Evolution of the T1 Retroposon Family in the
Anopheles gambiae Complex

Nora J. Besansky

Department of Biology, Yale University, and Malaria Branch, Centers for Disease Control

The T1 family of retrotransposable elements is interspersed and moderately repeated in five member species of the Anopheles gambiae sibling-species complex and has diverged little since the radiation of the complex. T1 includes two closely related but independent subfamilies, defined by the presence or absence of linked sets of restriction sites, in all but one species, although the relative abundance of the subfamilies differs within each. Sequence analysis of a 349-bp region from 21 clones isolated from A. gambiae confirmed the bipartite organization by revealing 19 coordinated nucleotide differences between the two subfamilies—T1α and T1β. Sequence divergence is not only greater between than within subfamilies, but divergence within T1β is less than that within T1α. Between-species comparisons of genomic consensus restriction maps revealed that T1α is fixed for species-diagnostic differences in all species. With one exception, these subfamilies account for ~70% of detectable T1 copies in the genome. The results support retroposition as the dominant mechanism underlying the evolution of the T1 family.

Introduction

Families of repetitive DNA are sets of genomic sequences capable of forming stable base-paired duplexes with one another (Klein et al. 1978). Although in most cases their biological function, if any, remains unknown, significant advances have been made in understanding the mechanisms responsible for their evolution. Some of this progress has resulted from investigations focused on the two dominant families of repetitive sequences in primates and rodents, the Alu and L1 families. Each is the best-studied representative of short and long eukaryotic dispersed repeats, SINES and LINES, respectively (Singer 1982). Although important differences distinguish Alu and L1, both appear to be propagated through RNA intermediates (Weiner et al. 1986), and both seem to be maintained primarily by retroposition rather than by gene conversion (Hwu et al. 1986; Casavant et al. 1988).

Recent analyses of an extensive sequence data base have shown the human Alu family to be divided into at least five subfamilies on the basis of correlated differences (Britten et al. 1988; Jurka and Smith 1988). Similarly, numerous heterogeneities in the L1 family indicate substructuring. In mice, these include linked restriction-site differences among major genomic L1 populations (Brown and Dover 1981; Jubier-Maurin et al. 1985) and two alternative nonhomologous 5' ends (Loeb et al. 1986; Mottez et al. 1986). In humans, there are two sets of cytoplasmic transcripts whose 3' ends differ at four diagnostic positions (Skowronski et al. 1988). Two current models for the history of the Alu (Britten et al. 1988, 1989) and L1 families (Hardies et al. 1988, 1989)

1. Key words: T1 family, retrotransposable element, Anopheles gambiae, evolution.
2. Present address and address for correspondence and reprints: Retrovirus Diseases Branch G-19, Centers for Disease Control, Atlanta, Georgia 30333.

© 1990 by The University of Chicago. All rights reserved.
0737-4038/90/0703-0003$02.00
1986) agree on the following points: The evolution of the family can be accounted for by a small number of genes evolving under selective constraint. Their transcripts are often reverse transcribed and continually inserted into the genome as pseudogenes, which degenerate relatively rapidly. Subfamilies, discrete sets of elements with common diagnostic positions, result from the transcription of different parental genes.

T1 is a recently described family of long dispersed sequences shared between members of the Afrotropical Anopheles gambiae sibling-species complex (Besansky 1990). Full-length elements are 4.7 kbp in length, over 92% of which is occupied by two long overlapping open reading frames. One of these possesses structural motifs characteristic of reverse transcriptases. Most elements are truncated at the 5' end, so that 3' sequences are four times as abundant in the genome as 5' sequences. At the 3' end is an (A)-rich tail consisting of tandem repetitions of the monomer (TGAAA). These features suggest that it is propagated through an RNA intermediate.

The present study was motivated by the discovery, made during an initial characterization of the T1 family, that two consensus restriction maps describe T1 in A. gambiae. Data are presented showing that both maps are the result of correlated restriction-site differences. Supporting the mapping data are sequence comparisons involving a 349-bp region from 21 T1 elements from A. gambiae. These revealed at 19 positions diagnostic differences that identify the two T1 subsets, T1a and T1b. The extent to which T1 structure has been conserved in the A. gambiae species complex was investigated. The time since separation is not known, but, because differentiation is thought to have been man influenced (Coluzzi et al. 1985), is likely to have been within the past 20,000 years. Morphologically indistinguishable, they are capable of producing fertile female hybrids in the laboratory and, to a limited extent, in nature (White 1970). Restriction-mapping analysis and Southern hybridization techniques revealed that in addition to conservation of overall sequence, organization, and copy number, both subfamilies are present in four of five species examined. One subfamily has a genomic consensus map that is diagnostic for each species. The findings suggest that this T1 subfamily is evolving independently, as a result of the retroposition of separate source genes. At present, it is unclear whether the bipartite structure of the family predated the radiation of the complex or developed following the branching-off of one sibling.

Material and Methods

Mosquitoes

Specimens were obtained from laboratory colonies maintained at the Centers for Disease Control. Geographical origins of colonies are listed in table 1.

DNA Preparation

DNA was extracted from single freshly frozen adult mosquitoes according to a method described by Livak (1984), with volumes doubled. Plasmid DNA was prepared by a standard boiling procedure (Holmes and Quigley 1981) without phenol:chloroform extractions.

DNA Probes

T1 elements from lambda phage 31b and 916 were isolated from an Anopheles gambiae G3 library according to a method described elsewhere (Besansky 1990). A 1.7-kb SalI/HindIII segment of 31b was subcloned into a Bluescript (Stratagene) plasmid and was denoted as BSH17.31b. A 1.7-kb EcoRI/XhoI segment of 916 was sub-
Table 1
Geographic Origin of Specimens Analyzed

<table>
<thead>
<tr>
<th>Species and Colony</th>
<th>Geographic Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anopheles gambiae:</td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>Gambia</td>
</tr>
<tr>
<td>GMMK6</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>AS46</td>
<td>Kenya</td>
</tr>
<tr>
<td>ZAN</td>
<td>Zanzibar</td>
</tr>
<tr>
<td>A. arabiensis:</td>
<td></td>
</tr>
<tr>
<td>ARZAG</td>
<td>Burkina Faso</td>
</tr>
<tr>
<td>KISU</td>
<td>Kenya</td>
</tr>
<tr>
<td>A. melas:</td>
<td></td>
</tr>
<tr>
<td>BRE</td>
<td>Gambia</td>
</tr>
<tr>
<td>BAL</td>
<td>Gambia</td>
</tr>
<tr>
<td>A. merus:</td>
<td></td>
</tr>
<tr>
<td>V12</td>
<td>Kenya</td>
</tr>
<tr>
<td>ZULU</td>
<td>Zululand</td>
</tr>
<tr>
<td>A. quadriannulatus:</td>
<td></td>
</tr>
<tr>
<td>CHIL</td>
<td>Zimbabwe</td>
</tr>
<tr>
<td>A. stephensi:</td>
<td></td>
</tr>
<tr>
<td>DELHI</td>
<td>India</td>
</tr>
</tbody>
</table>

cloned into Bluescript and was denoted as BEO17.916 (see fig. 2). Probes were prepared by isolating both 1.7-kb restriction fragments from the Bluescript vectors by fractionation on a 1.2% low-melting-temperature agarose gel (Seaplaque; FMC). The fragments were labeled with (\(^{32}\)P)dCTP (6,000 Ci/mmol; New England Nuclear) by random primer extension (Feinberg and Vogelstein 1984).

Southern Analysis

Single mosquito DNA was digested with at least 5 units of each restriction enzyme for at least 2 h according to manufacturer’s directions. Where possible, double digestions were performed simultaneously in the recommended buffer; otherwise they were performed sequentially in separate buffers. Digests were fractionated on 0.9% agarose gels and transferred to GeneScreen Plus (New England Nuclear). Membranes were hybridized, washed, and autoradiographed according to methods described elsewhere (Besansky 1990).

Densitometry

Differences in the abundance of the two variants within a given strain were estimated using Whole Band Analysis software with the Visage 2000 Image Analysis System (BioImage, a Kodak Co.). Optical densities for each band in a lane of the autoradiograph were measured, and the percentage of the total represented by each of the predominant bands within the lane was calculated.

Construction and Sequence Determination of Clones

Sixteen of the 21 sequences were cloned using the polymerase chain reaction (PCR) in a DNA thermal cycler (Perkin Elmer–Cetus). Template for the reactions was 1/100 of the total genomic DNA extracted from a single G3 specimen. Two oligonucleotide primers were synthesized for use in each reaction. Their sequence is
based upon the T1 consensus sequence (Besansky 1990). Primer 1, 5' GT(GGATCC)TTTCTCTCAGCTTCTCCT 3', contains a BamHI restriction-site linker (within the parenthesis). It anneals to T1 homologous sequences 176 bp upstream from H3 (see fig. 2). Primer 2, 5' GT(GTCGAC)GGACGAGGATTAACGTGAGCA 3', contains a SalI restriction-site linker. It anneals to T1 homologous sequences 57 bp downstream from H4. The primers are oriented to amplify the 446-bp target sequence between them. In each of two 100-μl PCR reactions, 100 ng of each primer was used. PCR was performed with reagents provided in the GeneAmp DNA Amplification Reagent Kit (Perkin Elmer–Cetus) by using the recommended conditions: 1 X reaction buffer, 200 μM each dNTP, 2.5 units DNA polymerase in a 100-μl volume. Denaturation was carried out at 94°C for 1 min, annealing was at 37°C for 2 min, and extension was at 72°C for 2 min for 25 cycles. The amplification products were pooled and purified with glass beads (GeneClean; Bio101). The DNA was then double digested with BamHI and SalI, inserted into Bluescript SK+, and used to transform XL1 Blue (Stratagene) competent cells. Isolates were chosen at random from the resulting T1 "minilibrary."

The remaining five sequences were determined from either M13mp19 (L34 and 31b) or Bluescript (52a, 914a, and 916e) subclones from phage of the same designations.

Sequences were determined using the dyeoxy chain termination method (Sanger et al. 1977) using reagents supplied in the Sequenase DNA Sequencing Kit (United States Biochemical). The programs Lineup and Pretty from the GCG Sequence Analysis Package, version 6, of the University of Wisconsin (Devereux et al. 1984) were used to create and display the sequence alignments.

Results

Restriction Analysis of T1 Subfamilies in Anopheles gambiae

In the first set of experiments, a consensus restriction map of T1 from A. gambiae was established. Genomic DNA was singly or doubly digested with every combination of a battery of six restriction endonucleases and was fractionated on an agarose gel, blotted, and probed sequentially with three internal fragments of T1. Examples probed with BSH17.31b or BEO17.916 are shown in figure 1. For any given enzyme or enzyme combination, a strongly hybridizing band indicated multiple T1 copies containing the same restriction sites. These restriction sites are conserved, consensus sites. In many cases, the strong signals corresponded to fragment sizes predicted from the restriction map of the clone from which the T1 probe was isolated. In other cases, however, bands were present that could not be reconciled with the probe map. Taken together, the results indicate the presence of two consensus restriction maps. The fragments detected by both probes, shown in relation to the restriction maps, are presented in figure 2.

Consider the hybridization profile obtained by probe BSH17.31b (fig. 1a). A SalI/HindIII codigestion results in a doublet, S-H3 and S-H4, rather than in S-H4 alone, which would have been expected on the basis of the map of the probe (see fig. 2). Similarly, a HindIII digestion produces H2-H3 and H2-H4 doublets instead of H2-H4 alone. Furthermore, after double digestion with HindIII/EcoRI, E2-H3 is observed as well as H2-H4. This suggests that sites E2 and H3 occur together but are only found in some of the repeats. Double digestion with HindIII/XbaI produces X1-H4, H2-H3, and, in much lower frequency, X1-H3. Thus, X1 and H3 do not often occur in the same repeat. Moreover, while a band is detected after SalI/XbaI (S-X2) and EcoRI/
Evolution of Tl Family

XbaI (E2-X2) codigestions, none is detected with XbaI alone. This implies that T1 repeats do not contain both X1 and X2.

Digestion with XbaI/XhoI and hybridization to BE017.916 (fig. 1b) produces a major O2-X2 fragment rather than O2-X1, which do not coexist in T1 repeats (fig. 2). Digestion with EcoRI/XhoI and hybridization to the same fragment detects O2-E2. The results of these two double digestions indicate that T1 repeats do not contain both E2 and X1. Double digestions with EcoRI and either Sall or XbaI demonstrate that E1 never occurs without E2, since only E1-E2 is detected instead of E1-S or E1-X1, respectively. The information provided by this analysis has been resolved into two similar but distinct restriction maps. These maps differ at several sites along the entire length of T1, yet almost no examples of recombination between them were observed. These two nonoverlapping populations of T1 repeats in the A. gambiae genome have been designated T1α and T1β.

Sequence Analysis of T1 Subfamilies in A. gambiae

In a second set of experiments, nucleotide sequences were determined from a 349-bp region of 21 T1 isolates. The region, which includes the conserved H4 HindIII site and the polymorphic H3 HindIII site, was PCR amplified from A. gambiae G3 DNA and directly inserted into a Bluescript plasmid (see Material and Methods). Of the 23 clones selected from this T1 “minilibrary,” one lacked an insert and another six contained anomalous inserts. DNA from the 16 remaining clones was digested with HindIII, a site expected in the insert but not in the vector (having been eliminated from the Bluescript polylinker during cloning). Seven clones contained both H3 and H4 (T1α), and nine contained only H4 (T1β). Nucleotide sequences were determined from five additional isolates selected previously from the G3 phage library (Besansky 1990). The 349 nucleotides sequenced from 11 T1α and 10 T1β isolates are aligned in figure 3.

The overall sequence similarity is quite high, ranging from 93% (25 mismatches) between 12bs and 14bs, 17bs, 18bs, 19bs, and 31b to 100% (no mismatches) between 14bs and 19bs. (With a genomic copy number of 100, it is possible that 14bs and 19bs represent the same element sequenced twice.) Almost all of the existing variation is due to nucleotide substitutions. Only two deletions were found, a 1-bp deletion from 916 and a 6-bp deletion from 13bs. No insertions or other rearrangements were noted.

The sequences presented in figure 3 are arranged to highlight the correlated differences between T1α and T1β. Although each clone has its own “private” mutations, there are 19 base positions diagnostic for each subfamily. On the basis of these coordinated differences, the first 11 sequences belong to T1α, and the remaining 10 belong to T1β. Two clones from T1α, 1bs and 13bs, may be intermediate forms, since they share the diagnostic differences of T1β at eight and five of the 19 positions, respectively. A consensus sequence was calculated for each subfamily separately. The difference between the two subfamily consensus sequences is 5.4%. The average divergence within T1α, on the basis of all 55 pairwise comparisons, is 2.2% (gaps were not included). The average divergence among T1β members (45 comparisons) is 0.8%.

The region sequenced is part of the second open reading frame of T1. Of the 19 base positions distinguishing T1α and T1β, 11 are third-position substitutions, seven are first-position, and one is second-position. These result in only six amino acid replacements, three of which are conservative (fig. 4). This is consistent with the T1α and T1β sequences having evolved under selection for function.
FIG. 1.—Southern analysis of *Anopheles gambiae* (G3) total genomic DNA digested with various restriction enzymes. S = *Sall*; H = *HindIII*; E = *EcoRI*; X = *XbaI*; O = *XhoI*. The blots were probed with (panel a) BSH17.31b and (panel b) BEO17.916. Double digests are indicated by a / (e.g., S/H).

Organization and Structure of the T1 Family in the Sibling Species

The same Southern analysis as described above, using the same probes and panel of enzymes, was applied to the genomic DNA of four sibling species of *A. gambiae* to construct consensus restriction maps of T1 for the complex (fig. 5). The results show that in four of five siblings T1 is divided into two subfamilies corresponding to T1α and T1β. *Anopheles merus* is an exception to this pattern, as it appears to contain only T1β-like sequences. The *A. merus* counterpart to T1α either is absent or is present in such low copy number as to be indistinguishable, by Southern analysis, from individual nonconsensus repeats. This finding, made using a recently established colony from coastal Kenya, was confirmed using a colony from South Africa (data not shown). Although a faint band is detectable at the position where the T1α subfamily would be expected to migrate (see fig. 6), it represents only 2% of the total activity in the lane. At this level it is indistinguishable from other presumably single-copy bands.

From the consensus maps, it appears that T1β has remained essentially unchanged since speciation occurred. Again, *A. merus* is an exception in that the *XbaI* site is
missing. The consensus maps of T1α, on the other hand, are unique for each species, because of the gain or loss of one or two sites.

Comparisons of distribution pattern and copy number were accomplished with a Southern blot containing SalI/BamHI codigestions of total DNA from pairs of mosquitoes in the complex probed with BSH17.31b (fig. 7). Also present on the Southern was the DNA from a congener, A. stephensi. No T1-hybridizing sequences were detected in DNA from this species. The complex patterns of hybridization to the probe by all siblings suggests that copies of T1 are dispersed rather than tandemly arranged in their genomes. Some variation in signal intensity between lanes, particularly between lanes 7 and 9, is due to differences in amounts of DNA extracted from the mosquito pairs. Nevertheless, the signals detected at high stringency confirm that T1 remains conserved at the sequence level among the species. Finally, the similar number of bands detected suggests that copy number is roughly equivalent among the species.

The relative abundance of the two subfamilies was estimated by densitometry of the autoradiograph of HindIII-digested DNA probed with BSH17.31b (fig. 6), on the assumptions that the 2.2-kb canonical band represents T1β and that the 2-kb canonical band represents T1α. While T1α and T1β are almost equally abundant in A. gambiae,
FIG. 2.—Restriction maps of two T1 subfamilies. The map of T1β and the T1β fragments detected by the probes are indicated above; the map of T1α and the T1α fragments detected by the probes are indicated below. Also shown is the map and subfamily origin of the probes, BEO17.916 (Ⅲ) and BSH17.31b (Ⅳ).
T1α is more abundant in *A. arabiensis* and T1β is more abundant in *A. quadriannulatus* (table 2). In *A. melas*, the amounts of both subfamilies are equal, but the proportion of copies present in both major bands of hybridization (only 9%) compared in the remaining detectable copies is much less than the approximately 70% detected in the other species. Although *A. merus* has only one predominant band, it accounts for almost 70% of the detectable hybridization.

**Discussion**

The independence of the T1α and T1β subfamilies is supported by two lines of evidence. The first is consensus mapping by Southern analysis. If T1 were simply highly polymorphic for the presence or absence of restriction sites, every combination of sites should have been observed. This was not the case. Instead, certain restriction sites are correlated such that the presence of one predicts the presence of the others. For example, E1, O2, and X2 cooccur but are not found in combination with X1 (fig. 2). The sequence analysis of T1α and T1β corroborates this result. Not only was sequence conservation greater within than between subfamilies, but the subfamilies also differed at a coordinated set of 19 nucleotide positions. This result is difficult to explain unless each population is independently evolving.

In principle, either gene conversion or retroposition might account for these observations. If gene conversion were primarily responsible, preexisting subfamily members would be corrected from a common source, through the pairing of sequences on homologous or heterologous chromosomes. Gene conversion would operate without altering copy number or location. Although it would be consistent with the observation that copy number has remained relatively constant between strains and species, it would be hard to reconcile with evidence that T1 elements occupy different positions between and even within a species (Besansky 1990). If gene conversion plays a role in the concerted evolution of T1 subfamilies, it cannot be an exclusive one. Gene conversion is unlikely for another reason. It has been documented in yeast that most meiotic conversion tracts are in the range of 1–2 kb, rarely longer (Judd and Petes 1988). In cultured mouse cells, conversion events are even smaller, often <400 bp in length (Liskay and Stachelek 1986). Yet, the consensus map of T1α shows that diagnostic restriction sites are spread over almost 5 kb. Therefore it would take multiple conversion events to correct a single copy. It is not possible to rule out gene conversion on a smaller scale, however. Clones 1bs and 13bs of T1α may have been involved in conversion events with T1β sequences. From the 5' to the 3' end, clone 1bs shares diagnostic positions 4, 6, 9, 11–12, 14, and 16–17 with T1β. Clone 13bs shares positions 1–2, 14, and 16–17 with T1β. Thus these two clones are more divergent from the T1α consensus than are the others.

In contrast to gene conversion, retroposition would amplify and disperse new copies via reverse transcription of an RNA transcript. There is strong circumstantial evidence from sequence data and Southern analysis that the T1 family is mobile (Besansky 1990), though it is not clear how recently movement might have occurred. That T1 propagates by retroposition is suggested by the similarity to reverse transcriptases at the amino acid level, by an A-rich tract at the 3' end, and by its clear relationship with other elements thought to move in the same way, e.g., the L1 elements. A difficulty with retroposition is that overall copy number would continually increase unless balanced by deletions. Deletion following recombination between target-site duplications is an unlikely mechanism for T1, since sequenced isolates lack such duplications. However, unequal crossover between T1 repeats is one possible solution, even though
FIG. 3.—Nucleotide sequence of T1 isolates from *Anopheles gambiae* (G3). The bs designation indicates PCR-cloned isolates. The consensus sequence is based on the most common nucleotide at each position. Only positions that differ from the consensus in each isolate are indicated. Dashes indicate gaps introduced to improve the alignment.
FIG. 3 (Continued)
Fig. 4.—T1α and T1β consensus sequences. Nucleotide substitutions and amino acid replacements of T1β are shown relative to the nucleotide and amino acid sequence of T1α.
Fig. 5.—T1α and T1β restriction maps from five members of the Anopheles gambiae complex. A. g. = A. gambiae; A. a. = A. arabiensis; A. q. = A. quadriannulatus; A. ml = A. melas; A. mr. = A. merus. E = EcoR; O = XhoI; H = HindIII; S = SalI; X = XbaI.
FIG. 6.—Hybridization of BSH17.31b to total genomic DNA from members of the Anopheles gambiae complex digested with HindIII. Lane 1, A. quadriannulatus (CHIL); lane 2, A. melas (BRE); lane 3, A. merus (V12); lanes 4 and 5, A. arabiensis (ARZAG and KISU); lanes 6–9, A. gambiae (G3, GMMK6, AS46, and ZAN). The bars to the left represent the position of HindIII fragments of lambda DNA of 23.6, 9.5, 6.6, 4.3, 2.3, 2.0, and 0.56 kb.

it would be expected to lead to deletion of intervening DNA sequences. Direct evidence of retroposition is lacking. Nevertheless, for the reasons outlined above it seems likely that retroposition, rather than gene conversion, has been the primary mechanism underlying the evolution of T1.

The T1 family is clearly conserved at the sequence level among the sibling species, as judged from the intensity of hybridization to Southern blots. To the extent that
change has occurred, it seems to be nonrandom. The species-diagnostic differences that are features of the T1α consensus restriction maps occur at the extreme 5' or 3' ends of the element, just outside of what would be the long open reading frames in *Anopheles gambiae*. It is tempting to speculate that this is due to relaxation of selection for protein function in the noncoding regions. For the 349-bp coding region compared between T1α and T1β within *A. gambiae*, nucleotide substitutions appear to be evenly distributed, but at the amino acid level silent substitutions outnumber replacements 13 to 6. If this were representative of the entire coding portion of the elements, it would imply that they are evolving under selection for protein function. That the T1α subfamily has remained functional since speciation occurred is evidenced by the spread of diagnostic changes throughout each species. The existence of T1α subfamilies with species-specific consensus maps probably reflects the activity of a single progenitor element that sustained the species-diagnostic mutations. Their existence also refutes the argument that the conservation of T1 among the siblings may simply be due to recency of the radiation or introgression between species.
Unlike $T_{1\alpha}$, interspecific comparisons of $T_{1\beta}$ consensus maps revealed no change except for the absence of an $XbaI$ site in $A. merus$. However, the apparent conservation of $T_{1\beta}$ restriction maps among species may be misleading. There is a paucity of restriction sites in $T_{1\beta}$ relative to $T_{1\alpha}$, particularly at the 5' and 3' ends, precluding a valid comparison at the interspecific level. On the other hand, on the basis of the sequenced regions from $A. gambiae$, the average divergence within $T_{1\beta}$ is only 0.8% compared with 2.2% within $T_{1\alpha}$. It is possible that the method of cloning by PCR preferentially sampled more recent, less divergent copies and that the bias was more pronounced for $T_{1\beta}$ than for $T_{1\alpha}$, since more sequences in this group were cloned by PCR. The greater divergence of $T_{1\alpha}$ is probably not a cloning artifact, however, since the five sequences not cloned by PCR are no more divergent than are those that were. One possible explanation for the smaller level of variation might be that the $T_{1\beta}$ subfamily has expanded more recently from a common source. Other, not necessarily exclusive possibilities are different rates of mutation or homogenization. Differences in the rate of homogenization, whether by gene conversion or retroposition, may also explain fluctuations in the relative proportions of subfamilies within species.

The evolutionary explanation for the absence of $T_{1\alpha}$ from $A. merus$ is uncertain. $A. merus$ may have branched off before the appearance of $T_{1\alpha}$, which would place it closer to the ancestor of the complex. On the other hand, $T_{1\alpha}$ may once have occupied the $A. merus$ genome and subsequently been lost. The current hypothesis concerning the phylogeny of the $A. gambiae$ complex is based on fixed paracentric inversions (Coluzzi et al. 1979). This hypothesis regards $A. quadriannulatus$—not $A. merus$—as nearest to the ancestral $gambiae$, on the basis of both its central placement in the chromosomal sequence and available bionomic data. Although the chromosomal evidence does not rule out an ancestor nearer to $A. merus$, which is at one end of the unrooted chromosomal sequence, that $A. quadriannulatus$ is ancestral is supported by a study of the chromosomal locations of ribosomal DNA in the siblings (S. M. Paskewitz and F. H. Collins, personal communication). This interpretation should be approached cautiously, however, because of the incidence of introgression between sibling species in areas of sympatry (White 1970). The alternative proposal suggests that $T_{1\alpha}$ might have been lost from the $A. merus$ genome, where "loss" refers to divergence beyond the sensitivity of genomic consensus mapping rather than to exact excision and disappearance. It is difficult to imagine the relatively rapid loss of a multicopy dispersed sequence. However, by analogy with the $Alu$ and $L1$ families, each subfamily may have only one or a few conserved source genes. The more numerous pseudogene copies would not be detectable by genomic consensus mapping once the source gene became inactive and the pseudogenes diverged.
Tlα from *A. merus* may even have lost function before having generated any pseudogene copies. Molecular data on evolutionary rates among *Anopheles* species are lacking. However, DNA reassociation and sequence studies from closely related *Drosophila* species suggest that rapidly evolving components of the genome contribute to overall rates of nucleotide substitution as high as 10%/Myr (Hunt et al. 1981; Martin and Meyerowitz 1986; Caccone et al. 1988). Rates of divergence for Tlα pseudogenes would have to be higher yet to account for their degeneration within the estimated time since speciation. It is interesting that, in spite of habitats on opposite coasts of Africa and in spite of considerable divergence based upon chromosomal inversions, *A. merus* and *A. melas* can be clustered together on the basis of (1) a common preference for breeding in saltwater and (2) rDNA and mtDNA restriction maps that are more similar to each other than to those of the other members of the complex (S. M. Paskewitz and F. H. Collins, unpublished data). That *A. melas* has very low levels not only of Tlα but also of Tlβ supports this clustering and favors the “loss” hypothesis. The resolution of this issue must await comparisons with a close relative outside the *A. gambiae* complex.

**Acknowledgments**

I am grateful to Frank Collins of the Centers for Disease Control, Malaria Branch, for providing the mosquito colonies and the laboratory in which this work was carried out and to Walter M. Fitch for many helpful comments on the manuscript. I thank Brian Holloway and the staff of the Centers for Disease Control, Biotechnology Core Facility, for synthesizing the sequencing primers.

**LITERATURE CITED**


ALAN M. WEINER, reviewing editor

Received October 5, 1989; revision received December 19, 1989

Accepted January 10, 1990