The Origin and Evolution of Mosquito APE Retroposons

James L. Crainey, Clare F. Garvey, and Colin A. Malcolm
School of Biological Sciences, Queen Mary, University of London, London, United Kingdom

The detection of horizontal transfer is important to understanding the origin and spread of transposable elements and in assessing their impact on genetic diversity. The occurrence of the phenomenon is in doubt for two of the three major groups of elements, but is disputed for retroposons, largely on the grounds of data paucity and overreliance on divergence estimates between host species. We present here the most wide-ranging retroposon data set assembled to date for a species group, the mosquitoes. The results provide no evidence for horizontal transfer events and show conclusively that four previously reported events, involving Juan-A, Juan-C, T1, and Q, did not occur. We propose that the origin of all known mosquito retroposons can be attributed to vertical inheritance and that retroposons have therefore been a persistent source of genetic diversity in mosquito genomes since the emergence of the taxon. Furthermore, the data confirm that the unprecedented levels of retroposon diversity previously reported in Anopheles gambiae extends to at least seven other species representing five genera and all three mosquito subfamilies. Most notably, this included the L1 elements, which are not known in other insects. A number of novel well-defined monophyletic groups were also identified, particularly, JM2 and JM3 within the Jockey clade, which included sequences from seven and five mosquito species, respectively. As JM3 does not contain an Anopheles element, this represents a good example of stochastic loss and the best out of at least four found in this study. This exceptionally diverse data set when compared with the wealth of data available for the many unrelated species with which mosquitoes have intimate contact through blood feeding ought to be fertile ground for the discovery of horizontal transfer events. The absence of positive results therefore supports the view that retroposon horizontal transfer does not occur or is far more exceptional than for other types of transposable elements.

Introduction

Transposable elements are an important source of genetic diversity and are, for example, thought to be responsible for over half of spontaneous mutations in Drosophila melanogaster laboratory strains (Arkhipova and Meselson 2005). They are, however, not always active and some classes may even be absent in some species, such as rotifers (Arkhipova and Meselson 2000). The mechanisms that sustain active elements in the genome and determine their gain or loss are therefore of considerable interest. Transposable elements are inherited vertically but on rare occasions may transfer horizontally from one species to another. Retroposons, also called non–long terminal repeat (LTR) retrotransposons, or long interspersed nuclear elements (LINEs), would appear to be no exception because there are many reports of horizontal transfer events (Mizrokhi and Mazo 1990; Mouches, Bensaadi, and Salvador 1992; Agarwal et al. 1993; Kordis and Gubensek 1995, 1998, 1999a, 1999b; Drew and Brindley 1997; Zupunski, Gubensek, and Kordis 2001; Kapitonov and Jurka 2003). Nevertheless, none of these are as convincing as those found in the two other major groups of elements, and it has been argued that retroposons are unable to transfer horizontally (Malik, Burke, and Eickbush 1999; Eickbush and Malik 2002). If the latter is true, a retroposon in a contemporary species predicts a progenitor in ancestors, which in turn indicates a long history of retroposon activity because this is essential to their survival and presumably a long-term impact on genetic diversity.

Retroposons divide into two types based on the proteins coded for, which may include either a restriction endonuclease (RE)–like domain or an apurinic-apyrimidinic endonuclease (APE)–like domain. The RE retroposons are older (Malik, Burke, and Eickbush 1999; Eickbush and Malik 2002).

The inheritance of two retroposition families, R1 (an APE retroposition) and R2 (an RE retroposition), which insert specifically into ribosomal RNA genes is widely accepted to have been vertical since the emergence of arthropods (Eickbush and Malik 2002). This is because the elements’ phylogenies are congruent with predicted host phylogenies, and inferred protein sequences show similar divergence rates (Burke et al. 1998; Malik, Burke, and Eickbush 1999). Focussed studies of this type are more difficult on retroposons that do not have a strong integration-site preference, which includes most APE retroposons. Where a large data set is not available the analyses rely more heavily on comparisons of protein divergence rates, which suffer from the problem that host taxa divergence estimates often vary. For example, protein evolution rates calculated for the mouse and rat L1 retroposons have been used to argue both their horizontal and vertical transmissions, with one group calculating rates based on the taxa sharing a common ancestor 15 MYA (Zupunski, Gubensek, and Kordis 2001) and another group using an estimate of 40 Myr (Malik, Burke, and Eickbush 1999; Eickbush and Malik 2002).

This difference in the interpretation of available data has lead Gubensek and colleagues to propose multiple retroposition horizontal transfer events, of which perhaps the most marked example is their contention that vertebrate RTE elements require at least two horizontal transfer events to explain their taxon distribution, one from vipers into ruminants and one into marsupials (Kordis and Gubensek 1998, 1999a, 1999b; Zupunski, Gubensek, and Kordis 2001). Eickbush and colleagues have disputed this and have repeatedly argued that retroposition inheritance may be strictly vertical (Malik, Burke, and Eickbush 1999; Malik and Eickbush 2000; Burke et al. 2002; Eickbush and Malik 2002).

In mosquitoes, there are multiple reports of horizontal transfer of transposable elements, including four involving...
retroposons (Mouches, Bensaadi, and Salvado 1992; Agarwal et al. 1993; Robertson and Lampe 1995; Rongnoparut et al. 1998; Kapitonov and Jurka 2003). Interestingly, the evidence for the retroposon cases was not actually based on divergence rates. For the Juan elements horizontal transfer was argued because of their within-family identity in both the Aedes aegypti and Culex pipiens genomes, which it was proposed indicated the absence of a common ancestor (Mouches, Bensaadi, and Salvado 1992; Agarwal et al. 1993). For the Anopheles gambiae elements, Q and T1, which had no known close relatives in arthropods, it was proposed that the elements had conserved esterase-like and homeo-like domains to facilitate horizontal transfer (Kapitonov and Jurka 2003).

The parameters involved in horizontal DNA transfer are generally not well understood. The LTR retrotransposon Gypsy is known to be able to enter and transpose in the D. melanogaster germ line if its infectious particles are mixed into larval feed (Song et al. 1994), but no such infectious particles have yet been reported for retroposons. Long periods of close proximity between foreign DNA and a recipient germ line are clearly one prerequisite for horizontal transfer. In mosquitoes this is facilitated by blood feeding, a trait found in nearly all three thousand plus contemporary species and dates back to the emergence of the taxon, an estimated 210 MYA (Rai and Black 1999). The mosquitoes ingest host blood to activate and provide for egg development, and as a consequence they also take up blood parasites. Given the evolutionary distance between mosquitoes and their blood meal hosts and blood parasites, horizontally transferred retroposons should be conspicuous. As mosquito hosts include humans and commercially, or medically, significant organisms, there are large amounts of publicly available DNA sequence and thus a large pool of retroposon sequences.

The work presented here is a survey of the representation of APE retroposons in mosquitoes. It provides a large data set to examine the previously proposed horizontal transfer events and to search for novel events. Two approaches were taken: one based on polymerase chain reaction (PCR) cloning and sequencing and the other based on bioinformatics. The survey demonstrates that mosquitoes possess elements from nearly all the previously described major clades and that close relatives of these elements can be found in species from five diverse genera. This includes Juan and T1 and Q and therefore contradicts the arguments for horizontal transfer of these elements.

Materials and Methods

PCR Survey

The following mosquito species were used in the study: Ae. aegypti (Ae Ae), An. gambiae (KWA), Anopheles stephensi (M2), and Toxorhynchites brevipalpis were from laboratory colonies originally maintained at the London School of Hygiene and Tropical Medicine; Aedes albopictus was collected in Malaysia; Anopheles sinensis was from a laboratory colony originally maintained at the Institute of Parasitic Diseases, China; and Culiseta annulata and Culex pipiens pipiens were collected in Stepney Green, East London.

The PCR-amplified region of the retroposon includes about 400 nt from just downstream of the 5’ end of the reverse transcriptase domain and is the most conserved part of all retroposons (Malik, Burke, and Eickbush 1999). As this region is sufficient to identify a retroposon it is referred to here as a retroposon sequence tag (RST).

The PCR primers were Jockey forward: AGYTAYCG-NCCNATH; CR1 forward: AAYTAYCNGNATHAC; CR1 and Jockey reverse: RCTNCCYTNGNGNAC; L1 forward: AA(C/T)TG(C/A)GNCCAT(T/A/C)/TC; and L1 reverse: (G/A)CTNC(C/T)/TGNGNAC. The four-step PCR program was as follows: (1) 5 min at 94°C; (2) 5 rounds of 90 s at 94°C, 90 s at 35°C, and 90 s at 72°C; (3) 30 rounds of 90 s at 94°C, 90 s at 45°C, and 90 s at 72°C; and (4) 10 min at 72°C. The three-step amplification program was: (1) 5 min at 94°C; (2) a variable number of rounds of 90 s at 94°C, then 90 s at 45°C, and 90 s at 72°C; and (3) 10 min at 72°C. All PCR reaction mixes were prepared in 50 µl total volume and contained 50 pmol of each primer; 1 mM MgCl2; 0.25 mM deoxynucleoside triphosphates; 5 µl of 10× NH4 buffer (160 mM (NH4)2SO4, 670 mM Tris-Cl [pH 8.8 at 25°C], and 0.1% Tween-20); 0.5 units of Taq; and 100 ng of template DNA. For negative controls template DNA was replaced with sterile nanopure water.

The PCR-amplified products (between 350 and 500 bp) were cleaned using the QIAIEN PCR clean-up kit and 150 ng ligated with 50 ng of GEM-T easy vector (Promega, Southampton, UK) following the manufacturer’s protocols. The ligation reactions were used to transform 50 µl of commercial Escherichia coli DHα competent cells, as directed by suppliers (Invitrogen, Paisley, UK). Positive colonies were picked and screened for the appropriate-sized insert using the PCR reamplification protocol as above. Initially, six PCR-positive clones from each ligation were sequenced. These data were then used to direct restriction enzyme digestions of PCR fragments from the remaining clones. Up to a further six clones were selected for sequencing based on digest results that were different from the first six. This was then repeated several times depending on the size of the library and diversity present. In total over 200 clones were sequenced.

Database Survey

The RST region from previously published sequences and sequences isolated from the PCR survey were used in Blast searches of public database deposits at the U.S. National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) and the Bioinformatics Center at Kyoto University’s Institute for Chemical Research (http://www.genome.ad.jp). The RST region contains 7 of the 15 residue positions conserved throughout retroposons (Malik, Burke, and Eickbush 1999). Only database sequences that included all seven of these residues were collected. Database sequences that contained stop codons or required frameshift mutations were also not collected. RST sequences generated from these searches were then used as query sequences in subsequent searches.

Sequence Analysis

Nucleotide sequences generated from the PCR survey were analyzed and edited using Chromas (McCarthy 1996);
Phylogenetic Analysis

Allocation of sequences into families was based on nucleotide identity. The inferred amino acid sequence from a representative sequence of each family generated from the PCR and bioinformatic surveys were then added to previously published data. Phylogenetic trees were prepared from this alignment using version 3.6 of the PHYLIP software package (Felsenstein 2002). Maximum parsimony trees were constructed using PROTPARS, using Jones-Taylor-Thornton amino acid change frequencies. Neighbor-joining trees were inferred from a distance matrix produced in PROTDIST, also assuming Jones-Taylor-Thornton amino acid change frequencies. The stability of both tree types was assessed with 1,000 bootstrap pseudoreplicates created using SEQBOOT. The trees are shown with the representative R2 clade retroposons as an out-group. The R2 clade has been repeatedly grouped as the most derived clade of the RE elements, and the RE retroposons have repeatedly been shown as ancestral to the APE retroposons and therefore were considered a good out-group (Malik, Burke, and Eickbush 1999; Malik and Eickbush 2000; Burke et al. 2002; Eickbush and Malik 2002).

Results

In total, 126 unique sequences were PCR amplified from within the retroposon reverse transcriptase domain (referred to here as RSTs) (European Molecular Biology Laboratory [EMBL] database accession numbers: AJ970181–AJ970306). Based on the analysis of the nucleotide data the sequences were grouped into 53 distinct families, in which no two members were less than 92% similar or more than 83% similar to a sequence belonging to a different family. Fourteen sequences contained internal deletions, of which nine were from Ae. aegypti, two from Cx. pipiens, two from Cs. annulata, and one from T. brevipalpis. In Ae. aegypti clones, these deletions meant that the average RST sequence was 6% shorter than the family consensus sequence. Database searches identified 59 An. gambiae and 52 Ae. aegypti retroposon families that were not represented in the PCR survey, as well as over 150 retroposon families from nonmosquito taxa.

Phylogenetic Analysis

As an initial test of the utility of the RST region in phylogenetic analysis, RSTs from the same retroposons used in an analysis based on the full-length reverse transcriptase domain by Malik, Burke, and Eickbush (1999) were used here. The topologies of the two trees were very similar in their major features (results not shown). A representative neighbor-joining tree inferred from the alignment of all 373 APE RSTs (providing as Supplementary Material online) is shown in figure 1. To simplify the figure, monophyletic groups that did not contain mosquito retroposons were truncated and labeled only with the number of elements they contain and a taxon name appropriate to the range of species represented. Mosquito elements are labeled with a number and an abbreviated species name, corresponding to an amino acid sequence appearing in the EMBL deposited alignment.

Ten of the 11 APE retroposon phylogenetic lineages previously described as clades are also generated in this tree (Malik, Burke, and Eickbush 1999; Malik and Eickbush 2000; and Burke et al. 2002). The only difference is that the Rex-1 group is nested within the metazoan lineage of the CR1 clade and not outside (Eickbush and Malik 2002). There are in addition a number of distinct, >50% bootstrap-supported, monophyletic groups that meet the criteria previously proposed for the definition of a retroposon clade (Malik, Burke, and Eickbush 1999). Two such mosquito exclusive monophyletic groups, Outcast and Loner, have been reported previously (Biedler and Tu 2003). Five more are simply labeled 1–5, in line with more recent nomenclature proposals (Eickbush and Malik 2002) that elements should be grouped into super clades or superfamilies.

Mosquito Jockey Clade Elements

The mosquito retroposons in the Jockey clade, representing all eight mosquito species examined, form three distinct monophyletic groups, JM1, JM2, and JM3, which together make up the largest of the three >50% bootstrap–supported, monophyletic groups (fig. 1). JM1 contains the previously reported Juan-A and Juan-C, as well as An. gambiae and Ae. albopictus elements added by this survey. JM1 has short branch lengths (<40% divergence) and represents three genera. It is thus likely to be limited to the Anopheinae subfamily.

JM2 contains 18 distinct, hitherto unreported, families of retroposons. Within the available mosquito data, this group should contain the best prospect for a horizontal transfer event because the genetic distances are all very short. However, consistent with the divergence of their host species groups, around 120 MYA, the anoheline and culicine retroposons form separate, high-bootstrap–supported, monophyletic groups. The lone Toxorhynchites element appears ancestral to the culicine elements, which is consistent with the current view that these two mosquito taxa are younger than the anophelines.

The JM3 elements are also novel. This group contains four retroposon families representing three genera of culicines. No two elements are more than 40% diverged and as might be expected the single family from Toxorhynchites is the most distantly related. There are no retroposons from anophelines, which given the near 100% sequence coverage of the An. gambiae genome, suggests it is unlikely that any will be found and that this is an example of stochastic loss. It is also notable that the chironomid element NLRCh1 outgroups JM3 with good bootstrap support yet with the shortest branch lengths between any mosquito monophyletic group and a nonmosquito retroposon family element.

The group labeled III has previously been reported as Loner (Biedler and Tu 2003). None of the An. gambiae elements belonging to this group were found in the PCR survey, and thus related elements from other mosquitoes would not be expected with the primers used.
Mosquito CR1 Clade Elements

The CR1 clade now contains 52 mosquito retroposons nested within two, bootstrap-supported, dipteran-exclusive monophyletic groups, labeled IV and V (fig. 1). The *Ae. gambiæ* Q and T1 elements proposed to be functionally adapted for horizontal transfer are indicated and nested within IV, which is represented with seven mosquito species representing five mosquito genera and three mosquito subfamilies. The T1 element is contained within a bootstrap-supported, monophyletic group within group IV. This group contains sequences from two *Ae. stephensi* retroposons isolated in this PCR survey, which, based on vertical inheritance, dates the group to before the divide of the *Neocellia* and *Pyretophorus* series of the *Cellia* subgenus; however, the branch lengths are long and may predate the anopheline culicine divide.

FIG. 1.—Figure 1 shows a representative neighbor-joining tree inferred from the RST region of 373 APE retroposons. The tree is displayed in two parts, separated at the node marked with a closed circle. Branches that are limited to mosquito elements are black; all other branches are in gray. The tree is rooted with the R2 clade RE retroposons as these elements have repeatedly been shown to be the APE retroposons’ closest relatives (Malik, Burke, and Eickbush 1999; Eickbush and Malik 2002). Monophyletic groups that do not contain mosquito representatives are truncated and labeled only with a taxon name, describing the narrowest possible taxon range of the group. These groups are also labeled with a number (in brackets) indicating the number of elements they contain. Previously described retroposon clades are indicated with lower case roman numerals and names. Mosquito monophyletic groups referred to in the text are labeled accordingly with uppercase roman numerals or names. Bootstrap-supported nodes, found on the 50% majority rule consensus tree, discussed in the text have their support shown as a percentage of 1,000 pseudo replicates. All mosquito retroposons are labeled with an abbreviated taxon name. Those that have been described previously are labeled.
FIG. 1. (Continued)
Two other CR1 groups contain An. gambaie and An. sinensis elements, therefore predating the divergence of the Anopheles and Cellia sub genera, but the branch lengths are short suggesting radiation after the divergence of the Anopheinae and Culicinae mosquito subfamilies.

Divergence Analysis

To present a simplified illustration of the congruence of the host and retroposon phylogenies evident in figure 1, the same data were used to make a series of comparisons between taxon groups. Figure 2 shows graphically how divergence of related retroposons increases as host species become more distantly related, without the need for specific estimates of the age of host divergence. The figure displays five sets of element comparisons: (1) elements from different culicine genera; (2) culicines versus Anopheles; (3) mosquito versus nonmosquito dipterans; (4) mosquito versus nondipteran neopterans; and (5) mosquito versus vertebrates. The data sets contained the following number of pairwise comparisons: (1) 431; (2) 1,278; (3) 802; (4) 639; and (5) 551. Sequence comparisons that had distance values greater than those observed between the closest related mosquito and yeast C. albicans retroposons were discarded, and the remaining data were expressed as a fraction of the whole data set.

Two other CR1 groups contain An. gambaie and An. sinensis elements, therefore predating the divergence of the Anopheles and Cellia sub genera, but the branch lengths are short suggesting radiation after the divergence of the Anopheinae and Culicinae mosquito subfamilies.

Divergence Analysis

To present a simplified illustration of the congruence of the host and retroposon phylogenies evident in figure 1, the same data were used to make a series of comparisons between taxon groups. Figure 2 shows graphically how divergence of related retroposons increases as host species become more distantly related, without the need for specific estimates of the age of host divergence. The figure displays five sets of element comparisons: (1) elements from different culicine genera; (2) culicines versus Anopheles; (3) mosquito versus nonmosquito dipterans; (4) mosquito versus nondipteran neopterans; and (5) mosquito versus vertebrates. The data sets contained the following number of pairwise comparisons: (1) 431; (2) 1,278; (3) 802; (4) 639; and (5) 551. Sequence comparisons that had distance values greater than those observed between the closest related mosquito and yeast C. albicans retroposons were discarded, and the remaining data were expressed as a fraction of the whole data set.

Two other CR1 groups contain An. gambaie and An. sinensis elements, therefore predating the divergence of the Anopheles and Cellia sub genera, but the branch lengths are short suggesting radiation after the divergence of the Anopheinae and Culicinae mosquito subfamilies.
mosquito species monophyletic groups are most closely related to an element from a species other than mosquito. In each case the sequence divergence is sufficiently large to consider this to be the consequence of data paucity. Consistent with this and with the strict vertical inheritance of mosquito retrotransposons, the divergence analysis (fig. 2) shows that element divergence increases with taxon divergence. Importantly too, it also shows clear baseline values for taxon comparisons that increase with the taxon comparisons, which can also be seen with mosquito element comparisons.

Fossil records suggest that mosquitoes evolved by the Jurassic (206–142 MYA) and the Culicinae subfamily by the Eocene, which began 60 MYA (Rai and Black 1999). Although it is generally accepted that the Anophelinae subfamily is ancestral to both the subfamily Culiciniae and the genus Toxorhynchites, the relationship between the latter two is unclear, and it may even be that Toxorhynchites is nested within the subfamily Culiciniae (Harbach and Kitching 1998; Mitchell, Sperling, and Hickey 2002). Consistent with the close relationship between genus Toxorhynchites and subfamily Culiciniae and with the subfamily Anophelinae as the most ancient subfamily, baseline similarity comparisons made between elements from the Toxorhynchites and Culiciniae are greater than those made between either taxon with elements from the Anophelinae. Comparisons of this sort may prove useful in future studies to test the likely significance of any particular monophyletic group containing a retroposon sequence from a species more distantly related than the others represented in the group and do not suffer from ties from divergence estimates that previous analysis of this sort have.

Mosquito Retroposon Stochastic Loss

With close to 100% genome coverage in the An. gambiae genome (Holt et al. 2002) and strong support for strict vertical inheritance of retroposons, it is possible to make predictions of retroposon stochastic loss. The targeted database searches performed here and more general searches performed by Biedler and Tu (2003) produced no evidence for the existence of An. gambiae elements in four lineages of RSTs derived exclusively from dipterans and containing sequences from other mosquitoes (fig. 1). While there is perhaps still some scope to find these missing elements the absences are most likely to be due to loss prior to the emergence of An. gambiae. The JM3 group is the clearest example because it contains sequences from four genera representing two mosquito subfamilies and the Chironomidae.

Implications for Mosquito Genome Evolution

The data collected here strongly support the vertical inheritance of Juan-C and Juan-A elements but does not entirely resolve the apparent absence of ancestral elements in the Culex and Aedes genomes. These may exist as unrecognizable relics, but it is also possible that nonfunctional elements are lost rapidly from the genome. Clues to the mechanics of this phenomenon might also have been provided here.

Despite actively excluding short DNA fragments from the PCR survey, 27 RST clones with internal deletions were still sequenced and none were found with conspicuous insertions. For the Ae. aegypti RST clones, the inclusion of deleted sequences produces an average sequence length 6% shorter than the consensus. In a broad variety of Drosophila species a genome-wide deletional bias has been reported and has been argued to play a key role in genome size determination (Petrov, Lozovskaya, and Hartl 1996; Petrov and Hartl 1998; Petrov et al. 2000; Petrov 2002). Intriguingly, the bias observed here was restricted to culicines, suggesting that loss from these taxa occurs at a higher rate than from anophelines, despite the fact that culicine genomes are up to eight times larger. A more consistent explanation is that the larger mosquito genomes, with a much higher proportion of repetitive DNA, are more dynamic and result in more frequent deletion events and therefore more rapid loss of inactive retroposons. The deletions could be balanced by amplification of other sequences, including the spread of active derived retroposons.

Conclusions

In conclusion, the data presented here are consistent with and support the strict vertical inheritance of all hitherto discovered mosquito APE retroposons, including those previously proposed to have arrived by between-species horizontal transfer. The data also show that the hitherto unprecedented levels of APE retroposon diversity discovered in An. gambiae also exist in Ae. aegypti and are likely to be common in all mosquito taxa. Together these findings show that APE retroposons have been a persistent source of genetic diversity in mosquito genomes since the emergence of the taxon.

Supplementary Material

Supplementary Material is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

C.F.G was funded by a Collaborative Awards in Science and Engineering Biotechnology and Biology Sciences Research Council studentship. We thank Paul Ready of the Natural History Museum for his contribution to the preliminary stages of this work and the design of the original PCR primers.

Literature Cited


Richard Thomas, Associate Editor

Accepted July 4, 2005