Evolution of the Transposable Element Pokey in the Ribosomal DNA of Species in the Subgenus Daphnia (Crustacea: Cladocera)

Erin H. Penton and Teresa J. Crease

Department of Zoology, University of Guelph, Guelph, Ontario, Canada

Pokey is a member of the piggyBac (previously called the TTAA-specific) family of transposons and inserts into a conserved region of the large subunit ribosomal RNA gene. This location is a “hot spot” for insertional activity, as it is known to contain other arthropod transposable elements. However, Pokey is unique in that it is the first DNA transposon yet known to insert into this region. All other insertions are class I non-LTR retrotransposons. This study surveyed variation in Pokey elements through phylogenetic analysis of the 3′ ends of Pokey elements from ribosomal DNA (rDNA) in species from the nominate subgenus of the genus Daphnia (Crustacea: Cladocera). The results suggest that Pokey has been stably, vertically inherited within rDNA over long periods of evolutionary time. No evidence was found to support horizontal transfer, which commonly occurs in other DNA transposons, such as P and mariner. Furthermore, Pokey has diverged into sublineages that have persisted across speciation events in some groups. In addition, a new highly divergent paralogous Pokey element was discovered in the rDNA of one species.

Introduction

Transposable element evolution within host genomes is a dynamic process that involves both vertical transmission within a species and horizontal transfer between species. A typical transposon life cycle is thought to consist of vertical transmission through progeny, the eventual silencing of active elements by the host, and stochastic loss of inactive elements. The latter two stages are both balanced by the reintroduction of elements horizontally through transfer (Lohe et al. 1995; Hartl, Lohe, and Lozovskaya 1997). Life cycles vary both within and between the different classes of transposons. Class I retrotransposons tend to exhibit long periods of vertical transmission (Malik, Burke, and Eickbush 1999; Stuart-Rogers and Flavell 2001), with rare interspecific transfer events (Gonzalez and Lessios 1999; Zupunski, Gubensek, and Kordis 2001). In contrast, class II DNA transposons, such as the P and mariner families, display more frequent cycles of loss and horizontal transfer, which, at times, leads to patchy distributions across species (Clark and Kidwell 1997; Hartl, Lohe, and Lozovskaya 1997).

Some transposons undergo stable vertical transmission, and, as a result, they are maintained within a species over long evolutionary periods (Eickbush and Eickbush 1995). Such cases illustrate not only that transposons can be a stable component of a host’s genome but also that horizontal transfer is not always required to evade extinction. Often, cases of stable vertical transmission involve retrotansposable elements that have high copy numbers per genome, most of which are defective. R1 and R2, which insert into a conserved region of the arthropod large subunit ribosomal RNA (LSU rRNA) gene, are two such examples. They are both non-long terminal repeat (non-LTR) retrotransposons that have been found in a multitude of divergent arthropods (Jakubczak, Burke, and Eickbush 1991; Burke et al. 1993, 1998). Moreover, their phylogenetic relationships are consistent with those of their host species (Eickbush and Eickbush 1995), which suggests that they have been present in arthropods from the origin of the phylum, some 500 MYA (Burke et al. 1999) and are stable components of their genomes. These two elements insert in a sequence-specific manner within the LSU rRNA gene, at sites only 74 bp apart (Eickbush 2002).

Several other highly divergent families of transposons have been documented in arthropods within the conserved region of the LSU rRNA gene occupied by R1 and R2 (Burke, Muller, and Eickbush 1995; Sullender 1993). The high frequency of elements specific to this gene could be explained if rRNA genes are a suitable “habitat” for the persistence of insertion sequences. Indeed, Kidwell and Lisch (2001) suggested that genomes are composed of various “ecological niches” that are exploited by different types of transposons. The LSU rRNA gene occurs in tandemly repeated ribosomal DNA (rDNA) units, which form a multigene family that undergoes concerted evolution through gene conversion and unequal crossing over (Coen, Thoday, and Dover 1982; Arnheim 1983; Dvorak, Jue, and Lassner 1987). As a result, the transposons that occupy this niche are subjected to the same homogenization forces as are the genes themselves (Eickbush and Eickbush 1995). In fact, Eickbush and Eickbush (1995) suggested that it is these characteristics that make the rDNA “niche” ideal for the propagation and long-term persistence of transposable elements.

The presence of another site-specific transposon, Pokey, has been detected in the “hot spot” region of the LSU rRNA gene (Sullender 1993). In fact, Pokey inserts only base pairs away from R1 and R2 (fig. 1). Pokey is a class II DNA transposon that possesses terminal inverted repeats of 16 bp and a single 1.5-kb open reading frame (ORF) that codes for a putative transposase (Penton, Sullender, and Crease 2002). It creates a 4-bp target-site duplication (TTAA) on insertion, which makes it a member of the piggyBac family of transposons (Cary et al. 1989; Wang, Fraser, and Cary 1989; Beames and Summers 1990; Penton, Sullender, and Crease 2002; Sarkar et al. 2003). Pokey was originally found within a specific TTAA site in approximately 10% of the LSU rRNA genes of the...
cladoceran crustacean *Daphnia pulex*, although *Pokey* also inserts at many other genomic locations (Sullender 1993; Sullender and Crease 2001).

Here, we report the results of a phylogenetic analysis of the 3' end of *Pokey* elements that were amplified from rDNA in species from the nominate subgenus of the genus *Daphnia* (Crustacea: Cladocera). In addition, we used sequences from two genes, the mitochondrial small subunit (mtSSU) rRNA gene and the nuclear LSU rRNA gene, to construct phylogenies of the “host” *Daphnia* species from which the elements are derived. We show that *Pokey* is widely distributed in this subgenus and that the phylogeny of the elements is consistent with that of the host species, which suggests that *Pokey* has been stably inherited within rDNA over long periods of evolutionary time. Furthermore, we find that *Pokey* has diversified into multiple sublineages that have persisted across speciation events.

**Methods**

**Daphnia Samples**

We analyzed 32 isolates representing 14 species of the subgenus *Daphnia* (table 1). Total genomic DNA was extracted by use of the Isoquick kit (Orca Research) from single animals that were flash frozen in liquid nitrogen in the field or from multiple animals that were propagated parthenogenetically from a single female in the laboratory by the standard culture technique of Hebert and Crease (1980).

**DNA Amplification and Sequencing**

We used the polymerase chain reaction (PCR) to amplify an approximately 1,820-bp fragment of the 3' end of *Pokey* elements located in rDNA by use of an internal *Pokey* primer, Pok5026F (5'-TCGAACCTGCAAGCGG-GAGGAATTTTCGACG-3'), and a primer located in the LSU rRNA gene about 200 bp downstream of the element insertion site, 28SR (5'-CGTCTCTTTCACTTGACATCCACCTC-3'). The *Pokey* primer is located in the ORF, which may code for a transposase (Penton, Sullender, and Crease 2002). A few samples that did not amplify well were amplified using an alternate reverse primer 28SR (5'-TCCATTTCCTGGCGCCGTCACTAATTAGTAC-3'), which is located only 46 bp downstream of the TTAA target site. All PCR reactions were of 50 μl total volume and contained 1.5 mM MgCl₂, 5 pmol of each primer, 40 μM dNTPs, 10 mM Tris-Cl (pH 8.3), 50 mM KCl, 10 to 50 ng of genomic DNA, and 1 unit of Taq DNA polymerase (Roche). The amplification reactions were performed in an MJ PTC-100 thermal cycler (MJ Research Inc.). The thermocycling profile consisted of 1 cycle of 1 min at 94°C, 35 cycles of 30 sec at 94°C, 30 sec at 55°C and 2 min at 72°C, with a final incubation of 5 min at 72°C.

To verify the taxon identification of each *Daphnia* DNA sample, we amplified a 596-bp region of the mtSSU rRNA gene by use of primers 12SA (5'-CCGATATCAGT- TACCTTGGTACGGCG-3') and 12SB (5'-AATCTGCC- CAGCCGTCGCGG-3') (Colbourne and Hebert 1996) and

---

**Table 1: Daphnia Samples Used in This Study**

<table>
<thead>
<tr>
<th>Species Label*</th>
<th>Collection Site</th>
<th>State or Province</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. ambiguа – FL</td>
<td>Lake Pleasant</td>
<td>Florida</td>
</tr>
<tr>
<td>D. arenata – OR</td>
<td>Florence</td>
<td>Oregon</td>
</tr>
<tr>
<td>D. catawba – ON</td>
<td>Echo Lake</td>
<td>Ontario</td>
</tr>
<tr>
<td>D. cheraphila – SD</td>
<td>Buffalo</td>
<td>South Dakota</td>
</tr>
<tr>
<td>D. minutula – MI</td>
<td>Watersmeet</td>
<td>Michigan</td>
</tr>
<tr>
<td>D. minutula – ON</td>
<td>Sault St. Marie</td>
<td>Ontario</td>
</tr>
<tr>
<td>D. obtusa NA1 – AR*</td>
<td>Princeton</td>
<td>Arkansas</td>
</tr>
<tr>
<td>D. obtusa NA1 – AZ*</td>
<td>Coconino</td>
<td>Arizona</td>
</tr>
<tr>
<td>D. obtusa NA1 – IL1*</td>
<td>Hey-a-pond, St. Joseph</td>
<td>Illinois</td>
</tr>
<tr>
<td>D. obtusa NA1 – IL2*</td>
<td>Nobody’s Pond, Morton</td>
<td>Illinois</td>
</tr>
<tr>
<td>D. obtusa NA1 – LA*</td>
<td>Bellwood</td>
<td>Louisiana</td>
</tr>
<tr>
<td>D. obtusa NA1 – MO*</td>
<td>Springfield</td>
<td>Missouri</td>
</tr>
<tr>
<td>D. obtusa NA1 – NV</td>
<td>Red Rock</td>
<td>Nevada</td>
</tr>
<tr>
<td>D. obtusa NA1 – OK</td>
<td>Lake Texoma</td>
<td>Oklahoma</td>
</tr>
<tr>
<td>D. obtusa NA1 – SD*</td>
<td>Buffalou</td>
<td>South Dakota</td>
</tr>
<tr>
<td>D. obtusa NA1 – TN</td>
<td>Natchez</td>
<td>Tennessee</td>
</tr>
<tr>
<td>D. obtusa NA1 – TX1*</td>
<td>Big Wells</td>
<td>Texas</td>
</tr>
<tr>
<td>D. obtusa NA1 – TX2*</td>
<td>Cisco</td>
<td>Texas</td>
</tr>
<tr>
<td>D. obtusa NA2 – IL2</td>
<td>Nobody’s Pond, Morton</td>
<td>Illinois</td>
</tr>
<tr>
<td>D. obtusa NA2 – OH</td>
<td>Coal Tipple Pond, Portsmouth</td>
<td>Ohio</td>
</tr>
<tr>
<td>D. parvula – AR</td>
<td>Calion Lake</td>
<td>Arkansas</td>
</tr>
<tr>
<td>D. parvula – ON</td>
<td>Columbia Lake, Guelph</td>
<td>Ontario</td>
</tr>
<tr>
<td>D. pileata – OK</td>
<td>Tescumsh Rd Pond, Norman</td>
<td>Oklahoma</td>
</tr>
<tr>
<td>D. pulex – IL</td>
<td>Smith Farm, Vermillion County</td>
<td>Illinois</td>
</tr>
<tr>
<td>D. pulex – IN</td>
<td>Portland Arch, Warren County</td>
<td>Indiana</td>
</tr>
<tr>
<td>D. pulex – ON</td>
<td>Essex County, Ontario</td>
<td>Ontario</td>
</tr>
<tr>
<td>D. pulex – PQ</td>
<td>La Tuque</td>
<td>Quebec</td>
</tr>
<tr>
<td>D. pulicaria – SK</td>
<td>Humbolt Lake</td>
<td>Saskatchewan</td>
</tr>
<tr>
<td>D. retrocarva – IN</td>
<td>Waubee Lake</td>
<td>Indiana</td>
</tr>
<tr>
<td>European D. pulex – GR1</td>
<td>Grebin</td>
<td>Germany</td>
</tr>
<tr>
<td>European D. pulex – GR2</td>
<td>Malente</td>
<td>Germany</td>
</tr>
<tr>
<td>European D. pulicaria – GR</td>
<td>Givson Binnensee</td>
<td>Germany</td>
</tr>
</tbody>
</table>

*Letters following the dash specify the state or province of origin, whereas numbers distinguish multiple sites from the same state or province.
sequenced the PCR product directly. The sequences were compared with those obtained by Colbourne and Hebert (1996) in their phylogenetic analysis of the genus *Daphnia* in North America. In addition, we amplified two regions of the LSU rRNA gene. A 694-bp region that spans the insertion site of *Pokey* was amplified by use of primers 28SF (5'-CTGCCAGTCTGTCAGTCAATGGAAG-3') and 28SCR (5'-GATGGTCGGCCCCAGTCAAACCTC-3'). A 1134-bp region upstream of the *Pokey* insertion site was amplified by use of primers 28SS7TF (5'-AACCTCGCCGGTGTAGCCC-3') and 28S1211R (5'-TCCGACGATCGTATGC-3'). The PCR conditions were identical to those described above.

All PCR products were electrophoresed on 0.8% TAE agarose gels, stained with ethidium bromide, and visualized under UV light. The DNA fragments were excised from the gel and purified using either a QIAEX II Agarose Gel Extraction kit (Qiagen) or a freeze/thaw method. The freeze/thaw protocol was as follows: after excision, the agarose slice was frozen in the top of a filter-plugged pipette tip (50 μL), thawed, and then spun at maximum speed for 10 min in a 1.5 mL microfuge tube. The resulting eluant was precipitated in ethanol. All purified samples were sequenced by use of 20 to 50 ng of template with 5 pmol of the primers PokF, 28SBR, 28SR, 12SA, 28SF, 28SCR, 28SS7F, and 28S1211R as appropriate, using the ABI Prism TaqFS dye terminator kit (PerkinElmer). The sequences were resolved on an ABI 377 automated sequencer. Sequences that were well resolved on the electropherograms were only sequenced in one direction. However, sequences longer than 700 bp were sequenced from both ends so that overlapping data were available for the middle region of the fragments. Samples that provided low quality sequence data were cloned by use of the TOPO TA Cloning kit for Sequencing (Invitrogen). For this process we used high fidelity JumpStart™ Taq DNA polymerase (Sigma-Genosys) to generate the cloning template to avoid error associated with the misincorporation of nucleotides by the Taq DNA polymerase. When necessary, we sequenced these clones with an internal primer, Pok5338F (5'-TGGCTCTGTTGACAGGTCGATGGATATGC-3').

Sequences of *Pokey* from the rDNA of *Daphnia pulexica* (D. pulexica-SK-C1 and D. pulexica-SK-C2) were taken from GenBank (accession numbers AY115589 and AY115590). The *Pokey* sequence *D. pulex*-IL+IN-consensus is the consensus of 20 cloned sequences from two isolates of *D. pulex* from IL and IN. In addition, the *Pokey* sequences D. pulex-PQ-C1, D. obtusa NA1-TN-C11, and European D. pulex-GR2-C9 were taken from full-length rDNA *Pokey* clones generated for another study (Penton, unpublished data). Sequences of the LSU rRNA gene for *D. pulexica* (accession number AF346514) and *D. ambigua* (accession number AF346513) were taken from Omilian and Taylor (2001).

Alignment and Phylogenetic Analyses

*Pokey*, mtSSU rRNA gene, and LSU rRNA gene sequences were all aligned by the Align program (Person et al. 1997) and/or by eye with the aid of Bioedit (Hall 1999). Use of the complete gene sequences reported by Omilian and Taylor (2001) in their phylogenetic reconstruction of daphniids facilitated alignment of the LSU rRNA gene sequences. Highly variable regions of expansion helices were removed from this alignment before phylogenetic analysis.

We performed a Bayesian phylogenetic analysis for all three data sets by application of MrBayes (http://morphbank.ubc.ca/mrbayes) (Huelsenbeck 2000). We first chose the model of DNA substitution by characterizing the sequences with Modeltest version 3.0 (http://ibio. byu.edu/Faculty/kac/crandall_lab/modeltest.htm) (Posada and Crandall 1998). The results of these analyses were then incorporated into the Bayesian phylogenetic analyses. The number of generations run for all three data sets was 500,000. All trees constructed before confluence (10,000 generations) were discarded as “burn-in” (Huelsenbeck 2000).

We performed cladistic analyses by implementation of the maximum-parsimony criterion in PAUP version 4.0b10 (Swofford 2002) and use of the program’s heuristic search algorithm and tree-bisection and reconnection (TBR) feature. The results of the Modeltest analyses were also incorporated into these analyses. Sequences were added randomly in 50 replicate trials, with one tree held at each step. Bootstrap values for maximum-parsimony trees were based upon 1,000 pseudoreplicates.

We used the Kimura two-parameter model (Kimura 1980) to estimate sequence divergence for all three genes in MEGA version 2.1 (Kumar, Tamura, and Nei 1993). Phenetic analysis of the resulting distance matrices was performed by use of the neighbor-joining (NJ) method in MEGA (Saitou and Nei 1987), with pairwise deletion of missing sites. The bootstrap percentages from 1,000 pseudoreplicates were calculated in MEGA. We constructed NJ trees separately for the different domains (coding versus noncoding) of the *Pokey* transposon to determine the presence of any region-specific differences in the pattern of *Pokey* sequence divergence.

We rooted all phylogenetic trees through the sequence of *D. ambigua* based on Colbourne and Hebert’s (1996) phylogenetic analysis of the genus *Daphnia* from North America. We performed tests of neutrality on the ORF region of the *Pokey* sequences by application of the Z-test (Nei and Kumar 2000) as implemented in MEGA. The Nei-Gojobori p-distance model with pairwise deletion of missing data was used to estimate the number of the synonymous and nonsynonymous substitutions between pairs of sequences.

Statistical Analysis of Phylogenetic Congruency

We used three statistical approaches to test whether the topologies constructed from the *Pokey* sequences and the gene sequences from the *Daphnia* “host” species are congruent. The first approach, ParaFit (Legendre, Desdevois, and Bazin 2002), which is available on-line at http://www. fas.umontreal.ca/biol/casgrain/en/parafit.html, evaluates the correlation between sequence divergence in the “parasite” *Pokey* and its host species. Both the overall coevolutionary structure (ParaFitGlobal statistic) and the significance of individual host-parasite links (ParaFitLink1
Results
Phylogenetic Analyses

All sequences generated for this study are available on GenBank. Accession numbers for the mtSSU rRNA gene sequences are AY626352 to AY626366, accession numbers for the LSU rRNA gene sequences are AY630599 to AY630618, and accession numbers for the Pokey sequences are AY630579 to AY630598. All of the alignments are available at the Molecular Biology and Evolution Web site as Supplementary Material online.

Pairwise sequence divergence between mtSSU rRNA gene sequences ranges from 0.2% to 21.2%. Of the total 502 nucleotide positions, 177 are polymorphic, 119 of which are phylogenetically informative. The phylogeny (fig. 2) produced from these sequences is consistent with that of Colbourne and Hebert (1996). However, support is low for a few nodes that define the relationships between major groups. Nonetheless, the monophyly of these major groups is well supported.

Because of the slow evolution of reproductive isolation in the genus Daphnia, interspecific hybridization is not uncommon (Colbourne and Hebert 1996). As mitochondria exhibit uniparental transmission, cases of hybridization could go undetected. In addition, support is low for some nodes in the mitochondrial trees. Thus, we also constructed a phylogeny based on a nuclear gene. For the combined LSU rRNA gene fragments, 142 nucleotides of a total 1,481 are polymorphic, 55 of which are phylogenetically informative. Pairwise sequence divergence for these sequences ranges from 0.1% to 6.4%. Phylogenies produced from the LSU rRNA gene (fig. 3) are generally consistent with the trees constructed from the mtSSU rRNA gene. Although support is low for a few nodes that define relationships among major groups, the monophyly of those major groups is well supported. The only major difference between the two Daphnia phylogenies concerns the sister group relationship between D. parvula/D. retrocurva and the D. obtusa group in the LSU rRNA gene tree. In the mtSSU rRNA gene tree, D. parvula/D. retrocurva clusters with D. catawba/D. minnehaha (fig. 2).

In the Pokey sequence alignment, 824 nucleotides of 1,711 are polymorphic, 525 of which are phylogenetically informative. The pairwise divergence between unique sequences ranges from 0.1% to 44.0%. This sequence contains approximately 355 bp of the 3' end of the ORF, which may encode a transposase and approximately 1,350 bp of noncoding DNA. The pairwise sequence divergence for the ORF is much smaller than that of the noncoding sequence and ranges from 0% to 18% (average of 6.5%). That of the noncoding region ranges from 0.1% to 54.7% (average of 23.2%).

Most Pokey sequences share the same stop codon. However, D. obtusa (NA1 and NA2) and D. catawba possess different deletions that both lead to the same downstream stop codon, which adds 7 and 8 amino acids (aa) to the proteins, respectively. On the other hand, D. pileata has a 1-bp insertion that leads to a premature stop codon, which truncates the protein by 3 aa. Finally, D. ambiguus has a large deletion at the 3' end of the ORF, which leads to the absence of a nearby stop codon. Because of these differences, an ORF data set was constructed that contained the 315 nucleotide positions up to and including the codon directly upstream of the large deletion in the sequence of D. ambiguus. This regions contains 78 variable nucleotide sites leading to 24 aa changes. The majority of the nucleotide changes (66.67%) are at third codon positions (52 of 78), whereas 17 (21.79%) and 9 (11.54%) are at first and second codon positions, respectively. The high proportion of third position differences suggests that the ORF is conserved because of functional constraint. Indeed, Penton, Sullender, and Crease (2002) and Sarkar et al. (2003) have shown that the putative protein encoded by this ORF is similar to the functional transposase of the piggyBac element. Moreover, Sarkar et al. (2003) found a putative DDD amino acid motif located in the middle of a conserved core region (D268, D346, and D447) of the transposases from the piggyBac
The widely documented DDD motif is believed to be the functional catalytic domain of the transposase proteins from the \textit{Tc1} and \textit{mariner} groups (Doak et al. 1994; Robertson and Lampe 1995). As this study only examined the 3' end of the ORF, we could not detect the entire DDD motif. However, the third aspartic acid (D) residue is present in all \textit{Pokey} sequences examined here.

An overall Z-test of neutrality that included all \textit{Pokey} sequences indicated that purifying selection, rather than positive selection, is responsible for the pattern of nucleotide substitution in the ORF across species (Z = 5.445, P \ll <0.001). Because some species were represented by multiple sequences, we repeated the analysis but only included one sequence for each species (\textit{D. pulex}-ON for \textit{D. pulex}, \textit{D. pulex}-IL+IN-consensus, \textit{D. pulex} IL, \textit{D. pulex} GR2-C9, \textit{D. obtusa} NA1-OK-C1, \textit{D. obtusa} NA2-IL2-C4, and \textit{D. parvula}-ON). Again, the results strongly suggest that purifying selection is responsible for the pattern of nucleotide substitution observed among the species (Z = 6.53, P \ll <0.001).

Trees constructed from the \textit{Pokey} sequences by different phylogenetic methods are generally consistent with one another and the support is high for most nodes (fig. 2). Two additional \textit{Pokey} trees were constructed; one based entirely on the coding regions from the ORF (see below) and the other based on the noncoding DNA (data not shown). These two phylogenies are generally consistent with the one shown in figure 2, which indicates a lack of region-specific differences in the pattern of \textit{Pokey} sequence divergence.

### Statistical Analysis of Phylogenetic Congruency

The ParaFit global test of an association between \textit{Pokey} sequence divergence and that of its host's mtSSU rRNA gene is highly significant (i.e., coevolutionary) (ParaFitGlobal = 0.02092, P = 0.001). The analysis of individual host-parasite links (fig. 2) shows only the \textit{D. cheraphila} link to be nonsignificant (i.e., random) (ParaFitLink1 = 0.00065, P = 0.115), even though the topology of the two trees is congruent in this region.
Fourteen *Daphnia* sequences appear on the mtSSU rRNA gene tree (Euro *D. pulex* was only included once in the topology) and 17 sequences appear on the *Pokey* tree used in the tree-mapping analysis. The reconciled *Daphnia* tree requires the addition of 33 nodes and the loss of 11 nodes. The frequency distribution of nodes added in 1,000 random *Daphnia* trees ranges from 66 to 151, and the distribution of nodes lost ranges from 31 to 74. Both of these ranges are significantly higher than the number of gains and losses observed in the reconciled *Daphnia* tree.

Thus, the association between the host mtSSU rRNA gene tree and the *Pokey* tree is highly nonrandom. Twelve *Daphnia* sequences appear on the LSU rRNA gene tree (Euro *D. pulex* was only included once in the topology) and 17 sequences appear on the *Pokey* tree used in the tree-mapping analysis. The reconciled *Daphnia* tree requires the addition of 33 nodes and the loss of 11 nodes. The frequency distribution of nodes added in 1,000 random *Daphnia* trees ranges from 66 to 151, and the distribution of nodes lost ranges from 31 to 74. Both of these ranges are significantly higher than the number of gains and losses observed in the reconciled *Daphnia* tree. Thus, the association between the host mtSSU rRNA gene tree and the *Pokey* tree is highly nonrandom.

The only major difference between the *Pokey* tree and the *Daphnia* LSU rRNA gene tree is the position of *D. pileata*, but two major differences exist between the *Pokey* tree and the *Daphnia* mtSSU rRNA gene tree: (1) the relationships among *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group. A separate SH test was used to determine whether each of these three differences was significant. A single representative of *Pokey* from each species was used in this analysis (Euro *D. pulex*-ON for *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group). A separate SH test was used to determine whether each of these three differences was significant. A single representative of *Pokey* from each species was used in this analysis (Euro *D. pulex*-ON for *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group). A separate SH test was used to determine whether each of these three differences was significant. A single representative of *Pokey* from each species was used in this analysis (Euro *D. pulex*-ON for *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group). A separate SH test was used to determine whether each of these three differences was significant. A single representative of *Pokey* from each species was used in this analysis (Euro *D. pulex*-ON for *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group). A separate SH test was used to determine whether each of these three differences was significant. A single representative of *Pokey* from each species was used in this analysis (Euro *D. pulex*-ON for *D. pulex/D. plicaria/D. arenata*, Euro *D. plicaria*, and Euro *D. pulex* and (2) the position of the *D. parvula/D. retrocurva* group).

**Intraspecific Variation in Pokey**

To determine the level of *Pokey* sequence divergence within a geographically widespread species, we analyzed a smaller segment (763 bp) of the 3′ end of *Pokey* from

![Pokey: parasite](image-url)

![LSU rRNA gene: host](image-url)

**Fig. 3.—** Neighbor-joining phylogenies based on variation in *Pokey* and the LSU rRNA gene from species in the subgenus *Daphnia*. Dotted lines connect *Pokey* isolates to corresponding host genomes. These relationships were not tested statistically and are included for display only. Numbers beside major nodes represent bootstrap support (1,000 pseudoreplicates) from neighbor-joining and maximum-parsimony analyses and clade credibility values from Bayesian analysis (NJ/MP/BA). Asterisks indicate no support in the MP or BA analyses for the topology shown. The scale bar indicates sequence divergence. Significant rate variation occurs across sites in the LSU rRNA gene based on the TrN model of DNA substitution (number of substitution sites = 6) and the invariant-gamma model (*P* = 0.000002). The gamma distribution shape parameter is 0.7347, the proportion of invariable sites is 0.7437, and the substitution rate matrix is R[A-C] = 1.0000, R[A-G] = 3.9170, R[A-T] = 1.0000, R[C-G] = 1.0000, R[C-T] = 7.9080, R[G-T] = 1.0000. Details for the *Pokey* tree are given in figure 2.
additional isolates of *D. obtusa* from across its North American range (table 1 and fig. 4). Based on analysis of the mitochondrial cytochrome *c* oxidase subunit I gene, Penton, Hebert, and Crease (2004) showed that *D. obtusa* in the United States diverged into two morphologically cryptic species, denoted NA1 and NA2, at least 12 MYA. Furthermore, *D. obtusa* NA1 radiated into four lineages with largely allopatric distributions during the Pleistocene (< 1 MYA). Even so, the mean sequence divergence between *Pokey* sequences from all *D. obtusa* isolates, including the two species, is only 0.7%. However, cloned sequences from two isolates, NA1-TN and NA2-IL2, are substantially more divergent from the others with a mean sequence divergence of 1.2%. When these two isolates are removed from the analysis, the mean divergence between the remaining *Pokey* sequences decreases to 0.4%.

We cloned and sequenced the larger *Pokey* fragment (1,618 bp) from isolates NA1-NV, NA1-OK, NA1-TN, NA2-IL2, and NA2-OH. The mean divergence between these larger *Pokey* fragments is 3.9%. This increase reflects the fact that the additional sequence includes a hypervariable region. In this case, the NA1-TN and NA2-IL2 *Pokey* fragments show 6.6% sequence divergence from the other three sequences, whereas the mean divergence among the other sequences is only 0.9%. These results indicate the existence of at least two lineages of *Pokey* elements in *D. obtusa* and that both occur in each of the two species.

Inspection of the electropherograms generated by direct sequencing of PCR fragments from several of the other *D. obtusa* isolates shows that “double peaks” often occur at nucleotide positions where the two types of *Pokey* elements differ. In fact, the sequence in one highly variable region where most of the differences occur was impossible to read because of the presence of both of these lineages within these individuals. This situation is similar to the one seen in *D. pulicaria*, where two different lineages of *Pokey* elements, differing by 6% sequence divergence, were PCR amplified and cloned from the genome of a single individual (Penton, Sullender, and Crease 2002). These two sequences are included in the *Pokey* tree in figures 2 and 3 (*D. pulicaria*-SK-C1 and *D. pulicaria*-SK-C2). Note that the sequences from *D. pulex* are clearly more closely related to the SK-C1 sequence. To date, no indication has emerged that the second lineage (SK-C2) is present in the *D. pulex* isolates that have been sampled, but a conclusion that it has been lost from this species altogether is clearly premature.

Two isolates of *D. parvula* (table 1) were also included in the *Pokey* analysis, one from Ontario (ON) and a second from Arkansas (AR), and the divergence between them is 0.8%. In contrast, the sequence divergence between *Pokey* elements from different species is often substantial (table 3). For example, sequence divergence ranges from 39.1% to 42.6% (average of 41.0%) between *Pokey* from the most divergent taxon, *D. ambigua*, and from all of the other taxa. Overall, this finding suggests that each species contains one or more highly homogenous lineages of rDNA *Pokey* elements whose interspecific sequence divergence is highly correlated with that of their hosts.

### A New Paralogous *Pokey* Lineage

Two of the cloned *Pokey* elements from *D. obtusa* NA2-OH that we sequenced were so divergent from the others that we were not able to completely align their noncoding regions. An NJ phylogeny produced from the ORF region (fig. 5) shows that these elements form a sister group to the sequences obtained from all of the other species in the subgenus, and, thus, they represent a new paralogous *Pokey* family in the rDNA of the subgenus *Daphnia*. This new family is hereafter denoted as *Pokey*B, and the original family is denoted as *Pokey*A.

We incorporated the ORF sequence of *Pokey*B into the alignment of known *piggyBac* elements (which includes *Pokey*A) generated by Sarkar et al. (2003), calculated a pairwise matrix of the number of amino acid differences, and then used it to generate an NJ tree (data not shown). *Pokey*A and *Pokey*B cluster with one another with strong bootstrap support (999/1,000 replicates) on this tree, which indicates that they are both members of the same major group of transposons. The *D. obtusa* NA2-OH isolate from which *Pokey*B was isolated also contains *Pokey*A elements (fig. 5), so

---

**Table 2**

<table>
<thead>
<tr>
<th>Topology</th>
<th>−lnL</th>
<th>Delta</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pokey (11 species)</td>
<td>8828.6307</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>LSU rRNA – <em>D. pileata</em></td>
<td>8845.3589</td>
<td>0.141</td>
<td></td>
</tr>
<tr>
<td>Pokey (12 species)</td>
<td>9520.9561</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>mtSSU rRNA – <em>D. pulex</em> group</td>
<td>9563.3885</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>mtSSU rRNA – <em>D. parvula</em></td>
<td>9559.4349</td>
<td>0.034</td>
<td></td>
</tr>
</tbody>
</table>

* An NJ tree of the *Daphnia* species was generated from the *Pokey* data and then compared with constrained trees based on the LSU rRNA and mtSSU rRNA genes. Only one *Pokey* sequence was used to represent each *Daphnia* species. The difference between the *Pokey* tree and the tree based on the rRNA gene sequences is indicated.

* An LSU rRNA gene sequence was not obtained for *Daphnia cheraphila*.

* −lnL denotes the likelihood score of the tree given the *Pokey* sequence data. The likelihood settings were number of substitution types = 2 (HKY85 variant) and a transition/transversion ratio = 2 (kappa = 3.98).

* Delta is the difference in likelihood score between the *Pokey* tree and a constrained tree.

* P is the probability that the two trees are congruent.
clearly these two families can coexist in the same genome. Analysis of the complete nucleotide fragment for PokeyB (fig. 6) reveals that the 3' end of its ORF is intact, although the location of its stop codon would produce a protein that is 11 aa shorter than the one produced by the PokeyA elements from most species. Sequence divergence between the consensus ORF of PokeyA and PokeyB is only 16%, whereas that of the noncoding DNA is greater than 50%. This finding suggests that the ORF is conserved across Pokey families because of functional constraint, and that PokeyB is likely to be an autonomous transposon family.

Discussion

Colbourne and Hebert (1996) found weak support for the relationships among major groups in their phylogenetic reconstruction of North American Daphnia. Similarly, the relationships between major groups are not well supported in either the mtSSU rRNA gene or the LSU rRNA gene.
phylogenies constructed here (figs. 2 and 3). A comparison between these two host phylogenies reveals one major difference, the sister group relationship of *D. parvula* // *D. retrocurva* to the *D. obtusa* group in the LSU rRNA gene tree. Hybridization is unlikely to be the explanation for this occurrence, as these events have only been documented between species possessing mtSSU rRNA gene sequence divergence of less than 14% (Colbourne and Hebert 1996). Overall, no conclusive evidence of horizontal transfer was detected, and the results of the ParaFit analysis (Hebert 1996) are consistent with vertical transmission. For example, Gentile et al. (2001) detected two *mariner* elements in *Daphnia* species, even in the complete absence of horizontal transfer and indeed, Capy, Anxolabehere, and Langin (1994) argue that phylogenetic inconsistencies may have been accurately attributed to horizontal transfer when they could be explained by other factors that are consistent with vertical transmission. For example, comparison of paralogous elements (i.e., different lineages), varying rates of sequence evolution of elements between species, and retention of ancestral polymorphisms (Capy, Anxolabehere, and Langin 1994) can lead to a transposon tree that is incongruent with that of the host species, even in the complete absence of horizontal transfer. The few discrepancies between the *Pokey* tree and the *Daphnia* gene trees in the present study are likely the result of one or more of the above factors, especially given the occurrence of multiple lineages of rDNA *Pokey* elements in the genome of both *D. pulicaria* (Penton, Sullender, and Crease 2002) and *D. obtusa*.

The occurrence of multiple, long-lived lineages such as we observed in *Pokey* has also been observed in the arthropod rDNA-specific non-LTR retrotransposons *R1* and *R2*. For example, Gentile et al. (2001) detected two paralogous lineages of *R1*, A and B, in five species groups of *Drosophila*. All 35 species surveyed contained the A lineage, whereas 11 species, from three of the five species
groups, also contained the B lineage. In addition, the A lineage had diverged into two sublineages, A1 and A2, in the melanogaster species group. All the species in this group had the A1 lineage, whereas the five species in the takahashii subgroup (which is within the melanogaster species group) also contained the A2 lineage. Results obtained by Burke et al. (1993) in an analysis of elements from five divergent insect species—Bombbyx mori (Lepidoptera), D. melanogaster (Diptera), Sciara coprophila (Diptera), Popillia japonica (Coleoptera), and Nasonia vitripennis (Hymenoptera)—suggest that multiple paralogous lineages of R1 can persist for even longer periods of time. Four lineages of R1 were detected in N. vitripennis, but they were not monophyletic. One of them was most closely related to the element obtained from B. mori, whereas the other three formed a monophyletic cluster that grouped with the elements from D. melanogaster and S. coprophila (figure 6 in Burke et al. [1993]). Based on this limited sampling of taxa, at least two paralogous lineages of R1 appears to exist in insects.

As is the case for Pokey elements in Daphnia, intraspecific sequence divergence between copies of R1 and R2 elements from the same lineage in Drosophila is generally less than 1%, whereas sequence divergence between lineages is much higher (Gentile, Burke, and Eickbush 2001). Originally, such homogeneity of elements was thought to be maintained by concerted evolution within rDNA, but this hypothesis made an explanation of how new lineages could evolve difficult. However, Pérez-González and Eickbush (2001, 2002) showed that the sequence homogeneity within R1 and R2 lineages is most likely a function of rapid turnover of elements, which includes both elimination and insertion. Individual variants are rapidly eliminated by concerted evolution, which is primarily a consequence of intrachromosomal crossing over (Pérez-González, Burke, and Eickbush 2003), and then replaced by new copies so that sequence homogeneity is mainly a function of recent retrotransposition. Consequently, active elements give rise to progeny that diverge from one another independently, which provides opportunities for lineage sorting. The presence of a transposable element in an LSU rRNA gene results in a nonfunctional copy. Thus, the number of LSU rRNA genes that can be inactivated by R1 and R2 without affecting the fitness of the host creates a limiting resource, especially at particular times during development when a very high proportion of active genes are required to meet the requirements of rRNA transcription. The resulting competition results in the persistence of only a small number of lineages within any one species (Pérez-González and Eickbush 2001, 2002).

Gentile, Burke, and Eickbush (2001) extracted DNA from single stocks of each of the Drosophila species that they analyzed, so they were not able to obtain information on geographic variation among R1 elements. The present study shows that sequence divergence within a lineage of Pokey is extremely low even across very broad geographic areas. This condition is difficult to attribute to common ancestry from a small pool of active elements, given the substantial divergence in allozymes (Hebert and Finston 1996) and mtDNA (Penton, Hebert, and Crease 2004) that is known to have occurred among the D. obtusa populations that were surveyed. More likely, concerted evolution accounts for such homogeneity, as it does for the rDNA itself. If this conclusion is correct, how then can new lineages of Pokey evolve and persist for such long periods of time?

Nothing is known about rates of Pokey transposition or its precise mechanism. However, if Pokey transposes by the same “cut and paste” mechanism used by piggyBac (Elick, Bauser, Fraser 1996; Lobo, Li, and Fraser 1999) then transposition is not replicative, as is retrotransposition. Thus, increases in Pokey’s copy number in rDNA could be facilitated by concerted evolution or by transposition of copies from other genomic locations, as active copies of Pokey are known to occur outside of rDNA (Sullender and Crease 2001). Thus, the possibility that such exchange could occur would also help explain how different lineages of Pokey are able to evolve despite the homogenizing effect of concerted evolution. Genomic copies of Pokey could act as a reservoir of new variants that have evolved independently of lineages currently occupying rDNA. A survey of sequence variation among rDNA and non-rDNA copies of Pokey screened from a cosmid library produced from a single individual of D. pulicaria is currently underway to determine if transposition does occur between rDNA and other genomic locations.

rDNA has been suggested to possess characteristics that make it ideal for the evolution of insertion sequences. Although the presence of a transposon within the LSU rRNA gene results in a nonfunctional copy, rDNA is highly repetitive, so individuals have more units than they need for survival. Thus, elements can exist to a certain threshold level in this specific location within a genome without any noticeable phenotypic effects at the organismal level (Eickbush and Eickbush 1995; Malik and Eickbush 1999). High levels of insertion of R1 and R2 have been exhibited in Drosophila mercatorum and D. melanogaster/D. hydei, where they are associated with the abnormal abdomen and bobbed phenotypes, respectively (Franz and Kunz 1981; Templeton et al. 1989; Lathe et al. 1995; Malik and Eickbush 1999). Overall, however, targeting a site within a multigene family reduces the risk that the element will insert into single-copy genes and adversely affect the phenotype. In addition, the rDNA genes are actively transcribed, ensuring expression of the element’s transposase, and highly conserved, ensuring a population of uniform future target sites (Eickbush and Eickbush 1995). This feature is a definite advantage as transposons with nonspecific insertion sites are at risk of transposing into areas of the genome that are not transcriptionally active, such as heterochromatin.

Overall, the characteristics of rDNA combine to define a genomic location that is advantageous for the propagation of elements. Indeed, phylogenetic reconstruction that involves complete R2 elements from various arthropod species revealed that they have been present in these lineages from the origin of the phyllum, some 500 Myr (Burke et al. 1999). A subsequent analysis of the RT domain from non-LTR retrotransposons identified 11 distinct clades and dated the origin of this superfamiliy of elements to the Precambrian era (Malik, Burke, and
Eickbush 1999). No evidence for horizontal transfer was found either within or between clades. Malik, Burke, and Eickbush (1999) suggested that the lack of horizontal transfer and subsequent vertical nature of inheritance of non-LTR retrotransposons could be caused by their target-primed reverse transcription mechanism. However, Zupunski, Gubensek, and Kordis (2001) recognized cases of horizontal transfer in the RTE clade of non-LTR retrotransposons, which indicates that this superfamily of elements is indeed capable of horizontal transfer. Thus, the lack of horizontal transfer in R1 and R2 may be more a function of their occupation of rDNA than of their transposition mechanism. Indeed, Pokey is a very different type of element than R1 and R2, and yet its pattern of evolution in rDNA is very similar to theirs, which suggests that the rDNA location itself has a major impact on the evolution of the elements that insert within it.

The presence of multiple, divergent elements within a specific region of the rDNA unit suggests that it is a “hot spot” for insertional mutations. This region is small, only about 100 bp in length (fig. 1), and occurs in a core region (domain V) upstream of the D9 expansion segment (Gray and Schnare 1990) in a highly conserved section of the LSU rRNA gene, about one-third of the way upstream from its 3’ end. The rDNA-specific transposons insert into both helices and into unpaired regions between helices. However, what specific features of this particular region of the rDNA unit that account for its susceptibility to insertion by transposons is still unclear. Sullender (1993) showed that Pokey inserts into one particular TTAA target site in this region, even though other target sites exist in the LSU rRNA gene; in fact, another TTAA is just downstream of the actual insertion site. If Pokey did insert there, we would be able to detect it with the PCR primers that we use.

Before the discovery of Pokey, no DNA transposon was known to have such site specificity (Eickbush and Malik 2002), although Pokey does not seem to be site specific in other genomic locations beyond the specificity for TTAA. We have sequenced several hundred base pairs of the 3’ flanking region of four non-rDNA Pokey elements screened from the D. pulicaria cosmid library (unpublished data) and found no primary sequence similarity among these sequences and the LSU rRNA gene immediately downstream from the TTAA target site. However, further analysis is required to determine whether these sequences share other attributes to which Pokey is attracted. Even so, the fact that Pokey does occur at other genomic locations may have provided the advantage that it required to establish itself in the rDNA “hot spot,” despite the occupation of this region by the highly successful R1 and R2. In fact, the insertion site of R2 begins one nucleotide downstream from Pokey’s TTAA site (fig. 1), which suggests the possibility of competition between these elements. Preliminary PCR analysis suggests that R2 does occur in Daphnia (unpublished data). Thus, future study of the interaction between these elements within the rDNA of a single species may provide important insights into competition and survival patterns of transposable elements.

**Supplementary Material**

Both sequential (FastA) and interleaved Nexus files of the sequences used in this study are available at the Molecular Biology and Evolution Web site as Supplementary data.

**Supplementary Data File 1.** supplement-mbe040092–01. txt: Nucleotide FASTA file of 15 Daphnia mtSSU rRNA gene sequences and Interleaved NEXUS alignment file of 15 Daphnia nucleotide mtSSU rRNA gene sequences.

**Supplementary Data File 2.** supplement-mbe040092–02. txt: Nucleotide FASTA file of 12 Daphnia LSU rRNA gene sequences and Interleaved NEXUS alignment file of 12 Daphnia nucleotide LSU rRNA gene sequences.

**Supplementary Data File 3.** supplement-mbe040092–03. txt: Nucleotide FASTA file of 22 Pokey sequences from Daphnia and Interleaved NEXUS alignment file of 22 nucleotide Pokey sequences from Daphnia.

**Acknowledgments**

Financial support for this work was provided by a research grant from the Natural Sciences and Engineering Research Council to T.J.C. J. Colbourne, P. Hebert, C. Prokopowich, D. Taylor, and L. Weider generously provided samples of Daphnia. We thank A. Holliss for sequencing, Jeremy DeWaard for his assistance with the SH tests, and Hope Hollocher, Denis Lynn, and two anonymous reviewers for comments on earlier drafts of the manuscript.

**Literature Cited**


Brian Golding, Associate Editor
Accepted May 27, 2004