Multiple Lineages of R1 Retrotransposable Elements Can Coexist in the rDNA Loci of Drosophila

Karen L. Gentile,1 William D. Burke, and Thomas H. Eickbush

Department of Biology, University of Rochester

R1 non–long terminal repeat retrotransposable elements insert specifically into the 28S rRNA genes of arthropods. One aspect of R1 evolution that has been difficult to explain is the presence of divergent lineages of R1 in the rDNA loci of the same species. Multiple lineages should compete for a limited number of insertion sites, in addition to being subject to the concerted evolution processes homogenizing the rRNA genes. The presence of multiple lineages suggests either the ability of the elements to overcome these factors and diverge within rDNA loci, or the introduction of new lineages by horizontal transmission. To address this issue, we attempted to characterize the complete set of R1 elements in the rDNA locus from five Drosophila species groups (melanogaster, obscura, testacea, quinaria, and repleta). Two major R1 lineages, A and B, that diverged about 100 MYA were found to exist in Drosophila. Elements of the A lineage were found in all 35 Drosophila species tested, while elements of the B lineage were found in only 11 species from three species groups. Phylogenetic analysis of the R1 elements, supported by comparison of their rates of nucleotide sequence substitution, revealed that both the A and the B lineages have been maintained by vertical descent. The B lineage was less stable and has undergone numerous, independent elimination events, while the A lineage has diverged into three sublineages, which were, in turn, differentially stable. We conclude that while the differential retention of multiple lineages greatly complicates its phylogenetic history, the available R1 data continue to be consistent with the strict vertical descent of these elements.

Introduction

A long-standing question in the study of transposable elements is that of the frequency with which they are transferred between species (reviewed in Kidwell 1993; Hartl, Lohe, and Lozovskya 1997). The most unambiguous evidence of horizontal (also referred to as lateral) events is found when elements from two different species have levels of nucleotide identity too high to be explained by vertical descent of the host genomes. Such evidence for horizontal transmission has been found in two major classes of eukaryotic transposable elements. The largest numbers of cases are found among the DNA-mediated transposons, i.e., those elements that move by a cut-and-paste transfer of their DNA. So numerous are the examples of horizontal transmission for the P and mariner elements that on a broad scale the phylogeny of these elements bears little relationship to the phylogeny of the organisms in which they reside (Robertson 1993; Lohe et al. 1995; Clark and Kidwell 1997). Evidence has also been mounting that the long terminal repeat (LTR) retrotransposons, elements which move by a reverse transcription mechanism similar to that of retroviruses, can jump between species. Recent examples of horizontal transfer have been found for the copia elements in Drosophila (Jordan, Matyunia, and McDonald 1999) and the SURL elements in sea urchins (Gonzalez and Lessios 1999).

Less persuasive is evidence for the horizontal transfer of elements from the third class of eukaryotic transposable elements, the non-LTR retrotransposons. These elements move by polymerizing their reverse transcript directly onto the DNA target site (Luan et al. 1993), a mechanism more similar to group II intron retrohoming (Zimmerly et al. 1995) and telomerase activity (Nakamura et al. 1997) than to retroviral retrotransposition. No evidence has been found for horizontal transmission of the two most extensively studied non-LTR retrotransposable elements, the L1 elements of mammals (Hutchison et al. 1989; Usdin et al. 1995; Verneau, Catze®is, and Furano 1998) and the R1 and R2 elements of insects (Eickbush and Eickbush 1995; Lathe and Eickbush 1997; Burke et al. 1998). While reports of the horizontal transmission of several other non-LTR elements have appeared, a recent analysis of the phylogeny and rates of evolution of non-LTR elements did not reveal unambiguous instances in which the sequence identities were too similar to be explained by descent (Malik, Burke, and Eickbush 1999 and references therein).

R1 and R2 elements insert specifically into the 28S rRNA genes of their host genomes (see fig. 1A). These elements are also detected outside the rDNA locus, but these non-rDNA copies are degenerate (Roita et al. 1981; Browne et al. 1984; Xiong et al. 1988; Adams et al. 2000). In many insects, single homogeneous families of R1 and R2 are found in the rDNA locus. In other instances, however, two or more divergent families of elements can be found in the same species. For example, the parasitic wasp Nasonia vitripennis contains at least four lineages of R1, and the beetle Popillia japonica contains at least five lineages of R2 (Burke et al. 1993). While the presence of highly divergent lineages in the rDNA locus of the same species is not in itself evidence for horizontal transfers, the alternative explanation, vertical descent, would seem unlikely. How could two or more families of R1 and/or R2 coexist over long periods while in competition for the limited number of insertion sites available in the rDNA locus?

To address this question, we extended our analysis of the R1 elements in the well-characterized genus Dro-
**A**  

Drosophila rDNA unit  

![Diagram of a typical Drosophila rDNA unit with a full-length R1 element insertion. Filled black boxes = regions encoding the mature rRNAs; thick line = internal and external transcribed spacer regions; open boxes = 5' and 3' untranslated regions (UTRs) of the R1 element; gray boxes = open reading frames (ORFs) of the R1 element. The 3' half and flanking 28S gene sequences have been expanded to indicate the locations of the seven PCR primers used in this study to identify R1 elements. R1 primers were named after the protein sequences they encode. The nucleotide sequence of each primer can be found in Materials and Methods.](image)

**B**  

D. takahashii R1 lineages

![Sequence similarity between the R1A1 and R1A2 elements of Drosophila takahashii. The R1 ORF2, 3' UTR, and downstream 28S sequences are indicated as in A. Sequence identity between the two R1 lineages is 58% in the region encoding ORF2 but >96% in the 3' UTR. Because the 3' UTRs of R1 elements evolve more rapidly than the ORF regions, a recombination event has been postulated to account for the divergence between these R1 lineages (see text). The nucleotide sequence of the R1A1 and R1A2 elements in the region of this postulated recombination is shown.](image)

Fig. 1.—Locations of R1 elements in the rDNA unit of Drosophila. A, Diagram of a typical Drosophila rDNA unit with a full-length R1 element insertion. Filled black boxes = regions encoding the mature rRNAs; thick line = internal and external transcribed spacer regions; open boxes = 5' and 3' untranslated regions (UTRs) of the R1 element; gray boxes = open reading frames (ORFs) of the R1 element. The 3' half and flanking 28S gene sequences have been expanded to indicate the locations of the seven PCR primers used in this study to identify R1 elements. R1 primers were named after the protein sequences they encode. The nucleotide sequence of each primer can be found in Materials and Methods. B, Sequence similarity between the R1A1 and R1A2 elements of Drosophila takahashii. The R1 ORF2, 3' UTR, and downstream 28S sequences are indicated as in A. Sequence identity between the two R1 lineages is 58% in the region encoding ORF2 but >96% in the 3' UTR. Because the 3' UTRs of R1 elements evolve more rapidly than the ORF regions, a recombination event has been postulated to account for the divergence between these R1 lineages (see text). The nucleotide sequence of the R1A1 and R1A2 elements in the region of this postulated recombination is shown.

Gentile et al.

While previous surveys have revealed a single lineage of R2 in this genus (Lathe and Eickbush 1997), multiple lineages of R1 were identified (Lathe et al. 1995). We report here that at least two of these R1 lineages can be explained by strict vertical descent, with some species retaining a single lineage and other species maintaining both lineages.

**Materials and Methods**

**Species Stocks**

Species stocks from the quinaria and testacea species groups were obtained from John Jaenike. All other species were obtained from the Drosophila stock center at Bowling Green University.

**PCR Amplification and Nucleotide Sequence Determinations**

Genomic DNA was isolated from a mixture of male and female adults of each species as previously described (Eickbush and Eickbush 1995). To insure the cloning of all R1 lineages, three separate oligonucleotide primers complementary to different locations in the R1 second open reading frame (ORF2) were used to PCR-amplify the R1 elements (see fig. 1A). The first primer, AYADD (5’-GCNTWGCNGAYGAY-3’), is complementary to a highly conserved sequence of the reverse transcriptase domain and has been used in several previous studies (Eickbush and Eickbush 1995; Lathe et al. 1995; Burke et al. 1998). The second primer, LTGHG (5’-TTYYTIYTIACIGGICAYGG-3’), was newly designed to be complementary to a conserved sequence located 90 amino acids upstream of the end of the R1 ORF2. The third primer, CLPVCRT (5’-TGYYTITCYCTITGAY/CGIAC-3’), was also newly designed to be complementary to a conserved sequence found in the cysteine-histidine motif approximately 200 amino acids upstream from the end of the R1 ORF2. All three primers were used in combination with the 28S gene primer (5’-GCCAGATTAGAGTCAAGCTC-3’), which is complementary to the coding strand sequence 80 bp downstream of the R1 insertion site. After amplification, the PCR products were directly cloned into a modified mp18 vector digested with XcmI to generate 3' T overhangs (Burke, Muller, and Eickbush 1995). Individual plaques from each transformation were grown and subjected to the complementarity tests (C-tests) typically used to screen clones for m13 cloning/dideoxy sequencing (Bethesda Research Laboratories). Complementary pairs of R1 clones were identified for each species, and at least two clones in each orientation were completely sequenced using the Universal Sequencing primer (U.S. Biochemicals), as well as additional oligonucleotide primers for incremental sequencing. For those elements that were only obtained using the LTGHG or CLPVCRT primers located nearer the end of ORF2, the regions between these primers and the AYADD sequences were amplified using element-specific primers in combination.
with the degenerate AYADD primer. All sequences were analyzed using the MacVector package of programs (Oxford Molecular Group, Inc.). Additional screens for the presence of the R1B family in each Drosophila species were conducted using a series of R1B-specific PCR primers. The EHVLCEC primer (5'-GAACATGTTCCTGTGAGTG-3') was designed to be specific for the R1B family in the testacea and quinaria species groups. Two degenerate primers, VGYSLA (5'-GNTNGGNNGTAYTTRGY-3') and QVKYLG (5'-G/ CARGTNAARTAYCTNGG-3'), were designed to be specific for R1B elements from any Drosophila species. These R1B-specific primers were also used in combination with the downstream 28S gene primer.

**R1 sequences have been submitted to GenBank under the following accession numbers:** Drosophila testacea R1A, AF248077; R1B, AF248078; Drosophila neotestacea R1A, AF248067; R1B, AF248068; Drosophila orientacea R1A, AF248069; D. orientacea R1B, AF248070; Drosophila recens R1A, AF248075; D. recens R1B, AF248076; Drosophila quinaria R1A, AF248073; D. quinaria R1B, AF248074; Drosophila palustris R1A, AF248071; D. palustris R1B, AF248072; Drosophila falleni R1A, AF248066; Drosophila curvispina R1A, AF248065; Drosophila brachynephros R1A, AF248064; Drosophila ananassae R1B, AF248063.

**Sequence Divergence and Phylogenetic Analysis**

Comparisons of the sequenced R1 elements were carried out using the multiple-alignment feature of CLUSTALW (Thompson et al. 1997), followed by minor manual adjustments to gaps. The nucleotide sequence alignment in the protein-coding regions were forced to conform to the amino acid alignment. The R1 sequence from Sciara coprophila (Burke et al. 1993) was then aligned to the Drosophila R1 elements using the profile alignment option. Phylogenetic trees were generated by the neighbor-joining and maximum-parsimony methods as implemented in PAUP*, version 4.0d64 (Swofford 1998). Maximum-parsimony analysis with heuristic options was conducted with tree bisection-reconnection branch swapping with the maximum number of trees saved at each step limited to five. Nucleotide sequence divergence between R1 and R2 elements was determined by the Kimura (1980) two-parameter method, also using PAUP*.

**Genomic Blots**

To determine the fraction of the rDNA units inserted with the different R1 families, 2-μg aliquots of total adult DNA were digested with Clal and either KpnI, PstI, or SstI, separated on a 1% agarose gel, and blotted onto nitrocellulose paper. Hybridizations were conducted as previously described (Eickbush and Eickbush 1995) using a 716-bp 28S gene probe located downstream of the R1 insertion site. The probe was labeled with P-32-dCTP using an Amersham Rediprime II random primer labeling kit. Quantitation of the hybridizing bands was conducted with a PhosphorImager (Molecular Dynamics).

**Results**

Our previous surveys of R1 elements in 16 Drosophila species (Eickbush and Eickbush 1995; Lathe et al. 1995) revealed a uniform population of elements in the rDNA loci of 14 species (average within-species sequence divergence of only 0.48%). The phylogenetic relationship of these R1 elements, referred to as the R1A lineage, was generally congruent with the species phylogeny but had little resolution above the species group level (Lathe et al. 1995). The two remaining Drosophila species were each found to contain two distinct families of R1 elements. In D. neotestacea, one family corresponded to the R1A lineage, while the second was >50% divergent in nucleotide sequence from the A lineage and was referred to as the R1B lineage. In D. takahashii, the families differed in sequence by >30%, but both were more related to the R1A lineage in D. neotestacea than the R1B lineage. These two families were referred to as the R1A1 and R1A2 lineages.

**Distribution of the R1B Family in Drosophila**

To determine if species closely related to D. neotestacea also contained elements from the R1B lineage, R1 elements were cloned and sequenced from the three other characterized species of the testacea group, as well as six species from the related quinaria species group. PCR amplification of all potential R1 lineages was attempted as shown in figure 1A using one primer to downstream 28S gene sequences in combination with the degenerate primer to the AYADD region of the reverse transcriptase domain (see Materials and Methods for the sequences of the various PCR primers used in this study). All characterized R1 elements from arthropods encode the AYADD motif (Burke et al. 1998); thus, this primer should be sufficiently degenerate to enable the amplification of all R1 lineages within each species, not just those closely related to the original R1 elements characterized from Drosophila. To obtain R1 elements that might have substitutions within the AYADD region or give rise to fragments that are unstable in the cloning vector, we used a second degenerate primer, this time to a highly conserved LTGHG motif near the carboxyl terminal end of the R1 ORF. DNA fragments generated by this second primer pair were only half the length of those generated with the AYADD primer pair and thus should be more stable in the cloning vector.

For each species, an average of 15 R1 clones were randomly selected from these two PCR amplifications. Because the individual copies of R1 were cloned into the single-stranded mp13 vector in either orientation, complementarity tests (see Materials and Methods) were then conducted to arrange all R1 copies from each species into one or more groups. At least two clones from each orientation of each complementary group were completely sequenced. As summarized in table 1, R1 elements were identified in all nine additional species of
Table 1
Drosophila Species Surveyed

<table>
<thead>
<tr>
<th>SUBGENUS</th>
<th>SPECIES GROUP</th>
<th>SPECIES</th>
<th>R1 LINEAGES IDENTIFIED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drosophila</td>
<td>Testacea</td>
<td>D. neotestacea</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. testacea</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. orientacea</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. patrida</td>
<td>-</td>
</tr>
<tr>
<td>Quinaria</td>
<td></td>
<td>D. falleni</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. curvispina</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. brachynephos</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. recens</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. quinaria</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. palustris</td>
<td>+</td>
</tr>
<tr>
<td>Repleta</td>
<td></td>
<td>D. mercatorum</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. hydei</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. buzzatii</td>
<td>-</td>
</tr>
<tr>
<td>Sophophora</td>
<td>Ambigua</td>
<td>D. ambigu</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. pseudobscura</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. miranda</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. persimilis</td>
<td>-</td>
</tr>
<tr>
<td>Melanogaster</td>
<td>melanogaster</td>
<td>D. melanogaster</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>subgroup</td>
<td>D. simulans</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. sechellia</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. mauritiana</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. teissieri</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. yakaba</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. erecta</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. orena</td>
<td>+</td>
</tr>
<tr>
<td>ananassae</td>
<td>subgroup</td>
<td>D. ananassae</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. varians</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. bipectinata</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. parabipectinata</td>
<td>+</td>
</tr>
<tr>
<td>takahashi</td>
<td>subgroup</td>
<td>D. malerkolliana</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. takahashii</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. lutescens</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. pseudotakahashii</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. prostipennis</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D. paralatea</td>
<td>+</td>
</tr>
</tbody>
</table>

NOTE.—All species were initially surveyed by PCR using a series of degenerate primers designed to amplify all R1 elements. The PCR products were cloned in either orientation, and individual copies were arranged into complementary pairs. Identification of the family group of each complementary pair was by nucleotide sequence determination.

the testacea and quinaria groups. Each species contained elements that had high sequence identity to the R1A elements of *D. neotestacea*. R1 elements with higher sequence identity to the R1B elements of *D. neotestacea* were recovered from only five species: *D. testacea*, *D. orientacea*, *D. recens*, *D. quinaria*, and *D. palustris*. As we have found in our previous studies of multiple R1 or R2 families within a species (Burke et al. 1993, 1998, 1999; Lathe et al. 1995), the different copies of R1 from either the A or the B lineage exhibited <1% intraspecies nucleotide divergence (data not shown).

Our inability to detect the R1B lineage in four species from the testacea and quinaria groups could have been a result of very low R1B copy numbers in these species. To confirm the distribution of R1B elements in the testacea and quinaria species groups, the R1B sequences from the six species were compared, and a third primer, EHVLCEC (fig. 1A), was designed that would specifically anneal to elements of the R1B lineage. Use of this R1B-specific primer in combination with the downstream 28S gene primer resulted in abundant PCR products in the six species identified above as containing R1B, but failed to reveal R1B elements in the other four species.

In a previous characterization of R1 elements in the repleta species group, no evidence for the R1B lineage could be found in *D. mercatorum*, *D. hydei*, and *D. buzzatii* (Malik and Eickbush 1999). In a similar manner, previously published and unpublished surveys of R1 elements from the eight members of the melanogaster species group and four species of the obscura group gave no indication of the R1B lineage (Eickbush and Eickbush 1995; Eickbush et al. 1995; Lathe 1998). In addition to extensive sequence analysis of PCR products, these studies of repleta, melanogaster, and obscura species included comprehensive genomic blot studies to account for all insertions in the 28S genes. Only in the case of the genomic blots with *D. ananassae* (ananassae species subgroup of the melanogaster group) were rDNA bands detected that could not be accounted for by the single R1 and R2 lineages previously described in this species (Lathe et al. 1995; Lathe and Eickbush...
Additional screening of PCR products from \textit{D. ananassae} obtained with the AYADD primer eventually revealed the presence of a second R1 lineage, which corresponded to the R1B lineage. Based on sequence conservation among the R1B elements from \textit{D. ananassae} and the testacea and quinaria species, two additional degenerate PCR primers (VGVSLA and QVKYLG) were designed that should specifically amplify the R1B lineage from any species throughout the Drosophila or the Sophophora subgenus. All species listed in table 1 were surveyed using these new R1B primers. Four additional species from the ananassae species subgroup, \textit{D. varians}, \textit{D. bipectinata}, \textit{D. parabipectinata}, and \textit{D. malerkotliana}, were found to contain the R1B lineage, indicating that the R1B family is widely distributed in the ananassae species subgroup. No evidence was obtained for R1B elements in any species previously stored as negative.

In summary, degenerate primers designed to amplify any arthropod R1 element, as well as primers specific to the R1B lineage, revealed the presence of the R1B lineage in three of five species groups from Drosophila. In each case, R1B elements were found in only a subset of the species in that group. This finding is in contrast to that for the R1A lineage, which was readily amplified from all tested species of Drosophila.

Distribution of the R1A2 Family in Drosophila

The AYADD and LTGHG primers, used to amplify all R1 elements in the previous section, revealed many R1A elements but did not identify additional elements from the R1A2 lineage originally found in \textit{D. takahashii} (phylogenetic analysis described in the next section). To determine if species closely related to \textit{D. takahashii} also contained the R1A2 lineage, four additional species within the takahashii species subgroup were screened. This subgroup is reported to be only 1–2.5 Myr old, with several species still capable of forming interspecific hybrids (Parkash, Outsa, and Vanda 1994; Inaba and Aotsuka 1995). A degenerate primer to the conserved sequence CLPVCRT (part of a cysteine-histidine motif of the R1A1 and A2 elements of \textit{D. takahashii}) was used in combination with the downstream 28S gene primer to amplify additional R1 elements from \textit{D. takahashii}, as well as from \textit{D. prostipennis}, \textit{D. lutescens}, \textit{D. para}lutea, and \textit{D. pseudotakahashii}. The 6–12 R1 clonal sequences from each species revealed that all species had both the A1 and the A2 lineages. The level of intraspecific nucleotide variation in these species averaged 0.45%, while interspecies sequence divergence within the A1 or A2 lineages ranged from 2% to 5%. It should be noted that in our original study (Lathe et al. 1995), both the A1 and the A2 lineages of \textit{D. takahashii} were further subdivided into lineages that differed in nucleotide sequence by 3%–5%. The strain of \textit{D. takahashii} used in this current study did not contain this variation. Because the level of nucleotide variation observed earlier was similar to the level we found between species of this subgroup, it is likely that our original stock of \textit{D. takahashii} contained a species contaminant.

We conclude that both the A1 and the A2 lineages are widely distributed in the takahashii species subgroup. However, because of the recent origin of this group, its poorly resolved phylogeny, and the possible introgression of elements due to interbreeding, we have not characterized the sequence relationship of these elements further.

The A1 and A2 lineages of \textit{D. takahashii}, however, did provide the only example we have found to date of a recombination between two R1 lineages. Throughout the protein-encoding regions, the A1 and A2 elements of \textit{D. takahashii} contain 42% nucleotide divergence. In contrast, the usually more variable 3’ untranslated regions (UTRs) of these two R1 lineages differed in nucleotide sequence by less than 4%. As shown in figure 1B, the change from low to high sequence identity occurred at a precise site a short distance downstream of the ORF2 termination codon. Sequence analysis of the A1 and A2 families from the four other species in this group revealed this same recombination, suggesting that it occurred in the common ancestor of these species. It is likely that the 3’ UTR of the R1A1 lineage replaced the A2 lineage in this recombination, because the sequences of this 3’ UTR show substantial similarity to the 3’ UTRs of R1 elements from other species of the melanogaster group (data not shown).

Phylogeny of the Drosophila R1 Elements

The phylogenetic relationships of the 35 Drosophila R1 elements we have sequenced in this and previous studies are shown in figure 2. The analysis is based on the nucleotide sequence of the 1,125-bp region encoding the last 375 amino acids of the R1 ORF2. The 3’ UTR of R1 could not be used in the analysis because of its variable length in different species (500–1,000 bp) and because it contains only short regions of nucleotide sequence common to all Drosophila R1 elements. The phylogeny presented in figure 2 is a distance neighboring (NJ) tree and has an identical topology to a tree rooted using as the outgroup the R1 element of a distant dipteran, \textit{Sciara coprophila} (fungus gnat) (Burke et al. 1993).

The phylogenetic analysis presented in figure 2 clearly suggests that R1B elements form a distinct lineage that diverged from the R1A lineage not long after their common ancestor separated from that of the fungus gnat. The phylogeny of the R1B elements is congruent with the phylogeny of the R1A elements found in the same species, as well as the species phylogeny itself based on both mitochondrial DNA and 28S rRNA gene sequences (G. Spicer, personal communication). Thus, the current distribution of R1B elements can be explained by vertical inheritance.

We have previously suggested that R1 (and R2) elements are undergoing nucleotide sequence evolution at relatively uniform rates that are only slightly faster.
Fig. 2.—Phylogeny of the sequenced Drosophila R1 elements obtained using the neighbor-joining methods. The analysis is based on the nucleotide sequence of the 1,125-bp region encoding the last 375 amino acids of ORF2. Numbers above branch nodes are the bootstrap values (shown as percentages) based on 1,000 bootstrap replications using the neighbor-joining methods. Numbers below branch nodes are the bootstrap values using maximum-parsimony (heuristic search) methods. In three places, the branching structure is left unresolved because neither distance nor parsimony approaches gave 75% bootstrap support. R1 sequences were either newly obtained for this report or taken from previous studies (Eickbush and Eickbush 1995; Lathe et al. 1995; Malik and Eickbush 1999). The Drosophila R1 phylogeny was rooted using the R1 element from the dipteran Sciara coprophila (Burke et al. 1993). Shown at right are the subgenus classifications of the species from which the elements were obtained. The four R1 lineages present in Drosophila have been named A1, A2, A3, and B.

than those of low-codon-bias genes within the Drosophila genome (Eickbush et al. 1995; Lathe et al. 1995). Therefore, if both the R1A and the R1B lineages are evolving by vertical descent, the level of nucleotide sequence divergence for R1A comparisons should equal the level of divergence for the R1B elements. Shown in figure 3 are the levels of nucleotide divergence (Kimura [1980] two-parameter method) for the R1A elements from any two species, as well as with R1 elements from the outgroup, Sciara coprophila, are plotted versus the divergence of the R1B elements from these same species. All R1A elements in this comparison are from the R1A1 lineage. The sequences compared correspond to the ~1,125-bp region encoding the segment from the YADD region of the reverse transcriptase to the end of ORF2 (see fig. 1A). The Drosophila species used were D. neotestacea, D. testacea, D. orientacea, D. recens, D. quinaria, D. palustris, and D. ananassae.

In the case of the R1A2 lineage found in the takahashii species subgroup, the phylogeny in figure 2 suggests that this lineage diverged from the A1 lineage prior to or early in the evolution of the Drosophila genus. Unfortunately, because no other species groups contain representatives of this lineage, our data provide no direct support for or against a model of vertical evolution.

One surprising finding was revealed by the better-resolved phylogeny of the R1A1 lineage made possible by the addition of new taxa (27 species total). While the R1A1 phylogeny was congruent with respect to relationships within each of the five species groups, the relationship of R1A1 elements from different species groups was not congruent with the generally accepted phylogeny of Drosophila (Lachaise et al. 1988; DeSalle and Grimaldi 1991; Pelandakis and Solignac 1993; Russo, Takezaki, and Nei 1995). In particular, R1 elements from members of the repleta group (D. buzzatii, D. hydei, and D. mercatorum) appear as a sister group to R1 elements from the other four species groups, when as members of the Drosophila subgenus, the repleta elements should be more closely related to elements from the quinaria and testacea species groups. In a second, less supported, violation of the phylogeny, R1A elements from the obscura species group (D. ambigua, D. miranda, D. pseudoobscura, and D. persimilis) are more closely related to elements from the quinaria and testacea species groups, when as members of the Sophophora subgenus, they should be more similar in sequence to elements from the melanogaster species group.

R1 Elements in the repleta Group Are a Separate Lineage

If the A and B lineages are undergoing vertical evolution at a uniform rate, then why is the phylogeny
Fig. 4.—Nucleotide substitution rate of the R1A lineage compared with that of the R2 lineage. The level of nucleotide divergence (Kimura [1980] two-parameter method) for the R1A elements from two species is plotted versus the divergence of the R2 elements from the same species. The R1 and R2 sequence comparisons correspond to the carboxyl terminal end of the open reading frame (approximately 1,125 bp in the case of the R1 elements and 525 bp in the case of the R2 elements). The R2 data were derived from Eickbush and Eickbush (1995), Eickbush et al. (1995), and Lathe and Eickbush (1997). The open data points correspond to comparisons of elements from the repleta group with elements from the other four species groups. The circled open points represent comparisons between the repleta group and the testacea and quinaria groups. The circled solid data points at the lower left reflect the increased rate of R1 element evolution in the simulans species complex (Eickbush et al. 1995). R1 and R2 element sequences were available from 16 species of Drosophila: D. melanogaster, D. simulans, D. mauritiana, D. sechellia, D. teissieri, D. yakuba, D. takahashii, D. ananassae, D. pseudoobscura, D. ambigua, D. persimilis, D. neotestacea, D. recens, D. fallens, D. mercatorum, and D. buzzatti. Comparisons involving the R1A2 element in D. takahashii are not shown.

of the R1A1 lineage as shown in figure 2 not congruent with the species phylogeny? Two possibilities exist: horizontal transfers have occurred but have not involved the species compared in figure 3, or the R1A1 lineage has undergone further subdivisions, and these various lineages have been differentially retained within Drosophila. One way to resolve these two possibilities is to compare the rate of nucleotide divergence of the R1A1 elements with that of typical nuclear genes. In previous reports, we have compared the R1 and R2 nucleotide substitution rates with that of the alcohol dehydrogenase (adh) gene (Eickbush et al. 1995; Lathe et al. 1995; Lathe and Eickbush 1997). These comparisons indicated that the rates of synonymous substitutions in R1 and R2 were 2–2.5 times those of adh, while the rates of non-synonymous substitutions were 5–7 times those of the adh gene. Unfortunately, common nuclear gene sequences are not available from D. takahashii, D. ananassae, or members of the quinaria and testacea groups.

In a previous study of R2 elements from Drosophila, it was shown that while the R2 phylogeny lacked resolution in some instances, it was congruent with the species phylogeny (Lathe and Eickbush 1997). We therefore compared the nucleotide sequence divergence of 16 R1A elements with the divergence of the R2 elements present in the same species. As shown in figure 4, most of the divergence comparisons between R1 and R2 suggest that the rate of nucleotide substitution in R1 elements is similar to the rate in R2 elements (slope of the line in fig. 4). In a number of the comparisons, however, the R1 elements were significantly more divergent than the R2 elements from the same species. This was particularly apparent in the comparison of R1 elements from the repleta group with elements from the testacea and quinaria groups (open circles enclosed by a circle). This was not due to a higher rate of R1 nucleotide substitution in the repleta group, because the repleta elements showed similar levels of sequence divergence when compared with elements from the melanogaster and obscura groups (remaining open circles). These data suggest that the R1A elements in the repleta group represent a different lineage of elements than the R1A lineage found in the other four species groups of Drosophila. We will refer to this repleta-specific lineage as the R1A3 lineage (fig. 2).

The second instance in which the phylogeny of the R1 elements appeared to violate the species phylogeny (fig. 2) was the closer-than-predicted relationship of the R1 elements in the obscura group with those from the quinaria and testacea groups. While this could represent yet another R1 lineage, the data favoring this model were not very convincing. First, the bootstrap values supporting the clustering of obscura R1A elements with those from testacea and quinaria was minimally significant in the maximum-parsimony analysis (55% bootstrap support). Second, pairwise comparison of the R1A elements with the R2 elements (fig. 4) revealed no consistent pattern of the sequence divergences being too high or too low. It is thus possible that R1 elements are undergoing somewhat different rates of nucleotide substitutions in the different species groups that have affected their apparent phylogenetic relationships. For example, we have previously documented an increased rate in the R1 elements of the simulans species complex (Eickbush et al. 1995) (circled points at lower left in fig. 4). Only the analysis of larger numbers of Drosophila species will resolve whether the R1A1 lineage has undergone further subdivisions.

Abundance of R1A and R1B Elements

Why is the R1A lineage present in all lineages of Drosophila, while the R1B lineage is present in only a limited number of species? One possibility is that R1B elements are present at lower levels within the rDNA locus and are occasionally lost through competition with the more abundant R1A elements. The fraction of rDNA units inserted with R1 elements in any species can be readily estimated by means of genomic blots probed with short 28S gene sequences located immediately downstream of the R1 insertion site (Jakubczak, Burke, and Eickbush 1991; Jakubczak et al. 1992). Based on the sequences of the R1 elements, restriction enzymes were selected which would generate different-sized bands for the R1A and R1B elements, as well as the uninserted rDNA units for each species. A fourth hybridizing band was also observed in each species. These bands corresponded to the R2 elements in D. ananassae,
D. recens, and D. neotestacea and were assumed to be R2 in the other four species.

Table 2 summarizes the abundance of R1A and R1B elements in the seven Drosophila species known to contain both lineages. The fraction of the rDNA units inserted in these species ranged from 0.04 to 0.31 (mean 0.20) for the R1A lineage and from 0.05 to 0.27 (mean 0.15) for the R1B lineage. In three species, the levels of the R1A elements are higher than those of the R1B elements; in three species, they are at similar levels; and in one species, the level of R1B elements is higher than that of the R1A elements. Thus, there is no evidence to suggest that the R1A elements are generally maintained at higher levels and are thus outcompeting the R1B elements. Of course, the data in table 2 are only a “snapshot” of the dynamic increases and decreases of element number in the rDNA loci of each species. The level of R1 elements varies considerably between and within populations of a species (Lyckegaard and Clark 1991; Jakubczak et al. 1992; C. Perez-Gonzalez, unpublished data). To directly demonstrate whether these two lineages of elements are in competition will require the analysis of multiple individuals from a species with both lineages.

Discussion

The most rigorous means of detecting the horizontal transmission of a mobile element is to identify elements in two species that have significantly higher levels of sequence identity than are indicated by a comparison of typical nuclear genes from the same species (Cummings 1994). Our various surveys of R1 and R2 elements in the rDNA loci of arthropods yielded no comparisons between elements that were too similar in nucleotide sequence to be explained by vertical descent of the host species. A second, less rigorous, criterion for horizontal transmission of mobile elements is the presence of an element lineage in one species that is absent in closely related species. To explain such a distribution without horizontal transfer requires that the element be maintained in one species lineage while it is lost from all neighboring lineages (i.e., differential stability). The more divergent the elements and the more restricted their distribution in a group of organisms, the more unlikely it becomes that differential stability is the explanation.

FIG. 5.—Summary of the evolution of the four R1 element lineages in five species groups of Drosophila. For simplicity D. curvispina, D. falleni, and D. brachynephros are referred to as the falleni complex; D. palustris, D. quinaria, and D. recens are referred to as the quinaria complex; and D. neotestacea, D. testacea, and D. orientacea are referred to as the testacea complex (see fig. 2). A, R1B lineage. B, R1A1 lineage. C, R1A2 lineage. D, R1A3 lineage. Each panel traces the preservation (+) or loss (−) of a specific R1 lineage in the generally accepted phylogeny of the species used in this study. The data in panels A and B are well supported based on the vertical inheritance patterns of the R1B and R1A1 lineages. The distribution of R1A2 and R1A3 is drawn reflecting a vertical evolution model; however, current data cannot rule out the horizontal transfer of a R1 element from outside the genus (arrow).

In this report, we extended our analysis of two R1 lineages with restricted distributions: the R1B elements of D. neotestacea and the R1A2 elements of D. takahashii (Lathe et al. 1995). An extensive sampling of elements in 35 species confirmed that the R1B lineage is relatively rare in Drosophila. R1B elements were detected in only a limited subset of species from three of the five species groups examined. Phylogenetic analysis of these R1B elements revealed that their current distribution was consistent with vertical descent. Our ability to conclude vertical descent was greatly strengthened by the uniform rate at which these elements accumulated nucleotide substitutions compared with the R1A1 lineage (fig. 3). Based on these rates, the R1A and R1B lineages diverged approximately 100 MYA, well before the origin of the Drosophila genus. Assuming strict vertical descent, the distribution of R1B elements suggests that there has been a minimum of five independent losses of the B lineage in Drosophila. As shown in figure 5A, these losses would have occurred in the progenitor of the repleta and obscura species groups and within various lineages of the testacea, quinaria, and melanogaster species groups.

Phylogenetic analysis of the R1A lineage was more complex, because it is composed of multiple sublineages, R1A1, R1A2, and R1A3. What is the origin of these multiple lineages? Based on our estimates of the rate at which R1 elements accumulate nucleotide sub-
stitions, these three lineages also arose prior to the separation of the drosophila and sophophora lineages of Drosophila. While the R1A1 lineage suggests vertical descent since the origin of Drosophila, the restricted distribution of the R1A2 and R1A3 lineages provides no evidence for or against their vertical descent or any indication whether they arose as horizontal events from outside the Drosophila genus. However, based on our accumulating data on the evolution of R1 and R2 elements, we suggest that the more likely model is that these lineages arose by divergence (separation) of the A lineage into multiple sublineages that have been differentially retained in the various species groups of Drosophila (fig. 5C and D). The only means to directly demonstrate this model would be to characterize additional members of the A2 and A3 lineages in other Drosophilidae groups and demonstrate, as we have done for the A1 and B lineages, that they have evolved by vertical descent. The 35 Drosophila species we have sampled to date represent only a small fraction of the diversity of the Drosophila genus. Thus, it would not be surprising if the survey of additional taxa revealed other species with these lineages, as well as additional R1 lineages.

To preserve a model of strict vertical descent in light of the multiple R1 or R2 lineages within a species requires that these elements have the ability to diverge into independent lineages within a species. The formation of multiple-transposable-element lineages within a group of organisms is not unusual: numerous examples have been documented in the study of retrotransposon elements from different groups of organisms (see, e.g., Hutchison et al. 1989; Marin et al. 1998; Wright and Voytas 1998; Gonzalez and Lessios 1999). However, the formation of such multiple lineages by site-specific elements like R1 and R2 might be more difficult. Concerted evolution of the rDNA units within each species maintains near perfect identity of the rRNA genes themselves (Schlotterer and Tautz 1994; Polanco et al. 1998). Presumably, such concerted evolution would prevent the accumulation of nucleotide substitutions that would give rise to distinct lineages of R1 and R2. In apparent agreement with this model, we found that the level of nucleotide divergence between different R1 copies from the same lineage was typically <1%.

How can divergent R1 lineages arise? One possibility is that active R1 elements can exist outside the rDNA locus. Non-rDNA copies would not be subject to concerted evolution and thus could diverge in sequences and repropagate the rDNA locus with R1 copies too divergent in sequence to recombine with the original lineage. Copies of R1 elements have been detected outside the rDNA locus in many organisms. In the two species for which they have been studied, D. melanogaster and Bombyx mori, all non-rDNA copies were found to have diverged in sequence from those within the rDNA locus, as well as to have acquired mutations that would indicate that they are inactive (Roiha et al. 1981; Browne et al. 1984; Xiong et al. 1988; Jakubczak et al. 1992; Eickbush and Eickbush 1995; Adams et al. 2000). Thus, the non-rDNA copies in these species appear to represent evolutionary dead ends, rather than potential founders that could repropagate the rDNA locus. Of course, these studies of non-rDNA copies were not conducted for the species with multiple R1 lineages, and even if they had been, such studies could never exclude the possibility that “active” non-rDNA copies formerly existed in these species or are still present but simply not cloned. It is extremely difficult to clone all copies of a mobile element. For example, even the completed genome project for D. melanogaster does not include the non-rDNA copies of R1 elements, as well as many other mobile elements, known to be present within the highly repetitive DNA at the centromere of each chromosome (Roiha et al. 1981; Jakubczak et al. 1992; Adams et al. 2000).

While formally difficult to disprove, we find the non-rDNA “master copy” model unlikely. First, there are now multiple examples of non-LTR retrotransposons in different taxa that have evolved sequence specificity for tandemly repeated genes (see Malik and Eickbush 2000). The only means to preserve through evolution a master element that generates copies for insertion elsewhere in the genome would be positive selection for these insertions. Second, in the case of the only element whose replication has been studied in detail, R2, all findings suggest that its retrotransposition mechanism is highly adapted to insert into and be expressed from the rDNA units (Luan et al. 1993; Luan and Eickbush 1996; Eickbush, Luan, and Eickbush 2000). Thus, it seems more likely that these many site-specific elements have become specialized for a limited but reliable niche in the genome, where they can maintain a set of active elements.

Finally, an explanation for how multiple lineages of R1 elements can arise within the same rDNA locus can be suggested by our more recent findings (unpublished data). Analysis of individuals within a single population of D. simulans revealed that rDNA recombination predominantly results in the elimination of individual R1 and R2 insertions, while expansions occur by retrotransposition. This recombinational preference to eliminate insertions suggests that the high levels of sequence identities we have found for the R1 and R2 lineages in a species is due to their short lifetime within the rDNA locus of a species (rapid turnover), rather than the concerted evolution of the rDNA repeats. As a result, because the progeny of each element are independent of the other elements, the stochastic process of elimination can occasionally lead to the origin of a new lineage within a species.

Once generated, the coexistence of more than one lineage of R1 elements in the same rDNA locus would be dependent on the size and structure of the species populations, as well as the size, structure, and recombination frequencies of the rDNA loci. The rDNA units of most species provide a few hundred potential insertion sites. All of these sites cannot be occupied, as each insertion inactivates the rDNA unit (Long, Rebbert, and Dawid 1979; Kidd and Glover 1981; Jamrich and Miller 1984). If each R1 lineage occupied but a small fraction of these rDNA units, then competition would be limited...
and multiple lineages could survive. However, in some species of Drosophila, the level of R1 insertions exceeds 50%, with some individuals in these species containing levels of unoccupied rDNA units that are near the minimum estimates required for full viability (Franz and Kunz 1981; Templeton et al. 1989; Jakubczak et al. 1992; Malik and Eickbush 1999). None of these species contain multiple R1 lineages. Surprisingly, three of the species we have shown to contain both R1A and R1B lineages in this report appear to have high levels of R1 insertion (table 2). One would presume that in these species there is competition between the R1 lineages for a limited number of target sites. Direct evidence for such competition will require a population-level analysis of the abundance of the two R1 lineages within one or more species.

Acknowledgments

This work was supported by a National Science Foundation grant (MCB-9601198) to T.H.E. We thank John Jaenike for various species stocks; Warren Lathe, Cesar Perez-Gonzalez, and Greg Spicer for their unpublished data; and Danna Eickbush and Cesar Perez-Gonzalez for their comments on the manuscript.

LITERATURE CITED


Diethard Tautz, reviewing editor

Accepted October 17, 2000