of reasons. Frist, transposable elements are currently used as mosquito germline transformation vectors yet none of the existing gene vectors perform efficiently, greatly limiting the use of this technology in

*Corresponding author: Center for Biosystems Research, University of Maryland, Biotechnology Institute, 9600 Gudelsky Drive, Rockville, MD 20850. E-mail: obrochta@umbi.umd.edu, Phone: 240-314-6343, Fax: 240-314-6255.

functional genomic studies (O'Brochta and Handler, 2008). Class II transposable elements that are highly active in mosquitoes would complement the efficient tools currently available for obtaining genomic DNA sequences and gene expression data. Second, transposable elements, by virtue of their abilities under some circumstances to rapidly invade host genomes and populations, have been considered as one of a number of options for introducing phenotype-altering genes into mosquito populations for the purposes of altering the intensity of mosquito-borne disease transmission (Aksoy, 2008). Highly active transposable elements with a strong tendency to replicate, increase in copy number and be transmitted to progeny in the next generation might be useful for disease transmission control strategies that aim to manipulate vector populations (Kidwell and Ribeiro, 1992; Atkinson, 2008). Finally, efforts to use transposable elements as gene drive agents will benefit from an understanding of the dynamics of transposable elements in target species. Understanding spatial and temporal patterns of movement of endogenous active elements will aid in predicting the behavior of elements intentionally introduced for the purposes of spreading anti-pathogen transgenes (O'Brochta et al., 2004).

Topi elements were discovered in *An. gambiae* as part of a concerted effort to identify *Tc1*-like transposable elements in this species using PCR-based strategies. Degenerate primers to highly conserved regions of *Tc1*-like elements were used to amplify a number of sequences that proved to be part of three endogenous *Tc1*-like elements, *Tsessebe, Topi* and *Tiang* (Grossman et al., 1999). Because apparently full-length copies of *Tsessebe* and *Topi* were found, in addition to their variable copy number and insertion site polymorphism among strains, they were thought to be transpositionally active transposable elements (Grossman et al., 1999). Coy and Tu (2005) expanded our understanding of *Topi* elements through their bioinformatic analysis of the *An. gambiae* genome (Holt RA et al., 2002) and reported that *Topi* consisted of three lineages (*Topi 1, 2* and *3*). *Topi 1, 2* and *3* were 91% identical and had similar but distinct inverted terminal repeat sequences of 24 bp. Multiple copies of these elements were reported with *Topi 1* comprised of 12 autonomous and 15 non-autonomous elements, *Topi 2* comprised of 10 autonomous and 43 non-autonomous elements and *Topi 3* comprised of one autonomous and 5 non-autonomous elements (Coy and Tu, 2005). Like Grosman et al. (1999), Coy and Tu (2005) suggested that these elements may have been "recently active". The level of element activity will be reflected in the degree of insertion site polymorphism in organisms with relatively large effective population sizes. The objective of this study was to quantify the amount of insertion site polymorphism and to assess copy number and nucleotide polymorphisms for the purposes of assessing the history of *Topi* element activity in *An. gambiae*. Based on this analysis we found no evidence for *Topi 3* elements in any of the insects sampled and only low levels of insertion site polymorphism in five distinct populations of *An. gambiae s.s.* High nucleotide sequence diversity among *Topi 1* and *2* elements in combination with insertion site polymorphism data suggested that *Topi* elements are long–time residents of this species with low levels of activity. Intact elements lacking internal deletions remain common, unlike the *P* element in *Drosophila* that rapidly evolve internally deleted forms as a consequence of element transposition. These data have implications for the use of transposable elements as gene drive agents for moving anti-pathogen trangenes into populations of *An. gambiae*.

2. Materials and Methods

2.1. Samples

Anopheles gambiae s.s. from five populations were used in this study with a sample size of 16 individuals each from Kisumu, Malindi and Zenet,15 from Furvela and 10 individuals from Bakin Kogi populations (Table 1). Samples from Malindi, Bakin Kogi, Zenet and Furvela have been previously described (Subramanian et al., 2007). Malindi is located in eastern Kenya and

was sampled in 1996 (Lehmann et al., 2003). Bakin Kogi is in north-central Nigeria and samples were collected in 1999 (Lehmann et al., 2003). Zenet is a village in northeastern region of Tanzania and was sampled in and around the village in 2004 (Meeraus et al., 2005). Samples from southern Mozambique (Furvela) were collected in 2003 and were earlier described (O'Brochta et al., 2006). Samples from Kisumu were collected in 2005 from two villages Iguhu and Kombewa in Western Kenya. All insects used in this study were S-form *An. gambiae*.

2.2. DNA isolation and whole genome amplification

Genomic DNA was isolated from individual mosquitoes as described (O'Brochta et al., 2006) and resuspended in 100 μl of distilled water and stored at −80°C. One hundredth of the genomic DNA from one mosquito (1 μl) was used in the whole genome amplification using GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Piscataway, NJ) following the manufacturer's recommendations. Amplified genomic DNA was resuspended in 20 μl of distilled water and stored at −80°C.

2.3. *Topi* transposable element display

The procedure used for transposable element display has previously been described (Guimond et al 2003, O'Brochta et al 2006) and was modified for the analysis of *Topi* transposable element as described below. Transposable element display was performed in triplicate using one eighth (2.5 μl) of the DNA obtained after the whole genome amplification of $1/100^{th}$ of the genomic DNA obtained from one mosquito (see below) for each replicate. Genomic DNA was digested for 4 hours in a volume of 20 μl at 37°C with 2 units of the restriction endonuclease *Dpn*II using conditions recommended by the manufacturer (New England Biolabs). *Dpn*II digestion products were ligated to 30 picomoles of adapters by adding 5 μl of 1X restriction enzyme buffer containing 5 mM ATP, 50 mM DTT (dithiothreitol), 10 μg BSA (bovine serum albumin), 4 units of *Dpn*II, 1 Weiss unit of T4 DNA ligase and incubated at 37°C overnight. To prepare the adapters, equimolar amounts of oligonucleotides MspIa (5′ GAC GAT GAG TCC TGA G 3') and DpnIIb (5' GAT CCT CAG GAC TCA TC 3') were heated to 100°C for 10 minutes and then allowed to very slowly cool to room temperature. The conditions used for the digestion/ligation reactions and also the design of the adapters allow the creation of only monomeric *Dpn*II-cut genomic DNA fragments with terminal adapters.

The next step was a polymerase chain reaction ("preselective PCR") with five microliters of the restriction/ligation reaction as the template in a 25 μl reaction volume containing 1X PCR Buffer II (Applied Biosystems), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (Applied Biosystems), and 24 pmoles of primer MspIa and primer TETopiR1 (5' GTT AGA ATG TGT TTT CG C 3'), a primer designed to amplify *Topi 1, 2* and *3* elements. The DNA polymerase was added as a complex with TaqStart™ Antibody (ClonTech) as described by the manufacturer for the purpose of "hot-starting" the reaction. The reaction conditions were 95°C/3 mins followed by 25 cycles of 95°C/15 sec, 54°C/30 sec, 72°C/1.0 min and a final cycle of 72°C/5 min. A second PCR was performed ("selective PCR") using 5 μl of the 20 times diluted preselective PCR products as template in a 20 μl reaction containing 1X PCR Buffer II, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 1 unit AmpliTaq® DNA polymerase (bound to TaqStart™ Antibody as above), 9 pmoles each of primers MspIa and Cy5™-labeled TETopiR2 (5' TAA ACA GTC CTT TTC AGG 3') which is also expected to amplify *Topi 1, 2* and *3* elements. The Cy5™-labeled primers were purified by HPLC prior to their use. Following an initial denaturation step at 95° C for 3 minutes, "touchdown" PCR conditions were used in which during the first 5 cycles the annealing temperature was decreased 1°C after each cycle with the first of these cycles being 95°C/15 sec, 59°C/30 sec, 72°C/1.0 min. Following these 5 cycles an additional 25 cycles were performed at 95°C/15 sec, 54°C/30 sec, 72°C /1.0 min with a final cycle of 72°C/5 min.

TETopiR1 and TETopiR2 are *Topi* element specific primers that anneal to sequences approximately 150 bp and 90 bp from the 3'end of the element.

5 μl of selective PCR products were mixed with 5μl of loading buffer (95% deionized formamide, 10mM EDTA) and the mixture heated to 95°C for 3 minutes, cooled quickly on ice and 6 μl loaded on a 6% polyacrylamide gel (19:1 acrylamide : bisacrylamide) containing 6.7 M urea in 1X TBE buffer (90 mM Tris-borate, 2 mM EDTA). ALFExpress™Sizer™50–500 (Amersham/Pharmacia) was used as a size standard. Electrophoresis was performed for 2.5 hours at a constant voltage of 70 watts. After which time the gel was transferred to 3MM filter paper and dried. The dried gel was scanned on a STORM 860 phosphoimager (Molecular Dynamics) to visualize the products of the transposable element display. The selective PCR products from the three independent replicates of a sample were run on the same gel to assist unambiguous calling of bands. A band was called as present or absent if it was present in at least 2 of the three replicates. From the three replicates, a single scoring matrix was obtained that was used in subsequent analyses. The advantage of this procedure is that it increased the accuracy of determining the presence of bands and minimized errors in subsequent analyses.

Transposable element display data was used to estimate the site-occupancy frequency distribution of *Topi* and by assuming the models of Charlesworth and Charlesworth (1983) these data were used to estimate the parameter β [equation 11, Charlesworth and Charlesworth (1983)]. The model parameter β measures the forces removing the elements from natural populations (drift, excision and selection). Because the model used in this analysis assumes that the copy number is in equilibrium, it also reflects the forces that tend to add elements to the population (replicative transposition). Estimation of β and the copy number of *Herves* per diploid genome were performed as described by Wright *et al*. (2001; equations 3 and 4, respectively) and O'Brochta e*t al*. (2006). Wright *et al*. (2001) considered the dominant nature of transposable element display signals and the application of the parameter estimation methods of Charlesworth and Charlesworth (1983) to diploid organisms. A one way- ANOVA and Tukey's HSD (honestly significant difference) test was used to compare the average diploid copy number among locations for statistical differences between different locations.

2.4. *Topi* transposase detection and sequencing

To analyze the structure and sequence of *Topi* elements, *Topi* transposase open reading frame was amplified using a Topi277F (5'-ATG GGT CGC GGA AAG CAC TG-3') primer that annealed to the 5' end of the open reading frame and a Topi1302R primer (5'- GCG GTG TTC CAC TGA GCG-3') that annealed to the DNA flanking the open reading frame. As with transposable element display primers these were designed to amplify the open reading frames of *Topi 1, 2* and *3* elements (Figure 1). One fiftieth of the genomic DNA from one mosquito (2 μl) was used as template in a 50μl reaction containing 1X ThermalAce™ (Invitrogen), 0.2 mM dNTPs (an equimolar mixture of dATP, dTTP, dCTP, dGTP), 2.5 mM $MgCl_2$, 2 units ThermalAce™ DNA polymerase (Invitrogen), and 24 pmoles of primer Topi277F and Topi1302R. The following conditions were used for the amplification reactions: 95°C/3 min followed by 30 cycles of 95°C/30 sec, 55°C/30 sec, 72°C/1min 30 secs and a final cycle of 72° C/10 min. Reaction products were fractionated in a 1% agarose gel. The 1kb amplification products from all samples and the approximately 600 bp products from 8 samples were eluted from the gel, precipitated, resuspended in 20 μl dH_2O and cloned into the pCR®-Blunt II TOPO vector (Invitrogen). One clone per individual was sequenced (2x coverage) and these sequences were used in subsequent analyses. From samples "Kisumu" (12), "Malindi" (8), "Zenet" (10), "Furvela" (11) and "Bakin Kogi" (8) a total of 49 sequences were obtained.

2.5. Sequence Analysis

Sequences used in this analysis were deposited in the National Center for Biotechnology Information's GenBank. Accession numbers EU284044–EU284092. Sequence alignments were done using AlignX, a ClustalW-base alignment program in VectorNTI Advance 10.0.1 (Invitrogen). Nucleotide diversity was estimated from average pair-wise number of differences between elements, π (Nei and Li, 1979)and from the number of polymorphic sites, θ (Watterson, 1975) π and θ were estimated using DnaSP 3 (Rozas and Rozas, 1995; Rozas et al., 2003). The silent-site diversity estimates were calculated using the Kumar method (Nei and Kumar, 2000) implemented in MEGA 3.1 (Kumar et al., 2004). Expected values of silent-site diversity were calculated described in Sanchez-Gracia et al (2005) (Sanchez-Gracia et al., 2005) and were the product of the haploid copy number and the average synonymous diversity (0.0209) from a sample of 35 nuclear genes (Morlais et al., 2004). The average nucleotide-sequence diversity, π, and, the average expected and observed silent-site diversity estimates were compared among locations using a one way-ANOVA. Post-hoc comparisons were made using Tukey's HSD test, $p < 0.05$ denoted a significant difference. An alignment of 14 sequences that did not have any pre-mature stop codons were used for estimating the number of synonymous substitutions per synonymous site (dS) and of non-synonymous substitutions per non-synonymous site (dN) and their ratio, ω= dN/dS using maximum likelihood (ML) methods employed by CODEML in PAML 3.13 (Yang, 1997). PAML can be used to examine the data using various codon substitution models that make different assumptions about the way selection pressure is distributed within the gene. We examined the data using three simple models: a one-ratio model (M0) that assumes one ω for all sites, a neutral model (M1) that assumes that there are two classes of sites within the gene; those that are conserved (p_0) with ω_0=0 and those that are neutral (p_1=1− p_0) with ω_1=1, and finally, a discrete model (M3) that assumes three classes of sites each having a unique value of ω that is estimated from the data (Yang, 1997). In each case a likelihood ratio was calculated which was used to compare and determine which model best reflected the observed data using a likelihood ratio test (LRT). The LRT statistic is twice the log-likelihood difference between two models being compared and has a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters between the two models (Yang et al., 2000).

2.6. Phylogenetic Analysis

Sequence relationships were inferred using the Neighbor-Joining method (Saitou and Nei, 1987) with 2000 bootstrap replicates (Felsenstein, 1985). Evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al., 2004). For the analysis of the *Topi* ORFs, codon positions included were 1^{st}+2^{nd}+3^{rd}+noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons. Phylogenetic analyses were conducted in MEGA4 (Tamura et al., 2007)

3. Results

3.1. Methods validation

Transposable element display is a DNA finger-printing method used to assess the copy number and position of transposable elements in the genome (Biedler et al., 2003; Guimond et al., 2003; Subramanian et al., 2007). We adapted the technique for the analysis of *Topi* transposable elements and estimated the copy number and site-occupancy distributions in five different populations of *An.gambiae s.s* in Africa. Because of the limited amount of genomic DNA available for analysis, a whole genome amplification method was employed to produce adequate amounts of DNA. Whole genome amplification is a method of uniformly producing microgram quantities of genomic DNA from small quantities of genomic DNA. Although shown by others to faithfully reproduce genomic DNA (Gorrochotegui-Escalante and Black,

2003), we confirmed the findings by comparing the results of transposable element display obtained using whole genome amplified DNA with those that were obtained using original, non-amplified genomic DNA. An analysis of 11 samples verified that the amplified genomic DNA reproduced the patterns of *Topi* insertion and copy number obtained from the original genomic DNA sample.

Transposable element (TE) display, as performed in this study does not result in the efficient amplification of fragments longer than 1kb because the extension time in the PCR reactions was only 1 minute. Because the *An.gambiae* genome is composed of 64.8% adenines and thymines and PCR templates for TE display were produced by digesting genomic DNA with *DpnII* (GATC), we expected only 7 % of the resulting fragments to be 1 kb or more in length. We assumed fragment sizes following *DpnII* digestion were distributed exponentially ($\lambda e^{-\lambda x}$) with the probability of a *DpnII* site (λ) = 0.00325174. This follows directly from the A/T, G/C composition of the genome and the recognition site for *DpnII*: $\lambda = 0.176^2\ 0.324^2$. All *DpnII* fragments displayed by TE display contain 90 bp of *Topi* sequences so we were concerned only with *DpnII* fragments 90 bp in length or greater. Because 74% and 5.2% of all fragments were greater than 90 bp and 1 kb, respectively, it follows that 7 % of all fragments greater than 90 bp were greater than 1 kb ($(^{0.052}/_{0.746}) \times 100\%$). Consequently we are confident of detecting 93% of the *Topi* elements present in the genome.

The specificity of *Topi* TE display was confirmed by eluting and sequencing 10 randomly selected bands from the gel. All the sequenced bands contained *Topi* elements as expected (Data not shown).

3.2. Copy number / Site Occupancy

In this study all individuals analyzed (n=73) had at least 2 copies of the *Topi* element and one mosquito sample from Malindi had 37 copies of the element. Mean element copy numbers in the five populations analyzed ranged from 10.2 – 33.8 per diploid genome. There was a statistically significant difference in copy numbers between all the locations ($p < 0.05$, Tukey's HSD test) except between Kisumu and Malindi, and Zenet and Bakin-Kogi (Table 1) (Figure 1). The copy number in Furvela was significantly lower than all the other locations analyzed. There were 19 and 17 elements with high site-occupancy frequencies that were present in more than 10 individuals in Malindi and Kisumu respectively. Furvela had the least number of high frequency occupied sites with only one that was present in 9 of 15 individuals.

We used the model of Charlesworth and Charlesworth (1983) (Charlesworth and Charlesworth, 1983) to analyze the observed site-occupancy distributions of the *Topi* element in *An.gambiae*. The model assumes the elements are at copy number equilibrium and that there are infinite insertion-sites within the genome. The model parameter β reflects the effects of forces other than drift that might be playing a role in shaping the observed distribution. According to the models, β values greater than one indicate that the forces of mobility and/or selection are responsible for the observed frequency distribution. We observed that all the locations except Bakin Kogi (β=1.5) showed β values less than one indicating that there has been little recent activity of *Topi* in *An.gambiae s.s* (Table 1) and that genetic drift is the major force shaping the distribution of site polyorphisms.

3.3. Structure of *Topi* elements

Autonomous Class II transposable elements code for functional transposase and can undergo transposition. Non-autonomous elements are usually deleted forms of the element that depend on transposase expressed from other elements in the genome. Class II elements like *P-elements* in *Drosophila* often exist in forms that have large internal deletions (Engels, 1989), however, *hAT* elements such as *Herves* in *An.gambiae* (Subramanian et al., 2007) and

Hermes in *Musca domestica* (L A. Cathcart, E S. Krafsur, P W. Atkinson, D A. O'Brochta and R A. Subramanian, unpublished) are rarely found with deletions. We analyzed the structure of *Topi* elements by amplifying the internal ~ 1kb *Topi* transposase coding region using PCR. We observed that all individuals analyzed (n= 73) had at least one copy of the 1 kb complete open reading frame and a ~600 bp deleted form (Figure 2). There were other less prevalent deleted elements of other sizes present in some of the individuals analyzed (Figure 2).

3.4. Nucleotide diversity of *Topi* elements

The 1kb complete *Topi* transposase coding region amplified was cloned and sequenced from 49 individuals to analyze the sequence diversity of the *Topi* elements in five different populations. Only one sequence per individual was obtained to give us the opportunity of sampling as many different elements as possible. All of the sampled 49 sequences were different from each other. The nucleotide sequence polymorphism ranged from $\pi = 0.029$ to $\pi = 0.062$ in the various locations of analysis with the average being $\pi = 0.051$ (Table 2). The π values were only significantly different between Malindi and Furvela, and Zenet and Furvela ($p < 0.05$, Tukey's HSD test). Eight deleted forms of *Topi* were recovered and analyzed. Two sequences were of Form A (828 bp), two were of Form E (572 bp), one each of Form B (785 bp), Form C (758 bp) and Form D (732 bp) (Figure 2). Deleted forms had ~ 200 bp to ~400 bp deletions in different regions of the *Topi* open reading frame (Figure 2). Form E had ~600 bp deletion when compared to the "canonical *Topi* ORF", however both the sequences had an extra 175 bp which was not similar to the canonical element (Figure 3).

Comparing the levels of silent-site diversity of transposable elements with that of single-copy host genes can be useful when looking for evidence of a lateral transfer event sometime in the history of the element and to understand when such an event might have occurred (Sanchez-Gracia et al., 2005). Here, we compared the silent-site diversity (π_s) of *Topi* elements with the average silent diversity of 35 nuclear genes in *An.gambiae* (Morlais et al., 2004); the observed silent-site diversity of *Topi* elements was significantly lower than the expected average silent-site diversity seen in the 35 nuclear genes. The observed π_s was 3 to 5-fold lower than the expected π_s in all locations analyzed (Table 3). Comparisons between populations showed that the observed π_s in Furvela was significantly lower than the π_s observed in Kisumu, Malindi and Zenet; and while the expected π_s in Bakin-Kogi was significantly lower than that in Malindi ($p < 0.05$, Tukey's HSD test). The expected silent-site diversity in Furvela was significantly lower than π_s at all other locations and the expected π_s in Bakin-Kogi was lower than all the other locations except Zenet ($p < 0.05$, Tukey's HSD test). The expected silent-site diversity in Kisumu and Malindi, and Zenet and Bakin-Kogi were not significantly different from each other.

3.5. Natural Selection

We tested for evidence of selective constraints within the *Topi* transposase coding region of the 14 sequences which had no pre-mature stop codons by estimating $\omega = d_N/d_S$ using Maximum Likelihood as described (Yang, 1997; Yang et al., 2000). The ω values ranged from 0.45 to 0.51 under all models (M0, M1 and M3) showing evidence of purifying selection. Even though, the discrete model (M3) fit the data better than the neutral model (M1) the Likelihood Ratio Test (LRT) statistic, $2\Delta l$ ($2\Delta l = 2(-2371.47 - (-2366.7))$, for this comparison was 9.54, which was less than the critical value of $\chi^2_{[0.001, 2]} = 13.816$.

3.6. Phylogeographic analysis

Neighbor joining analysis of the nucleotide sequence of the 49 transposase coding regions isolated in this study along with *Topi 1, 2* and *3* and *Tc1* as an out-group resulted in two diverse clades of elements (Figure 4) corresponding to *Topi 1* and *Topi 2* elements. Support for these two clades was moderately good (bootstrap value of 83%) but within these groups there was

a high degree of diversity. No *Topi 3* elements were recovered in this study. There was some correlation between location and element type in that all elements recovered from Nigeria, with one exception, were *Topi 2* elements while all elements recovered from Mozambique, with one exception, were *Topi 1* elements. Elements recovered from Kenya and Tanzania belonged to both the *Topi 1* and *Topi 2* families.

4. Discussion

An. gambiae contains some 30 different types of IS630-Tc1-*mariner*-type transposable elements (Grossman et al., 1999; Hill et al., 2001; Coy and Tu, 2005). Earlier studies of *Topi* elements in *An. gambiae* suggested that these *Tc1*-like elements might be active because the authors observed the presence of apparently autonomous forms of the elements. Based on the evolution of *P* elements in *D. melanogaster* where internally deleted forms of the element evolve rapidly, the presence of autonomous forms of transposable elements is often taken to indicate recently introduced, functional elements (Rio, 2002). In addition there were at least three lineages of *Topi* elements reported in *An. gambiae* (Coy and Tu, 2005). Based on insertion site polymorphism we found little evidence for recent *Topi* element movement and evidence for only two *Topi* element lineages within the *An. gambiae* populations examined here. The studies reported here were based on sampling 73 individuals from five populations of *An. gambiae* while the earlier studies of *Topi* were based on a much more limited sample.

One of the outstanding features of the *Topi* elements analyzed in this study was their high degree of sequence diversity ($\pi = 0.051$). Notably, of the 49 transposase-coding regions that were sequenced we found no two sequences to be identical and the two most closely related sequences had a pair-wise nucleotide diversity of 0.002. Class II transposable elements move by a copy-and-paste mechanism in which element replication relies on the cell's DNA polymerase and repair machinery. Because these processes are not particularly error-prone Class II transposition tends to maintain a low element diversity within a population. This is unlike Class I elements that rely on error-prone reverse transcriptase in which element movement increases sequence diversity. For example, *P* elements have been in *Drosophila melanogaster* for only approximately a century, during which time they have been very active. Yet, within this species there is no sequence diversity ($\pi = 0$) (O'Hare and Rubin, 1983; Sakoyama et al., 1985). *Herves* elements appear to have entered the *Anopheles* lineage prior to the diversification of the *An. gambiae* species complex and show evidence of activity in the recent past (Subramanian et al., 2007). Sequence diversity among *Herves* elements within *An. gambiae* s.s. was ten fold lower than observed among *Topi* elements ($\pi = 0.0046$) (Subramanian et al., 2007). These sequence diversity data are consistent with the idea that *Topi* elements have been in this lineage at least as long as *Herves* and probably much longer.

Phylogenetic analysis of the nucleotide sequences of the transposase coding regions of 49 elements using distance-based methods revealed the presence of two very diverse groups of elements that appeared to represent two lineages described previously by Coy and Tu (2005) and referred to by them as *Topi 1* and *Topi 2*. While support for the *Topi 1* and *Topi 2* lineages was moderately good (bootstrap value of 83%) the elements comprising each of these lineages were highly diverse (Figure 3). Interestingly almost all elements sampled from Bakin Kogi, Nigeria were *Topi 2*-like elements while almost all elements sampled from Mozambique were *Topi 1*-like elements. *An. gambiae* collected in Tanzania and Kenya had elements belonging to both the *Topi 1* and *Topi 2* clades. The distinction among elements from the extreme geographical locations sampled in this study (Nigeria and Mozambique) may reflect a long history of element immobility (see below) and perhaps very limited gene flow between insects from these locations (Lehmann et al., 2000; Lehmann et al., 2003).

Topi 3 was originally described by Coy and Tu (Coy and Tu, 2005) and based on the data presented here it is clear that this is a very unusual element. We found no other elements that were closely related to this element and consequently *Topi 3* remains somewhat mysterious. *Topi* 3 was originally detected in the genome of the PEST strain, a strain of *An. gambiae* created in the 1990s by crossing a laboratory strain started from mosquitoes collected in Nigeria and containing the *pink eye* mutation on the X chromosome to the offspring of wild-caught *An. gambiae* from western Kenya. Mosquitoes with pink eyes were selected and repeatedly (three times) outcrossed with mosquitoes from western Kenya. Thus the PEST strain contains a hybrid genome derived from mosquitoes from eastern and western Africa (Holt RA et al., 2002). Interestingly, mosquitoes sampled in the present study included insects collected from both western Kenya and Nigeria but *Topi 3* was not detected. Indeed, the elements in mosquitoes from Nigeria had limited diversity and were all *Topi 2*-like elements, with one exception.

Although we are confident that the PCR primers used to detect *Topi* elements would anneal to the ORF of the *Topi 3* elements described by Coy and Tu (2005) (Supplemental Figure 1) we cannot rule out the possibility that some forms of *Topi 3* went undetected because of polymorphisms in the priming region. Assuming our sampling of *Topi 3* was fair (no primer bias) the frequency of individuals with *Topi 3* elements would have to be less than 0.1 since at this frequency we would be 95% certain of observing at least one *Topi 3*-containing individual in a sample of 49 individuals. The low frequency of *Topi 3* could be the result of a long history of inactivity within the genome and its drift to extinction. Alternatively, the low frequency could reflect its recent addition to the genome and its incomplete invasion of this species.

Long-time transposable element residents of genomes accumulate mutations and internal deletions, as is seen in the case of *Topi 1* and *2* elements that consist of mostly mutated forms with occasional elements that appear autonomous. Although it is formerly possible that the last remaining elements in the *Topi 3* family appear to be autonomous, this seems unlikely. More likely is that *Topi* 3 has been recently added to the genome of *An. gambiae s.s.* and has not had time to invade all genomes. More data are required to clearly distinguish between these two hypotheses including deeper sampling using a variety of strategies and directly testing the functionality of existing ‘autonomous’ *Topi 3* elements.

If *Topi* 3 is an element that is invading the *An. gambiae s.s.* genome it could prove very valuable to efforts directed toward the development of genetic control strategies for *Plasmodium* transmission through the introduction of ‘resistance’ genes into mosquito genomes on self-mobilizing transposable elements. Elements presently invading *An gambiae s.s.* will provide unique opportunities to study transposable element invasion in natural populations of *An. gambiae* and to study how population structure will impact this process.

Insertion site polymorphism data, also referred to as site occupancy data, has been used to infer the history of activity of elements (Charlesworth and Langley, 1990). Elements undergoing transposition increase the amount of site occupancy polymorphism while elements that are inactive will drift to fixation or extinction, lowering site occupancy polymorphism observed in a population. High insertion site polymorphism data are taken to reflect recent activity and while this is reasonable the exact temporal dimension of this activity depends on the strength of other forces acting on the elements and the effective population size (Montgomery and Langley, 1983; Leigh-Brown and Moss, 1987; Charlesworth and Langley, 1990; Charlesworth et al., 1992; Biemont et al., 1994; Nuzhdin, 1995; Labrador et al., 1998; Nuzhdin, 1999; Wright et al., 2001; Neafsey et al., 2004). Neutral polymorphisms will persist in populations in proportion to the effective population size. The age (in generations) of neutral polymorphisms will be some small multiple of the effective population size. Effective population sizes of *An. gambiae* populations have been estimated at 10^4–10^5 and if there are 10 generations a year

then polymorphisms could persist within populations 10^3–10^4 years (Lehmann et al., 1998; Taylor et al., 1993). This provides a very rough window of time in which to consider the dynamics of *Topi*.

We applied common models of transposable element dynamics (Charlesworth and Charlesworth, 1983; Langley et al., 1983; Charlesworth and Langley, 1990) as part of an effort to infer the activity levels of *Topi* elements. In this study the value of 'Langley-Charlesworth'-β in all of the populations tested was approximately one, indicating that genetic drift has been the major force that has shaped the site-occupancy distribution of the *Topi* elements observed in this study (Charlesworth and Langley, 1990). There is little evidence that *Topi 1*- and *2*-type elements are currently active or that they have been active in the last 10^3–10^4 years. Consistent with this conclusion is the high degree of element diversity that was observed and the complete absence of any elements that were identical among the elements sampled. While the models assume the elements are at transposition/excision equilibrium, failure of this assumption will not lead to alternative interpretations of the data reported here. Situations where the assumption of equilibrium does not hold are likely to lead to overestimation of the model parameter β that might give the impression that the element has been more active than it actually has. This type of error would not change the basic conclusion of this study, that *Topi 1* and *2* are not highly active elements in *An. gambiae*.

Taken together the data presented in this study support the idea that *Topi 1* and *Topi 2* are long-time residents of this *An. gambiae* lineage that have not experienced high levels of transposition. Grossman et al. (1999) reported the presence of *Topi* in other members of the *An. gambiae* species complex suggesting that the origins of this element predate the divergence of this species complex. We do not know, however, the type of *Topi* elements present in these other species. Horizontal transfer is thought to play a central role in the evolution and distribution of transposable elements (Hartl et al., 1997; Robertson, 2002) and different methods have been devised for recognizing such events. One method relies on comparing the silent site diversity of transposable element transposase coding regions and the coding regions of nuclear genes whose histories have not involved horizontal transfer (Sanchez-Gracia et al., 2005). Transposable elements with short histories within a host genome will display less silent site diversity relative to that displayed in nuclear genes under some conditions (Sanchez-Gracia et al., 2005). In this study we observed that *Topi* transposase coding regions had less silent site diversity than nuclear genes but the difference was quite small, only 3–5 fold lower than nuclear genes. Others who have used silent site diversity to infer historical horizontal transfer of transposable elements have typically reported much greater differences for elements thought to have undergone horizontal transfer. For example, Sanchez-Garcia et al. (2005) reported 100-fold less silent site diversity within 13 transposable elements in *D. melanogaster* thought to have been introduced into the species between 5 –12 million years ago.

Topi elements displayed two characteristics in common with the *An. gambiae hAT* element, *Herves* (Arensburger et al., 2005; Subramanian et al., 2007). First, like *Herves*, numerous *Topi* elements (40%) have transposase coding regions that appeared capable of producing functional transposase (Subramanian et al., 2007). These elements however are not stimulating widespread element movement in this species and how such sequences can exist in a transposable element system that shows no evidence of recent movement remains unknown. Second, like *Herves*, there is an abundance of 'intact' *Topi* elements (Subramanian et al., 2007). The evolution of transposable elements in *D. melanogaster*, in particular the *P* element, involved the rapid increase in element copy number when the element is active and the creation of defective elements through the deletion of internal sequences (Engels, 1996). The rapid creation of populations of *P* elements with internal deletions is a consequence of the frequent reliance of *D. melanogaster* germ cells on homology-dependent gap repair mechanisms following element excision and the apparently infrequent use of non-homologous end joining

mechanisms. In *An. gambiae* intact *Topi* and *Herves* elements are abundant and there are relatively few internally deleted elements (Figure 2). The internally deleted *Topi* elements that were recovered are consistent with their creation during homology-dependent gap repair. The relative abundance of both intact *Topi* and *Herves* elements in *An. gambiae* may reflect some differences in the relative importance of certain gap repair mechanisms in the germ cells of *An. gambiae* compared to *D. melanogaster*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

ANOVA, analysis of variance
ATP, adenosine triphosphate
bp, basepair
C, centigrade
dATP, 2'-deoxyadenosine 5'-triphosphate
dCTP, 2'-deoxycytosine 5'-triphosphate
dGTP, 2'-deoxyguanisine 5'-triphosphate
DNA, deoxyribonucleic acid
dNTP, 2'-deoxynucleotide 5'-triphosphate
dTTP, 2'-deoxythymidine 5'-triphosphate
EDTA, ethylenediaminetetraacetic acid
HSD, honestly significant difference
kp, kilobasepairs
LRT, likelihood ratio test
$MgCl_2$, magnesium chloride
min, minutes
ML, maximum likelihood
mM, millimolar
ORF, open reading frame
PCR, polymerase chain reaction
pmoles, picomoles
sec, second
TBE, Tris-borate EDTA
TE, transposable element
µg, microgram
µl, microliter

Acknowledgements

Support for this work came from the National Institutes of Health (GM48102), the World Health Organization's Special Programme for Research and Training in Tropical Diseases (A60169) and the MacArthur Foundation.

References

Aksoy, S. Advances in experimental medicine and biology. Vol. 627. Austin: Landes Bioscience; 2008. Transgenesis and the management of vector-borne disease.

Arensburger P, Orsetti J, Kim Y-J, Aluvihare C, O'Brochta DA. A new active transposable element, *Herves*, from the African malaria mosquito *Anopheles gambiae*. Genetics 2005;169:697–708. [PubMed: 15545643]

Atkinson, PW. Proposed uses of transposons in insect and medical biotechnology. In: Aksoy, S., editor. Transgenesis and the management of vector-borne disease. Vol. 627. Austin: Landis Bioscience; 2008.

Biedler J, Qi Y, Holligan D, della Torre A, Wessler S, Tu Z. Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. Insect Molec. Biol 2003;12:211–216. [PubMed: 12752653]

Biemont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, Aulard S, Pasyukova EG. Population dynamics of the *copia, mdg1, mdg3, gypsy* and *P* transposable elements in a natural population of *Drosophila melanogaster*. Genet. Res 1994;63:197–212. [PubMed: 8082837]

Charlesworth B, Charlesworth D. The population dynamics of transposable elements. Genet. Res 1983;42:1–27.

Charlesworth, B.; Langley, CH. Population genetics of transposable elements in Drosophila. In: Selander, RK.; Clark, AG.; Whittam, TS., editors. Evolution at the molecular level. Sunderland: Sinauer Associates, Inc.; 1990. p. 150-177.

Charlesworth B, Lapid A, Canada D. Distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster*. II. Inferences on the nature of selection against elements. Genet Res 1992;60:115–130. [PubMed: 1334900]

Coy MR, Tu Z. *Gambol* and *Tc1* are two distinct families of DD34E transposons: analysis of the *Anopheles gambiae* genome expands the diversity of the *IS630-Tc1-mariner* superfamily. Insect Molecular Biology 2005;14:537–546. [PubMed: 16164609]

Engels, WR. P elements in *Drosophila melanogaster*. In: Berg, DE.; How, MM., editors. Mobile DNA. Washington D.C.: American Society for Microbiology; 1989. p. 439-484.

Engels, WR. Transposable Elements. Vol. 204. 1996. P elements in Drosophila; p. 103-123.

Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 1985;39:783–791.

Gorrochotegui-Escalante N, Black WC. Amplifying whole insect genomes with multiple displacement amplification. Insect Molecular Biology 2003;12:195–200. [PubMed: 12653941]

Grossman GL, Cornel AJ, Rafferty CS, Robertson HM, Collins FH. *Tsessebe, Topi and Tiang*: three distinct *Tc*1-like transposable elements in the malaria vector, *Anopheles gambiae*. Genetica 1999;105:69–80. [PubMed: 10483094]

Guimond N, Bideshi DK, Pinkerton AC, Atkinson PW, O'Brochta DA. Patterns of *Hermes* Transposition in *Drosophila melanogaster*. Molec. Gen. Genet 2003;268:779–790.

Hartl DL, Lohe AR, Lozovskaya ER. Modern thoughts on an ancient marinere: function, evolution, regulation. Annu Rev Genet 1997;31:337–358. [PubMed: 9442899]

Hill SR, Leung SS, Quercia NL, Vasiliauskas D, Yu J, Pasic I, Leung D, Tran A, Romans P. *Ikirara* insertions reveal five new *Anopheles gambiae* transposable elements in islands of repetitious sequence. Journal of Molecular Evolution 2001;52:215–231. [PubMed: 11428459]

Holt RA, et al. The genome sequence of the malaria mosquito *Anopheles gambiae*. Science 2002;298:129–149. [PubMed: 12364791]

Kidwell M. Horizontal transfer of P elements and other short inverted repeat tansposons. Genetica 1992;86:275–286. [PubMed: 1334912]

Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 2004;5:150–163. [PubMed: 15260895]

Labrador M, del Carmen Seleme M, Fontdevila A. The evolutionaly history of *Drosophila buzzatii.* XXXIV. The distribution of the retrotransposon *Osvaldo* in original and colonizing populations. Mol Biol Evol 1998;15:1532–1547. [PubMed: 12572617]

Langley CH, Brookfield JFY, Kaplan N. Transposable elements in mendelian populations. I. A theory. Genetics 1983;104:457–471. [PubMed: 17246142]

Lehmann T, Blackston CR, Besansky NJ, Escalante AA, Collins FH, Hawley WA. The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: implications for population structure. J. Heredity 2000;91:165–168.

Lehmann T, Hawley WA, Grebert H, Collins FH. The effective population size of *Anopheles gambiae* in Kenya: Implications for population structure. Mol Biol Evol 1998;15:264–276. [PubMed: 9501493]

Lehmann T, Licht M, Elissa N, Maega BTA, Chimumbwa JM, Watsenga FT, Wondjii CS, Simard F, Hawley WA. Population structure of *Anopheles gambiae* in Africa. J. Heredity 2003;94:133–147.

Leigh-Brown AJ, Moss JE. Transposition of the *I* element and *copia* in natural populations of *Drosophila melanogaster*. Genet Res 1987;49:231–237.

Meeraus WH, Maxwell C, O'Brochta DA. *Herves* transposable elements in Tanzanian *Anopheles gambiae* - potential uses in malaria control. Trans. Royal Soc. Trop. Med. Hyg 2005;99:942–956.

Montgomery EA, Langley CH. Transposable elements in Mendelian populations. II. Distribution of three *copia*-like elements in a natural population. Genetics 1983;104:473–483. [PubMed: 17246143]

Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D. Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. Am. J. Trop. Med. Hyg 2004;71:795–802. [PubMed: 15642974]

Neafsey DE, Blumenstiel JP, Hartl DL. Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. Mol Biol Evol. 2004

Nei, M.; Kumar, S. Molecular evolution and phylogenetics. New York: University Press; 2000.

Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 1979;76:5269–5273. [PubMed: 291943]

Nuzhdin SV. The distribution of transposable elements on X chromosomes from a natural population of *Drosophila simulans*. Genet. Res 1995;66:159–166. [PubMed: 8522157]

Nuzhdin SV. Sure facts, speculations, and open questions about the evolution of transposable element copy number. Genetica 1999;107:129–137. [PubMed: 10952206]

O'Brochta, DA.; Handler, AM. Perspectives on the state of insect transgenics. In: Aksoy, S., editor. Transgenesis and the management of vector-borne disease. Vol. 627. Austin: Landis Bioscience; 2008.

O'Brochta DA, Sethuraman N, Wilson R, Hice RH, Pinkerton AC, Levesque CS, Bideshi DK, Jasinskiene N, Coates CJ, James AA, Lehane MJ, Atkinson PW. Gene vector and transposable element behavior in mosquitoes. Journal of Experimental Biology 2004;207:1263–1263.

O'Brochta DA, Subramanian RA, Orsetti J, Peckham E, Nolan N, Arensburger P, Atkinson PW, Charlwood D. *hAT* element population genetics in *Anopheles gambiae s.l.* in Mozambique. Genetica 2006;127:185–198. [PubMed: 16850223]

O'Hare K, Driver A, McGrath S, Johnson-Schlitz DM. Distribution and structure of cloned *P* elements for the *Drosophila melanogaster P* strain π2. Genet Res 1992;60:33–41. [PubMed: 1333435]

Rio, DC. P transposable elements in *Drosophila melanogaster*. In: Craig, NL.; Craige, R.; Gellert, M.; Lambowitz, AM., editors. Mobile DNA II. Washington, D. C.: ASM Press; 2002. p. 1204

Robertson, HM. Evolution of DNA transposons in eukaryotes. In: Craig, NL.; Craige, R.; Gellert, M.; Lambowitz, AM., editors. Mobile DNA II. Washington, D. C.: ASM Press; 2002. p. 1093-1110.

Rozas J, Rozas R. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci 1995;11:621–625. [PubMed: 8808578]

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 2003;19:2496–2497. [PubMed: 14668244]

Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molec. Bio. Evol 1987;4:406–425. [PubMed: 3447015]

Sakoyama Y, Todo T, Ishiwachigusa S, Honjo T, Kondo S. Structures of defective P transposable elements prevalent in natural-Q and Q-derived M-strains of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 1985;82:6236–6239. [PubMed: 16593607]

Sanchez-Gracia A, Maside X, Charlesworth B. High rate of horizontal transfer of transposable elements in *Drosophila*. Trends in Genetics 2005;21:200–2003. [PubMed: 15797612]

Shao HG, Tu ZJ. Expanding the diversity of the *IS630-Tc1-mariner* superfamily: Discovery of a unique DD37E transposon and reclassification of the DD37D and DD39D transposons. Genetics 2001;159:1103–1115. [PubMed: 11729156]

Subramanian RA, Arensburger P, Atkinson PW, O'Brochta DA. Transposable element dynamics of the *hAT* element *Herves* in the human malaria vector, *Anopheles gambiae, s.s.* Genetics 2007;176:2477–2487. [PubMed: 17603116]

Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molec. Bio. Evol 2007;24:1596–1599. [PubMed: 17488738]

Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA 2004;101:11030–11035. [PubMed: 15258291]

Taylor CE, Toure YT, Coluzzi M, Petrarca V. Effective population size and persistence of *Anopheles arabiensis* during the dry season in west Africa. Med. Vet. Entomol 1993;7:351–357. [PubMed: 8268490]

Watterson GA. On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol 1975;7:256–276. [PubMed: 1145509]

Yang XH, Ma H, Makaroff CA. Characterization of an unusual *Ds* transposable element in *Arabidopsis thaliana*: insertion of an abortive circular transposition intermediate. Plant Molecular Biology 2004;55:905–917. [PubMed: 15604724]

Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci 1997;13:555–556. [PubMed: 9367129]

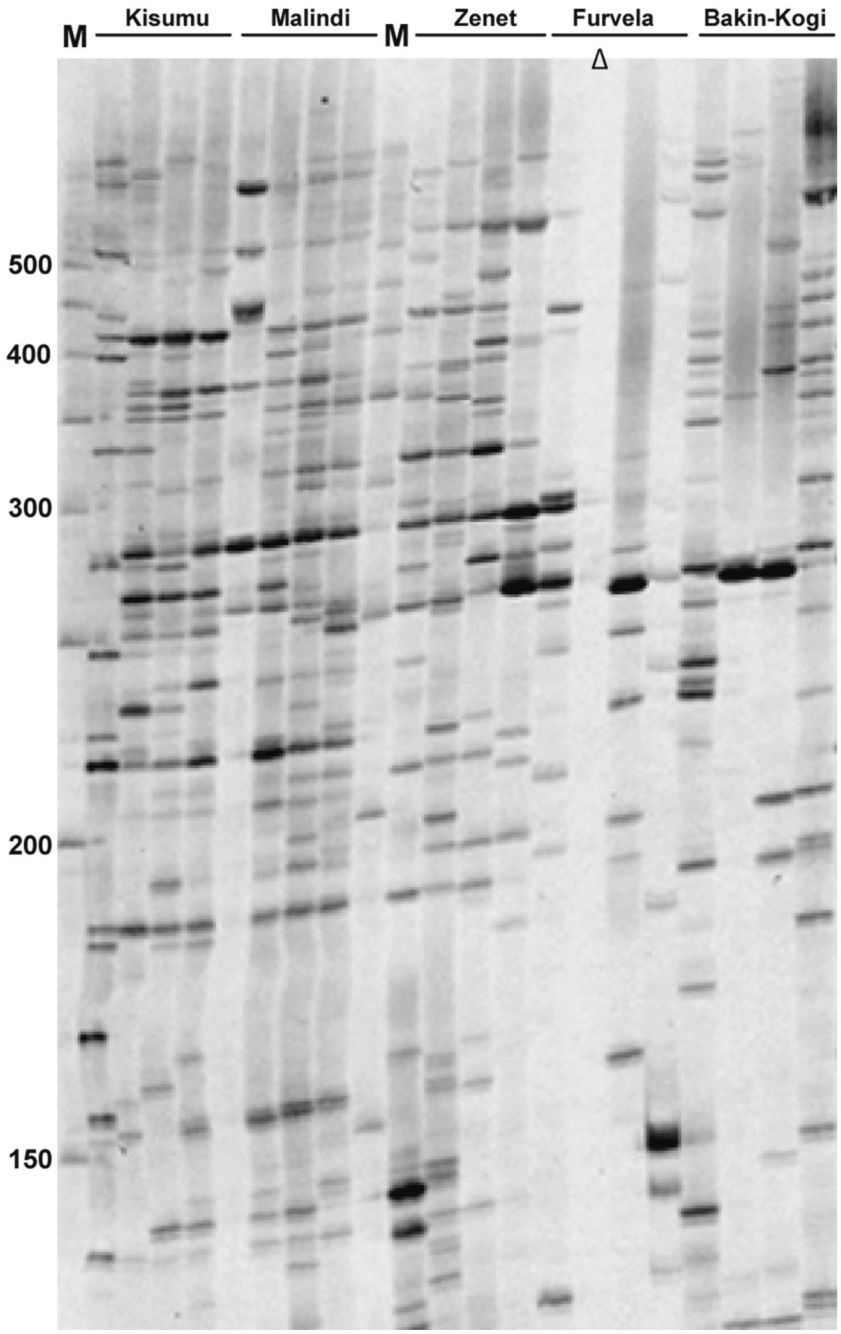

Figure 1. Transposable element display of the right end of *Topi* elements

A sample of transposable element display results obtained from five different locations is represented. Molecular weight markers (M) in base pairs are shown on the left side. $^{\Delta}$ Empty lane

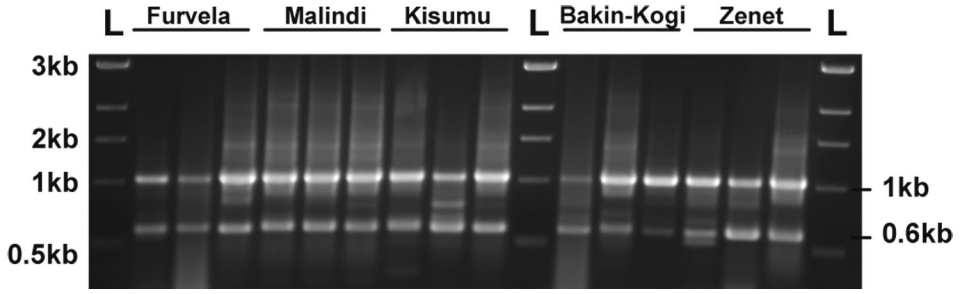

Figure 2. Structure of *Topi* elements

PCR products of a sample of individuals from five different locations used to analyze the structure of *Topi* elements are shown. Numbers at the top of each lane refer to sample names. Molecular weight markers (M) are shown on the left side in kilobase pairs. The ~1kb complete *Topi* transposase open reading frame is indicated on the right side. Approximately, 0.6 kb deleted form observed in all individuals is also indicated on the right side

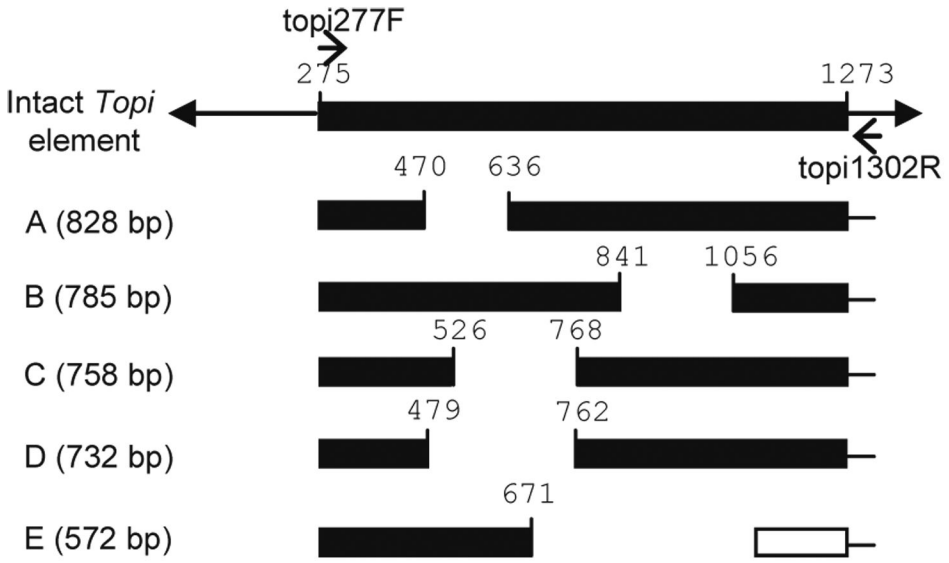

Figure 3. Structure of deleted forms of *Topi* elements

The position of the deletion corresponding to the full length *Topi* element is shown for each form. The position of deletion and the additional 175 bp of sequence that is not similar to the full length element is also shown for Form E. The position of the primers, topi277F and topi1302R that were used to amplify these forms are also shown.

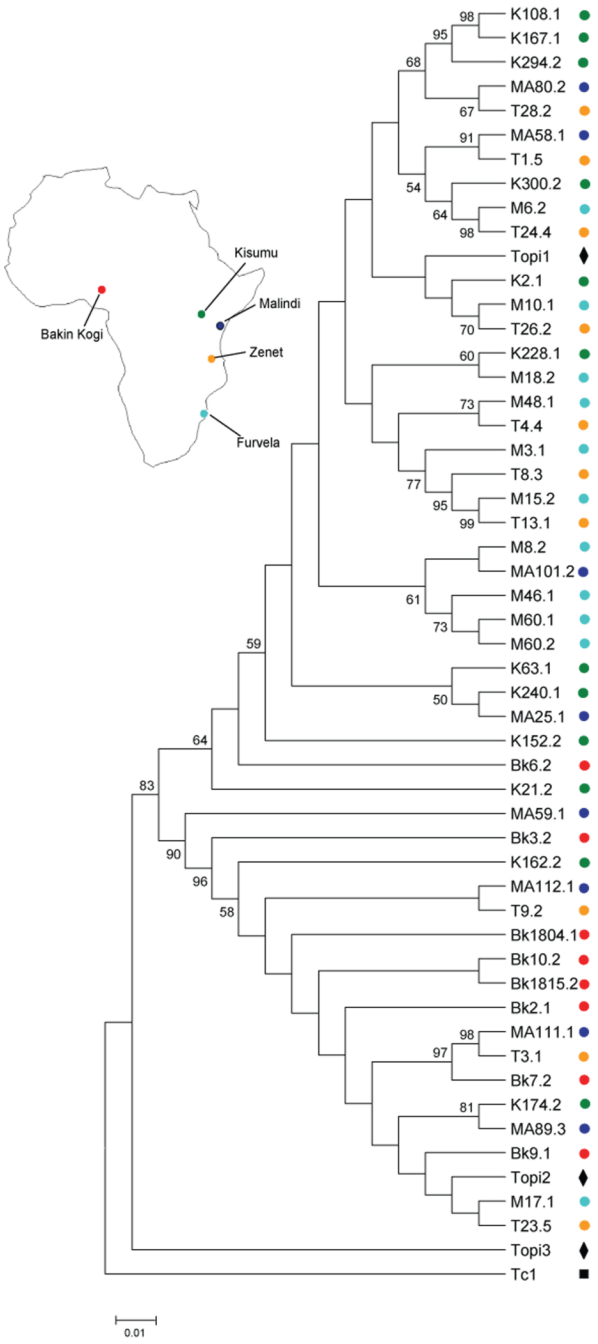

Figure 4. Phylogeographic analysis of *Topi* transposase coding region nucleotide Sequences

Evolutionary history inferred by the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.66170490 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (2000 replicates) are shown next to each branch in cases where it was 50% or more. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The Maximum Composite Likelihood method was used to compute evolutionary distances and are in units of the number of base substitutions per site (Tamura et al., 2004). There were a total of 1192 positions in the final dataset. Phylogenetic analyses were

conducted in MEGA4 (Tamura et al., 2007). *Topi 1*, *2* and *3* from Coy and Tu (2005) are indicated with diamonds. Colored dots indicate the location where sequences were recovered.

Table 1

Site occupancy

Location	N^{*}	κ^{Δ}	dcn†	$\beta^{\dagger\dagger}$
Kisumu (k)	16	72	31.3^{zfb}	0.6
Malindi (m)	16	78	33.8^{zfb}	0.7
Zenet (z)	16	56	18.2^{kmf}	0.5
Furvela (f)	15	59	10.22^{kmzb}	0.8
Bakin Kogi (b)	10	63	18.4^{kmf}	1.5

*Individuals analyzed by transposable element display

$^{\Delta}$Number of unique chromosomal sites containing *Topi*

†Diploid copy number of *Topi* (Wright et al., 2001)

$^{\dagger\dagger}4N_e(v+s)$ from Charlesworth and Charlesworth (1983)

$^{k\,m\,z\,f\,b}$the copy number was significantly different from the indicated location at a significance level of $p < 0.05$

Table 2

Nucleotide sequence polymorphism in *Topi* open reading frame

Location	Seqs*	Poly†	π^{Δ}	$\theta^{¶}$
Kisumu (k)	12	168	0.0452 (0.009)	0.056 (0.022)
Malindi (m)	8	149	0.062 (0.0077)^{f}	0.060 (0.026)
Zenet (z)	10	145	0.0521 (0.0099)^{f}	0.053 (0.022)
Furvela (f)	11	130	0.029 (0.0106)^{mz}	0.049 (0.019)
Bakin-Kogi (b)	8	122	0.047 (0.013)	0.067 (0.029)
Combined	49	227	0.051 (0.0032)	0.086 (0.024)

*Number of sequences analyzed

†Number of polymorphic positions

ΔPairwise nucleotide diversity (Nei and Li, 1979); standard deviation in parenthesis

¶Nucleotide diversity based on segregating sites (Watterson, 1975); standard deviation in parenthesis

$f\ m\ z$ π was significantly different from the indicated location at a significance level of $p< 0.05$

Table 3

Silent-site diversity of *Topi* elements from different locations

Locations	Haploid copy number	π_s †		
		Observed	Expected[Δ]	Observed/Expected[*]
Kisumu (k)	15.65	0.063 (0.041)[f]	0.33 (0.064)[zfb]	0.193
Malindi (m)	16.9	0.081 (0.044)[fb]	0.353 (0.075)[zfb]	0.227
Zenet (z)	9.1	0.063 (0.042)[f]	0.190 (0.082)[f]	0.332
Furvela (f)	5.11	0.022 (0.023)[kmz]	0.107 (0.066)[kmzb]	0.220
Bakin Kogi (b)	9.2	0.044 (0.043)[m]	0.192 (0.087)[kmf]	0.229
All	11.13[#]	0.051	0.238	0.219

† πs represents the average pairwise nucleotide diversity at synonymous sites

Δ see Materials and Methods

Average haploid copy number from all locations

k m z f b π_S was significantly different from the indicated location at a significance level of $p < 0.05$

* the observed π_S was significantly lower than the expected π_S at all locations at a significance level of $p < 0.05$