Phylogenetic Analysis of a Retrotransposon with Implications for Strong Evolutionary Constraints on Reverse Transcriptase

Bryant F. McAllister and John H. Werren

Department of Biology, University of Rochester

This study examines the evolutionary dynamics of a retrotransposon in a group of parasitoid wasps. A region containing the reverse transcriptase (RT) domain was sequenced for 43 elements from the genomes of nine different wasp species. Phylogenetic analysis of the elements revealed concordance with taxonomic classification of the host species, and the pattern was consistent with that expected for vertical transmission of a multicopy element during differentiation of the species. Twenty-three of the 43 elements had comparable intact open reading frames in the amplified region, and these were used in an analysis of evolutionary constraint on the amino acid sequence. As previously documented for retroelements, closely related elements exhibited nearly equal substitution rates at nonsynonymous and synonymous sites, but relative nonsynonymous substitution rates decreased as increasingly divergent elements were compared. A statistical test indicated that the decrease was not due to saturation of weakly selected sites. The pattern is most likely caused by a "pseudogene effect." Individual elements are not subject to purifying selection, and therefore, synonymous and nonsynonymous substitutions accumulate at equal rates. Comparisons among closely related elements are influenced strongly by this pseudogene evolution, whereas comparisons among distantly related elements reveal selection on the actively replicating lineages connecting the elements. These distant comparisons more accurately reflect the constraints on the amino acid sequence, and the comparisons among elements in this study indicated strong constraints on RT.

Introduction

Mobile genetic elements have been identified in the genomes of many organisms, including representatives of bacteria, fungi, plants, and animals. Mobile elements are classified into two major groups based on their mode of replication (Charlesworth, Sniegowski, and Stephan 1994). Retroelements (Class I) replicate via an RNA intermediate, whereas "classic" transposable elements or short inverted repeat elements (Class II) replicate via a DNA intermediate. The major biological difference between these two modes of replication is that, once inserted, retroelements remain in the genome (barring excision by recombination) and RNA transcripts are used as templates for replication. On the other hand, Class II elements replicate by excising from the genome and reinserting into another location.

The long-term evolutionary dynamics of mobile elements remain a matter of debate (Kidwell 1992; Capy, Anxolabéhère, and Langin 1994; Cummings 1994). Are mobile elements solely transferred vertically where they are confined within their host's genome and only inherited by the progeny of that host, or are mobile elements frequently transferred horizontally where they escape the confines of their host's genome and establish themselves in the genome of another species? Analyses of the phylogenetic patterns exhibited by mobile elements are providing evidence that distinguishes between these two possibilities. By examining the phylogenetic pattern of mobile elements present in many host species relative to the phylogenetic pattern of their hosts, the evolutionary history of the elements can be determined. However, the evolutionary complexity of multicopy elements requires a systematic survey to accurately reveal their long-term evolutionary history (Cummings 1994).

Recent sequence analyses of mobile elements have provided solid evidence to support both horizontal and vertical modes of evolution. Two Class II elements, *P* and *mariner,* appear to have undergone recent horizontal transfer. Similarity in nucleotide sequence indicates that the *P* element of *Drosophila melanogaster* was recently transferred into this species from a species in the willistoni group of *Drosophila* (Clark, Maddison, and Kidwell 1994). Based on sampling of the *mariner* element from widely divergent insect taxa, this element apparently has undergone widespread horizontal transfer (Robertson 1993; Robertson and MacLeod 1993). On the other hand, analyses of retrotransposons provide a number of examples where the element has been maintained by vertical transmission during the radiation of their host species. The elements which have been studied are: *LINE1* elements in rodents (Hardies et al. 1986; Casavant, Sherman, and Wichman 1996), *R1* and *R2* elements in *Drosophila* (Eickbush and Eickbush 1995; Lathe et al. 1995), *copia*-like elements in cotton (VanderWiel, Voytas, and Wendel 1993), and *SURL* elements in echinoids (Springer et al. 1995). One interesting correlation which is emerging from these studies is that RNA-mediated elements appear to be vertically maintained within host lineages, while DNA-mediated elements appear to undergo frequent horizontal transfer.

Although the pattern of retrotransposon evolution over a relatively narrow time period is indicative of vertical transmission, arguments have been made for horizontal transfer of retrotransposons over a long-term scale. All retroelements encode reverse transcriptase (RT), because this protein converts an RNA transcript into DNA during the replication cycle (Eickbush 1994; Flavell 1995). In comparisons among RT amino acid sequences from distantly related retroelements, the re-

Key words: retrotransposon, retroelement, molecular evolution, reverse transcriptase, Pteromalidae.

Present address and address for correspondence and reprints: Bryant McAllister, Department of Ecology and Evolution, 1101 E. 57th Street, The University of Chicago, Chicago, Illinois 60637-1573. E-mail: bryant@pondside.uchicago.edu.

Mol. Biol. Evol. 14(1):69–80. 1997

lationships among the elements generally do not reflect the phylogeny of their host species (Xiong and Eickbush 1990; Flavell 1992; Springer and Britten 1993; Eickbush 1994). This incongruence may represent horizontal transfer. Alternatively, phylogenetic analyses of the currently recognized elements may be inadequate for verifying long-term vertical maintenance of the elements (Cummings 1994). When elements from distantly related taxa are compared, it is difficult to determine orthology or paralogy among elements. Furthermore, comparisons among very divergent elements are confined to amino acid sequences, and unsubstantiated assumptions have been made concerning the rate of RT evolution (Flavell 1992). A measure of the constraints on RT will reveal the relative evolutionary rate of this protein sequence, thus providing a useful evaluation of the evidence for horizontal transfer.

This study analyzes the phylogenetic pattern of the retroelement *NATE* (NAsonia Transposable Element) in the genomes of nine parasitoid wasps, all occurring within the chalcid family Pteromalidae. *NATE* is a 11-kb dispersed repeat and contains a 1-kb long terminal repeat (LTR) oriented in the same direction at both ends of the element (McAllister 1995). Based on structural features and phylogenetic analysis of an internal amino acid sequence which is similar to RT sequences of other retroelements, *NATE* was described as a member of the *gypsy/Ty3* group of LTR retrotransposons (McAllister 1995). This group of elements is very interesting, because it apparently represents the sister group to retroviruses (Doolittle et al. 1989; Xiong and Eickbush 1990; Eickbush 1994) and may even be capable of infectious transfer (Kim et al. 1994). Although this group of elements is affiliated with infectious agents, a study of one member of this group supports its maintenance by vertical transmission (Springer et al. 1995).

NATE was originally isolated as an insertion into the supernumerary PSR (Paternal Sex Ratio) chromosome in *Nasonia vitripennis* (McAllister 1995). Study of its distribution and phylogenetic pattern was initiated to reveal the evolutionary history of PSR. The implications of the *NATE* phylogeny on the evolution of this chromosome will be treated with more detail elsewhere. In this paper, we examine the evolutionary dynamics of *NATE,* general patterns of retrotransposon sequence evolution, and evolutionary constraints on the RT domain. A critical finding of this analysis is that, although individual elements apparently evolve as pseudogenes, comparisons among distantly related elements reveal strong effects of purifying selection.

Materials and Methods

Primer Design

Procedures for isolating the original copies of *NATE* from the PSR chromosome have been described (McAllister 1995). These elements (PSR c41b, c17, c5, and c16) were isolated from a λ-clone genomic library constructed from PSR-carrying males in the MI strain of *N. vitripennis* (Reed et al. 1994). Sequences of the complete RT domain from these elements have been reported (GenBank accession U29470–U29473; McAllister 1995), and for this study were extended to an approximately 1,100-bp region. Cross-hybridizing elements were isolated from a *N. longicornis* genomic library. This *N. longicornis* λ-library contains genomic DNA of the IV-14 strain partially digested with *Sau*3AI and ligated into the EMBL3 λ-vector (Stratagene). A plasmid clone (P41-S5) containing the RT region from a PSR-inserted element (PSR c41b) was used as a probe to screen three genome equivalents of the *N. longicornis* library. This screen yielded many clones which hybridized to the probe and eight clones were isolated for DNA purification. The *N. longicornis* clones were examined by restriction digestion patterns and a preliminary sequencing analysis was performed. An approximately 1,100-bp region encompassing the RT domain was subcloned into plasmids and sequenced from two of the *N. longicornis* λ-clones (c21 and c105, GenBank accession U69494 and U69495). These four PSR and two *N. longicornis* reference sequences were obtained by directly subcloning into plasmids and sequencing both DNA strands.

Comparisons among the sequences from the four PSR elements and the two *N. longicornis* elements revealed regions of sequence conservation, thus identifying suitable primer sites which could be used to amplify this region from elements in genomic DNA. Two different primer sets were developed (fig. 1). The g-primer combination (N6, 5′ TAC ACC TCA AAG GCA CAG; N14, 5′ AGA CCA GCT TCG TTT ATC C) is a general primer set having high sequence conservation among all the reference sequences, and the product obtained from this reaction is approximately 770 bp. The p-primer set (N12, 5′ GCC TCC CTC TCG TCT GCA; N13, 5′ CAA TGG GTT TGA CTC GTT CG) was designed to amplify elements related to the elements on PSR and exclude the *N. longicornis* c105 element from amplification, and the reaction product is about 840 bp. These primer combinations were tested in control PCR reactions using reference elements in λ-clones as template. As expected, the general primer set (N6/N14) amplified a product from all the reference elements (PSR c41, *N. longicornis* c21, and *N. longicornis* c105), whereas the more specific p-primer set (N12/N13) did not amplify a product from *N. longicornis* c105.

PCR Amplification, Cloning, and Sequencing

Twelve wasp strains were screened for the presence of *NATE.* The taxonomic associations of these wasp species, collection localities of the strains, and providers are presented in table 1. Wasps were maintained in the laboratory on *Sarcophaga bullata* or *Musca domestica* pupae. Mass quantities of adult wasps (about 200) from each strain were used for DNA isolation and standard procedures were followed. Wasps were homogenized and solubilized in a solution containing 0.15 M NaCl, 0.1 M EDTA, 0.5% SDS, and 150 μg/ml proteinase K. Following phenol:chloroform (1:1) extractions, the DNA was precipitated with isopropanol. The DNA was resuspended in 50 mM Tris and 10 mM EDTA and treated with RNAse A (200 μg/ml). Additional phenol:chlo-

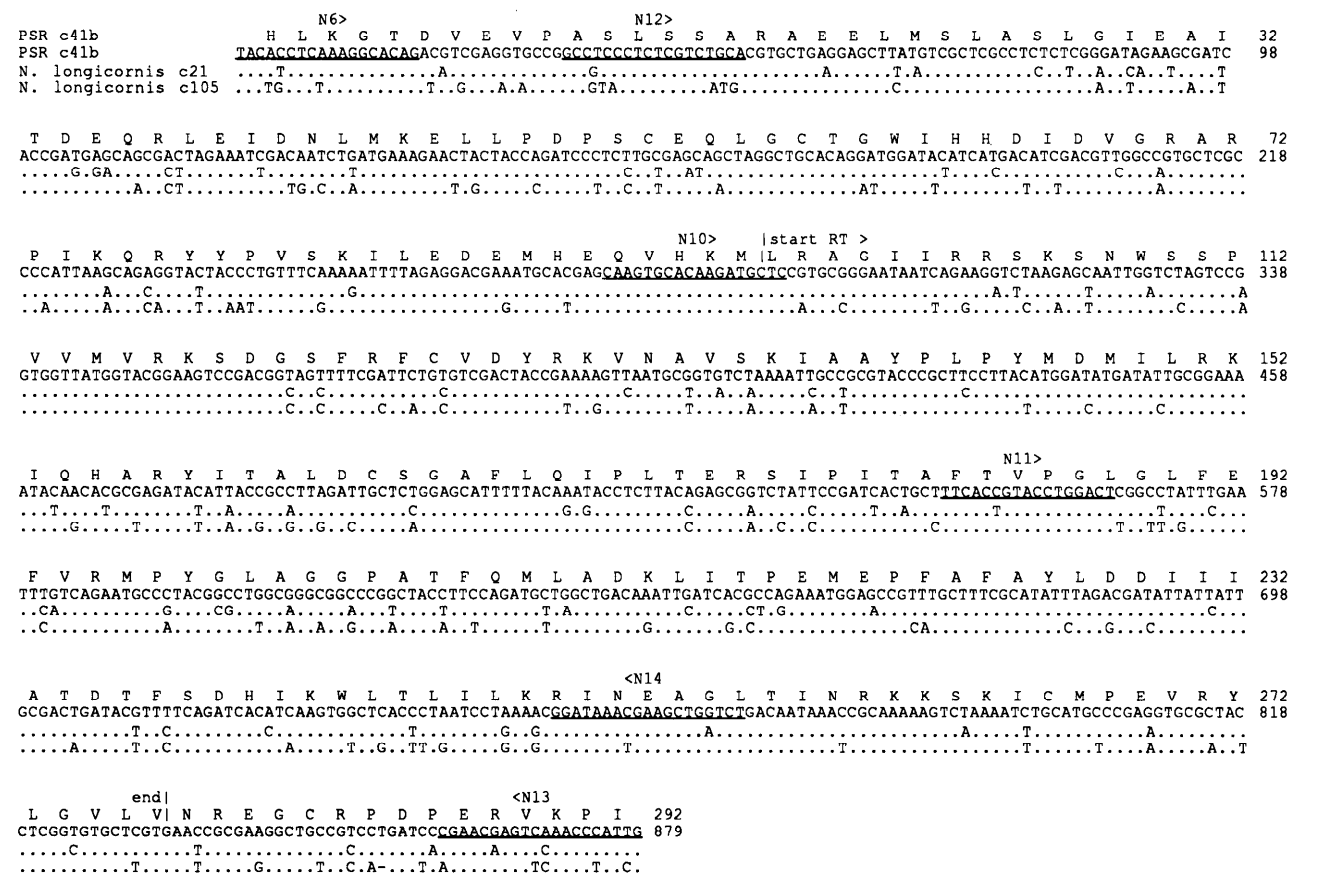

FIG. 1.—Nucleotide sequences for three reference elements in the amplified region of *NATE*, along with the conceptually translated amino acid sequence of PSR c41. Sites of the PCR primers N6, N12, N14, and N13 and the internal sequencing primers N10 and N11 are underlined. The region of the RT protein is indicated.

roform extractions and a chloroform extraction were performed, and the DNA was precipitated with ethanol.

Amplification of the RT region from elements in the genomic DNAs was performed by PCR. Reaction volumes were 50 μl and contained 1/10 reaction buffer (BRL), 2.5 mM $MgCl_2$, 100 nM each dNTP, 200 nM each primer, 2 units *Taq* polymerase (BRL), and approximately 5 ng template DNA. Reaction parameters were as follows: g-primers, two times (95°C, 2 min; 50°C, 1 min; 72°C, 2 min) followed by 30 times (95°C,

Table 1
Species Used for Isolation of Genomic DNA

Classification[a]	Collection Locality	Provider
Family Pteromalidae		
Subfamily Pteromalinae		
Tribe Pteromalini		
Nasonia vitripennis (PSR/MI)	Utah, USA[b]	J. Werren (U. Rochester)
Nasonia vitripennis (MI)	Illinois, USA	G. Saul (Middlebury College)
Nasonia vitripennis (LbII)	Leiden, Netherlands	J. van den Assem (U. Leiden)
Nasonia longicornis (IV-7)	Utah, USA	J. Werren (U. Rochester)
Nasonia giraulti (RV-2)	Virginia, USA	J. Werren (U. Rochester)
Trichomalopsis americanus	Nebraska, USA	J. Petersen (U. Nebraska)
Trichomalopsis dubius	New York, USA	D. Rutz (Cornell U.)
Muscidifurax raptor	New York, USA	D. Rutz (Cornell U.)
Muscidifurax uniraptor	Puerto Rico	E. Legner (U. C. Riverside)
Muscidifurax raptorellus	Nebraska, USA	D. Rutz (Cornell U.)
Urolepis rufipes	New York, USA	D. Rutz (Cornell U.)
Tribe Pachyneurini		
Pachycrepoideus vindeminae	New York, USA	D. Rutz (Cornell U.)
Subfamily Spalaniinae		
Spalangia cameroni	New York, USA	D. Rutz (Cornell U.)

[a] Classification based on Kogan and Legner (1970), Burks (1979), Hoebeke and Rutz (1988), and Darling and Werren (1990).
[b] Locality where PSR was collected.

30 s; 54°C, 1 min; 72°C, 1 min); and p-primers, two times (95°C, 2 min; 56°C, 1 min; 72°C, 2 min) followed by 30 times (95°C, 30 s; 60°C, 1 min; 72°C, 1 min). An Amplitron II thermocycler (Thermolyn) was used for cycling. Reaction products were assayed on 1.0% agarose gels.

In samples where a positive PCR product was obtained, about 15 μl of the PCR product was gel purified by electrophoresis in a 1.0% NuSieve GTG agarose (FMC) gel, and the band of the expected size was excised from the gel. DNA was purified from the agarose using the Gene-Clean kit (Bio 101, Inc.). The product was ligated into a modified M13mp18 vector containing T-tails (Burke, Müller, and Eickbush 1995). Single-strand DNA was isolated from at least two recombinant clones containing antisense strands of the PCR product from each wasp species.

Sequencing reactions were initiated on single-strand DNA using the −40 priming site in the vector. Reactions were performed using the Sequenase kit (USB) by labeling with [α^{35}S]dATP and eletrophoresed through buffer-gradient gels. The following internal primers were used to extend the sequences (fig. 1): N10, 5′ CAA GTG CAC AAG ATG CTC; N11, 5′ TTC ACC GTA CCT GGA CT. The PCR product sequences were read in one direction from a single strand and compared to the reference elements that were sequenced on both DNA strands. Most of the sequence was read from two gels, either by reading from the same reaction subjected to electrophoresis for two time frames (short and long runs), or by reading the overlapping frame produced by different sequencing primers. The PCR products contain a mixture of sequences representing different copies of the *NATE* in the genome; therefore, each unique sequence potentially, but not necessarily, represented a different element. Each unique sequence was treated as an individual element, although some sequences may represent the same element that appears different due to PCR error. Additionally, PCR error (estimated at about 0.2%, Eickbush and Eickbush 1995) can increase the apparent divergence of different elements. Where the artifacts of these methods may have affected the results, the implications are discussed.

Sequences were manually entered into the computer using ESEE ver. 2.0 (Cabot and Beckenbach 1989) where the sequences were visually aligned. Once the sequences were globally compared, all nucleotide substitutions and indels were verified by re-examining the sequencing gels. Phylogenetic analyses were performed with PAUP ver. 3.1.1 (Swofford 1991). To obtain the most parsimonious tree incorporating the sequence differences among all the elements, a heuristic search was performed with sequences added randomly 50 times. Confidence in the nodes was determined by 250 bootstrap replications using the heuristic search option with five random sequence additions. Trees were rooted at the midpoint. Phylogenetic analyses were also performed using the neighbor-joining algorithm in MEGA ver. 1.01 (Kumar, Tamura, and Nei 1993).

Synonymous/Nonsynonymous Evolution

Analyses of nucleotide substitutions among the elements with the same intact reading frames were used to determine the constraints on the amino acid sequence. Because the reading frames in elements with frame shifts and/or stop codons could not be unambiguously compared, these were not used in the analysis of RT evolution. A phylogenetic analysis was performed on the subset of elements with intact reading frames. The heuristic search option of PAUP ver. 3.1.1 was used for identifying the most parsimonious tree connecting the sequences. A 50% majority-rule bootstrap consensus tree, rooted at the midpoint, was used to infer the relationships among the sequences and identify the pattern of nucleotide substitution. Also, pairwise total sequence divergence among the elements was estimated according to the Jukes-Cantor method (Jukes and Cantor 1969). Proportional differences at synonymous (silent) sites and nonsynonymous (replacement) sites were estimated with MEGA version 1.01 (Kumar, Tamura, and Nei 1993) according to the algorithm of Nei and Gojobori (1986). These proportional differences were corrected according to Jukes and Cantor (1969). All pairwise comparisons were performed, and the means and standard deviations were calculated for groups of comparisons. The groups consisted of all pairwise comparisons among elements isolated from a single strain and in the same basal clade, and all pairwise comparisons between elements from separate strains. The mean number of synonymous substitutions per site (K_s) and nonsynonymous substitutions per site (K_a), K_s/K_a ratio of the means, and total mean divergence were calculated.

Results

To investigate diversity of *NATE* in *Nasonia* and related wasps, two sets of primers were used for amplifying a region of the element from genomic DNAs. The g-primers were designed for amplification of all the reference elements obtained through library screening; p-primers were more specifically designed to amplify from elements related to those found on the PSR chromosome (McAllister 1995). Eleven wasp species from six genera were examined: *Nasonia* (*N. vitripennis, N. longicornis, N. giraulti*), *Trichomalopsis* [=*Eupteromalus*] (*T. americanus, T. dubius*), *Urolepis rufipes, Muscidifurax* (*M. raptor, M. raptorellus, M. uniraptor*), *Pachycrepoideus vindeminae,* and *Spalangia cameroni. Nasonia, Trichomalopsis, Urolepis,* and *Muscidifurax* are relatively closely related wasps in the tribe Pteromalini, whereas *Spalangia* and *Pachycrepoideus* are more distantly related (Burks 1979). Under normal stringency reaction conditions (see *Materials and Methods*), the p-primers yielded a product from all Pteromalini species except *N. vitripennis* and *M. raptor.* The g-primers yielded a product from all Pteromalini species tested. Neither primer set amplified a product from the more distantly related wasps *P. vindeminae* and *S. cameroni.*

Sequences were obtained for 43 elements (including the four PSR and two *N. longicornis* elements obtained from λ-clones) representing the PSR chromosome

and nine wasp species (GenBank accession U29470–U29473, U69494–U69532). Element names are given as the species name (except for PSR elements) followed by the isolation method (c = genomic library cloning, g = PCR with g-primers, or p = PCR with p-primers) and an identification number. The consensus region present in these 43 elements contained a total of 700 bases (flanked by the N12 and N14 primer binding sites, fig. 1) and is presumably part of an extended open reading frame (ORF) that is transcribed, translated, and subsequently processed into different active proteins (Varmus and Brown 1989). The region represents 240 bases preceding and 460 bases into the 543-bp RT domain, using the definition of the RT domain of Xiong and Eickbush (1990) as previously applied to the *NATE* sequence by McAllister (1995). The protease domain is expected to precede the RT domain, based on the structure of related LTR retrotransposons in the *gypsy/Ty3* group (Smyth et al. 1989; Evgen'ev, Corces, and Lankenau 1992). Although 80 amino acid residues preceding the RT domain were examined, no similarity to the protease domain was detected in the inferred amino acid sequence. So, it appears the amplified region preceding RT is an intervening sequence between these two protein domains. An intact ORF was evident in 27 elements throughout the entire 700 bases, but 4 of these had compensatory frame-shift mutations. Insertions, deletions, or nucleotide substitutions causing stop codons in the ORF were observed in 16 of the elements.

These sequences isolated from nine taxa provided data for examining several aspects of retrotransposon evolution. First, relationships among the elements were compared to the taxonomy of the host species to infer the evolutionary history of the element. The phylogeny of the elements was examined for evidence of horizontal transfer or a pattern of vertical transmission during the differentiation of these wasp species. If the elements have been maintained by vertical transmission, the phylogenetic relationships among elements are expected to reflect the taxonomic status of the species. Second, the process of molecular evolution of *NATE* was examined. Incorporation of nucleotide substitutions at silent and replacement sites helped reveal the pattern of element evolution and provided evidence for strong constraints on RT.

Phylogenetic Analysis

Evolutionary relationships among these 43 elements were examined by phylogenetic analysis of the approximately 700-bp region sequenced for all the elements. A parsimony analysis of the data set revealed 16 trees with 791 evolutionary steps depicting the most parsimonious relationships among the elements. Differences among the 16 equally-likely trees were minor, and limited to placement of terminal taxa within the tree. Bootstrap analysis was used to determine the confidence in the nodes of the tree (Felsenstein 1985). The bootstrap consensus tree, containing nodes supported in more than 50% of the 250 bootstrap replicates, is presented in figure 2. The tree is rooted at the midpoint. Two major subfamilies of elements were revealed and designated subfamily I and subfamily II (fig. 2). Elements in subfamily I were represented in all the wasp species that yielded a PCR product, except for *N. vitripennis* and *M. raptor*. Subfamily II contained elements from the three *Nasonia* species and a single element from *M. raptor*. Analysis of the same data set using the neighbor-joining algorithm (Saitou and Nei 1987) produced a similar phylogenetic pattern.

Phylogenetic relationships among the *Nasonia* elements can be directly compared to an established phylogeny based on independent molecular markers. Analyses of sequence data from the ITS2 and 28s D2 regions of rDNA (Campbell, Steffen-Campbell, and Werren 1993) and a region of the 16s gene in mtDNA (unpublished) have established that *N. vitripennis* is more distantly related to the sister species *N. longicornis* and *N. giraulti*. Both analyses placed *T. dubius* as an outgroup to the *Nasonia* species; however, none of the other wasp species examined in this study were included. The phylad of *Nasonia* elements in subfamily II reflects the phylogeny of those species (fig. 2). Elements from *N. longicornis* and *N. giraulti* are closely related to each other, whereas elements from *N. vitripennis* are more distantly related. Only elements from *N. longicornis* and *N. giraulti* are presented in subfamily I, and these formed a clade separate from the other genera. Partial sequences of two elements from *N. vitripennis* were obtained, but are not presented in this analysis. In an analysis using sequences from a smaller region, these two *N. vitripennis* elements formed a clade with the other *Nasonia* elements in subfamily I. Again, these were more distantly related to the *N. longicornis/N. giraulti* elements. Contrary to the phylogenetic relationships among these species is that elements from the supernumerary PSR chromosome of *N. vitripennis* are contained within the group of elements from *Trichomalopsis* (fig. 2). This pattern relates to the evolutionary history of PSR.

Because systematic relationships among all of these wasp species have not been established, the remaining elements must be examined relative to the taxonomic status of their host species. The pattern exhibited by the elements is complex, but generally consistent with the taxonomy. Except in cases of divergent lineages being present in a single species, elements generally grouped within species and genera boundaries (fig. 2). Some of the species-specific clades contain very closely related elements which may be artifacts of PCR. A total of eight sequence pairs were less than 1.0% different; however, except for two pairs, other more divergent elements were present in the species-specific clades, indicating validity of the clades. Elements from each genus grouped into a strongly supported clade that was quite separate from the other genera. Relationships among these genera were not supported with high confidence, because at this level of comparison synonymous nucleotide positions were approaching saturation, thus creating noise in the data. By analyzing only nonsynonymous sites, it may have been possible to obtain a more reliable branching pattern among these genera. However, in the absence of a complete species phylogeny, there is no null hypothesis for comparing the relationships among

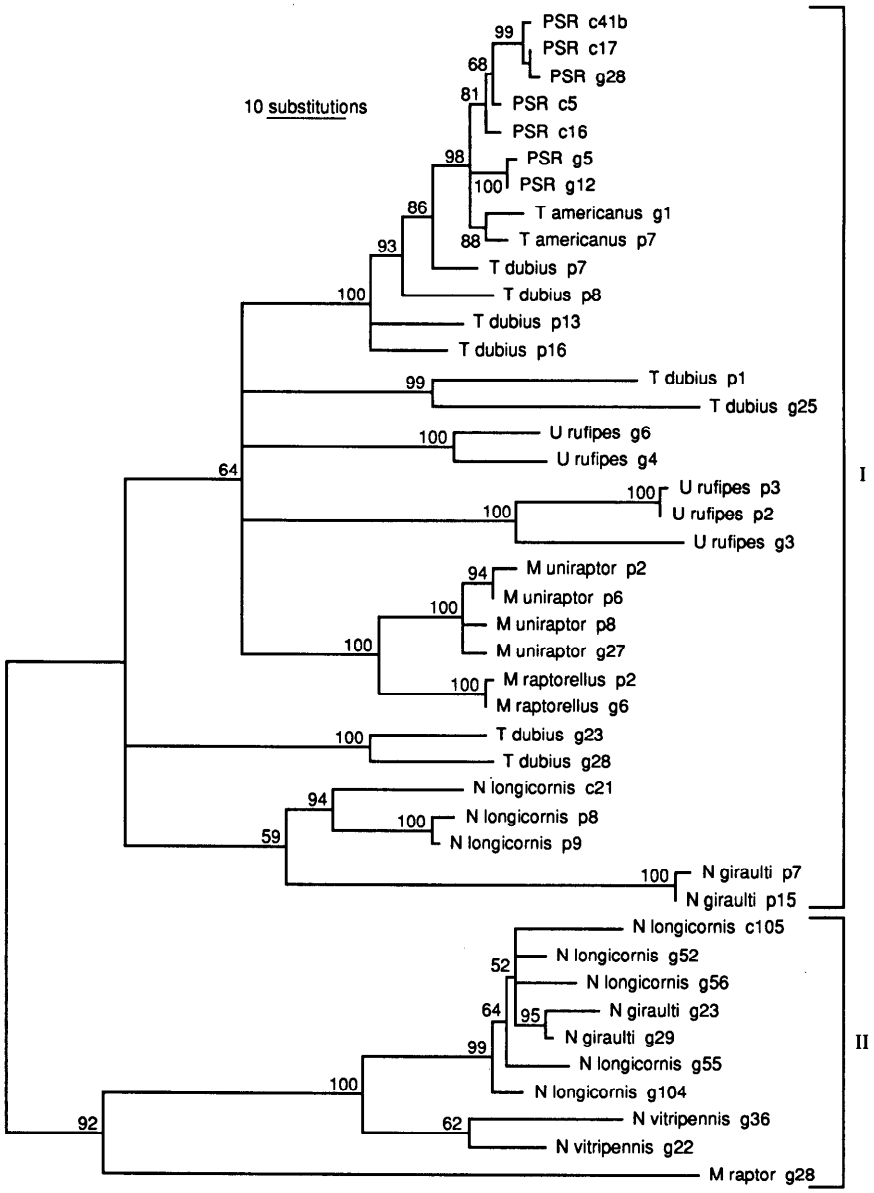

FIG. 2.—Phylogenetic relationships among the elements isolated from nine species of parasitoid wasps. The tree represents the bootstrap consensus following 250 replicates, and nodes with confidence values greater than 50% are indicated.

elements from the different genera. Nevertheless, the general pattern representing the taxonomy of the species is evident. Additionally, wasp species that yielded a PCR product are all classified in the tribe Pteromalini, whereas the two species classified outside of this tribe did not produce a product.

The overall pattern exhibited by these elements was somewhat obscured by the presence of multiple lineages within individual genomes (fig. 2). These divergent lineages apparently represent instances of long-term maintenance of multiple element lineages within an individual genome. Although some of these paralogous lineages appear old, none are unique to a single group of wasps. For example, the two divergent lineages in the *Nasonia* species apparently diverged prior to the divergence of this entire group of wasps, but neither was unique to *Nasonia*. Additionally, none of the paralogous lineages show evidence of horizontal transfer among these species, because all are more divergent than the basal splitting of the genera.

Two cases were observed where closely related elements from a single species did not clearly resolve into a distinct phylad. The *N. longicornis* elements in subfamily II do not clearly resolve into a separate phylad from the *N. giraulti* elements (fig. 2), although the two *N. giraulti* elements did significantly resolve into a distinct clade. The same pattern is apparent with the *T. dubius* elements and the *T. americanus*/PSR phylad in subfamily I (fig. 2). In both cases, the elements from the species that did not resolve had comparatively long branch lengths representing a large number of autapomorphic nucleotide substitutions. It appears as though these elements duplicated from a common element prior to the divergence of the species, and have remained in

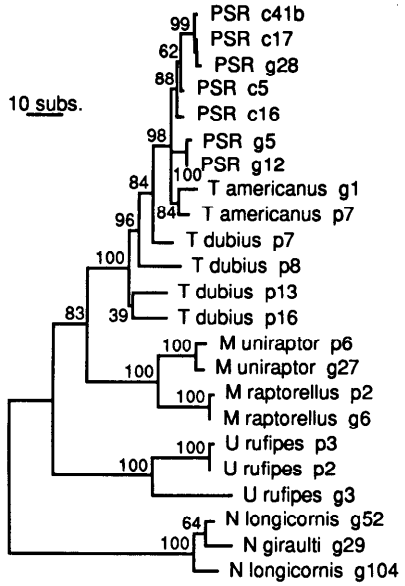

FIG. 3.—Phylogenetic relationships among the 23 elements used to analyze the substitution patterns. These elements all had intact reading frames in the region sequenced. The tree represents the 50% majority-rule bootstrap consensus tree incorporating 312 nucleotide substitutions.

the genome of the one species (*N. longicornis* and *T. dubius*) following the speciation event. In both cases, this element diversity was not detected in the genome of the closely related species (*N. giraulti* and *T. americanus*). More extensive sampling of these species may have revealed multiple clades representing the divergence between the two species.

Reverse Transcriptase Evolution

A total of 43 elements were represented by the sequences obtained from the λ-clones and PCR products. The sequenced region encompasses 80 amino acids (aa) preceding and 153 aa of the RT domain (fig. 1). Of the 43 elements, 23 had comparable protein reading frames, 4 had compensatory frame-shifts in the reading frame, and 16 had stop codons caused by base substitutions or frame-shifts. A conservative approach was taken to determine evolutionary constraints on the coding sequence by only analyzing the 23 sequences without frame-shifts or stop codons in the reading frame. A phylogenetic analysis was performed on these sequences, and the 50% majority-rule bootstrap consensus tree is presented in figure 3. To investigate the constraints on the amino acid sequence in the sequenced region, the ratio of synonymous (K_s) to nonsynonymous (K_a) substitutions was examined. This measure of the relative substitution rate at these sites is an indicator of purifying selection on a coding sequence. In the absence of purifying selection, the K_s/K_a ratio is expected to equal unity, and as the intensity of purifying selection increases, the ratio also increases. A total of 253 pairwise comparisons were performed, and the mean K_s, K_a, K_s/K_a, and total distance was obtained for all the intraspecific and interspecific comparisons. All intraspecific comparisons were within the same basal clade (fig. 3). As expected for a coding sequence, rates of substitution at synonymous sites were generally higher than at nonsynonymous sites. However, the K_s/K_a ratio increased as the evolutionary distance between the sequences being compared increased. This curious pattern has been observed in other studies of retrotransposons (Springer, Davidson, and Britten 1991; Lathe et al. 1995; Springer et al. 1995). Analyses were performed to quantify the pattern, test for its statistical significance, and test a proposed cause.

Using the mean distance measures among groups of elements, the ratio of K_s/K_a was examined relative to the total divergence between the sequences being compared. The ratio of synonymous/nonsynonymous substitutions per site versus the overall sequence divergence was plotted (fig. 4). A clear pattern emerges upon plotting these data; the ratio of K_s/K_a is dependent on the level of divergence between the sequences being compared. Sequences which are very closely related have

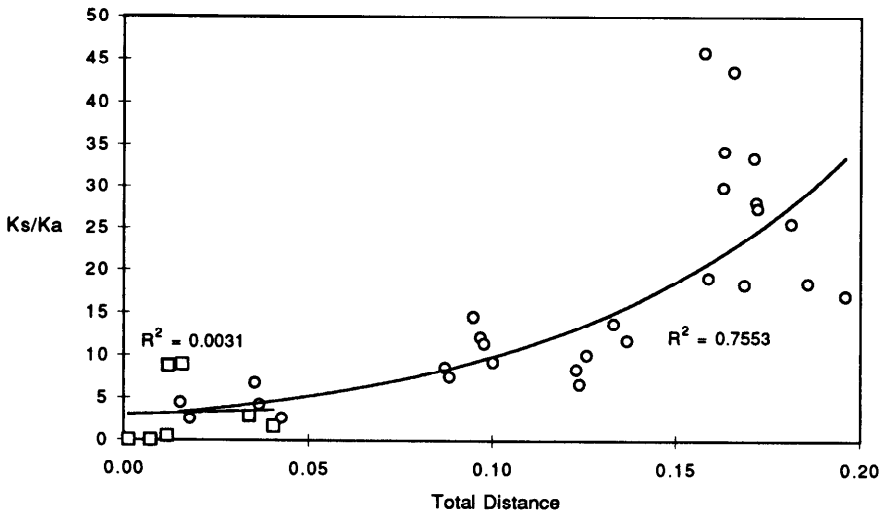

FIG. 4.—Relationship between the K_s/K_a ratio and total distance. Boxes represent means for comparisons among element clades within each species, and circles are means for each between-species comparison. Regression lines were fit independently for within- and between-species comparisons.

Table 2
Distributions of Nucleotide Substitutions

A. **Substitution Patterns at Different Levels of the Inferred Phylogeny**

	Silent	Replacement	Silent/Replacement	Weighted[a]
Single elements	37	61	0.61	1.85
Polymorphic in species	16	15	1.07	3.25
Fixed in single species	42	17	2.47	7.51
Present in multiple species	113	11	10.27	31.23

B. **Nucleotide Substitutions Resulting in the Same Amino Acid Replacements**

	Replacement Substitutions at a Nonsynonymous Site				
	0	1	2	3[b]	>3[b]
Observed sites	432	86	6	2	0
Expected sites[c]	431.6	85.3	8.4	0.6	0.1

[a] Silent/replacement substitutions weighted by the approximate number of sites.

[b] Cells were combined for χ^2 test.

[c] Calculated assuming the 104 replacement substitutions were Poisson distributed over the 526 nonsynonymous sites.

nearly equal rates of substitution, whereas very distantly related sequences have low rates of nonsynonymous relative to synonymous substitution. Regression lines were fit independently to the within-species and between-species comparisons (fig. 4). No significant correlation was observed for the within-genome comparisons (r^2 = 0.0031), whereas a positive correlation was observed for the between-genome comparisons, and an exponential regression (r^2 = 0.7553) fit the data better than a linear regression (r^2 = 0.5343). Although the pattern was evident and supported by the regression lines, assumptions about independence among the data are violated for regression analysis.

A different approach was applied to determine the statistical significance of the pattern observed in figure 4. This test was based on similar assumptions concerning the constancy in the rates of silent and replacement substitutions through time as the statistical test proposed by McDonald and Kreitman (1991) for the analysis of single-copy protein-coding genes. Three minimal-length trees of 312 steps were obtained following parsimony analysis of the 23 sequences. All differences among the trees were in the placement of *T. dubius* p13 and p16. The majority-rule bootstrap tree (fig. 3) was used to map the 312 nucleotide substitutions on the gene tree connecting the 23 sequences. Each silent or replacement substitution (convergent changes treated independently) was categorized into four classes (table 2*A*); (1) occurring in a single element (unique), (2) polymorphic within the elements of a clade from a single species, (3) fixed in all elements of a clade from a single species, or (4) present in elements from more than one species. These different categories were intended to represent persistence of nucleotide substitutions through increasing lengths of evolutionary time. As expected, with increased evolutionary time between elements, the number of accumulated silent substitutions increased (table 2*A*). However, replacement substitutions exhibited the opposite pattern. A total of 61 replacement substitutions were present in single elements and only 11 replacement substitutions were shared among elements isolated from more than one species. Patterns of silent and replacement substitution over this representation of evolutionary time were significantly ($P < 0.001$, $\chi^2 = 77$, df = 3) different. Some nucleotide substitutions observed in single elements (class 1) could have resulted from polymerase error during PCR. These would have occurred at silent and replacement sites proportional to the number of sites (approximately 1:3 for this sequence), thus biasing the ratio of silent to replacement substitutions in this class toward neutrality. A test of independence excluding the single-element class still revealed a statistically significant result ($P < 0.001$, $\chi^2 = 30.22$, df = 2).

One possible explanation for this phenomenon is that saturation has been reached at weakly selected amino acid sites when distantly related elements are compared (Springer, Davidson, and Britten 1991; Lathe et al. 1995; Springer et al. 1995). If so, the same nucleotide substitutions resulting in replacements should be occurring at weakly selected aa sites in different clades (convergent evolution to the same amino acid sequence). Weakly selected sites become saturated as more distantly related elements are compared, thus decreasing the observed rate of nonsynonymous substitution. To test this hypothesis, each of the 104 replacement substitutions was examined and identified as unique, or as occurring as independent events in different clades in the inferred phylogeny (table 2*B*). If the pattern represents saturation, nonsynonymous sites with two or more independent substitutions resulting in the same replacement should be more numerous than the random expectation. The expected distribution of the 104 replacement substitutions over the 526 nonsynonymous sites was calculated using the Poisson distribution ($\bar{x}$ = 0.1977). Fit between the observed and expected distributions was not significantly different ($\chi^2 = 3.668$, $P = 0.30$, df = 2). This result indicates that an excess of convergent nucleotide changes resulting in the same amino acid replacements were not present in these sequences.

Putative "dead" elements can also be informative of retroelement evolution. Of the 43 sequences, 16 had stop codons and 4 had compensatory frame-shifts in the reading frame. In 8 of the 20 elements with problems in the reading frame, the mutation causing the problem was unique to that element. The other 12 elements had mutations which were shared, so there were 6 element pairs that had the same mutations causing problems in the reading frame. Identifying different elements that share potentially deleterious mutations is important for revealing which elements in the genome are capable of replication. Three cases of shared mutations may be artifacts from the methods of obtaining sequences. In two cases, the sequence differences between the elements was low (<0.6%) suggesting that the sequences sharing defects may have amplified from the same dead element. One case represents the inability to resolve an inferred deletion in a sequencing compression. The other three cases were not attributed to error, and thus apparently

represent the replication of elements with mutations in their amino acid sequences. The elements *T. dubius* g23 and g28 had the same indels causing a compensatory frame-shift over 5 aa, and were 4.0% different at the sequence level. Two elements (g4 and g6) from *U. rufipes* also shared indels causing a compensatory frameshift over 35 aa; the two elements were 3.3% different in sequence. A nucleotide substitution resulting in a stop codon was shared by *N. vitripennis* g22 and g36, which were 4.3% different at the sequence level. These three sequence pairs suggest the existence of "parasitic" elements that can replicate with proteins encoded by functional elements.

Discussion

When examining the phylogeny of a transposable element in a group of host organisms, the null hypothesis is that the element has been transmitted vertically through the germ line (Cummings 1994). Except for the elements residing on the supernumerary PSR chromosome, no patterns were observed in the phylogeny of *NATE* in this group of parasitoid wasps that are inconsistent with vertical maintenance of the element during the divergence of these species. Elements were isolated from nine species representing four different genera. The genome of each species generally yielded at least one monophyletic group of elements, and elements from multiple species within the same genus all formed monophyletic groups representative of the genus. Because the grouping of elements by taxonomic status of these wasp species was evident, the evolutionary history of *NATE* apparently represents an additional case of vertical maintenance of a retroelement. Several aspects of the *NATE* phylogeny were not representative of a single-copy nuclear gene, but are typical of interspecific comparisons among multicopy elements. More specific analyses of these patterns await a more complete molecular phylogeny of these genera.

The one pattern in the *NATE* phylogeny inconsistent with vertical transmission was the position of elements isolated from the supernumerary PSR chromosome of *N. vitripennis*. Elements on PSR were closely related to elements amplified from the genome of *Trichomalopsis americanus,* and the PSR elements were contained in a clade of elements represented by both *T. americanus* and *T. dubius*. Because the PSR elements were contained within this *Trichomalopsis* clade, loss of these elements within the autosomal complement of the three *Nasonia* species with retention on PSR is not sufficient to explain the relationship between the PSR elements and the *Trichomalopsis* elements. The PSR chromosome has only been detected in *N. vitripennis* populations in the great basin region of North America, although the species occurs worldwide (Werren 1991; Beukeboom 1992). This limited distribution is suggestive of a recent origin for PSR in *N. vitripennis*. The relationship between elements on PSR and in the genome of *Trichomalopsis* apparently represents transfer (presumably by intergeneric hybridization) of the intact PSR chromosome or a chromosomal fragment which evolved into PSR. These copies of *NATE* are not the only regions of PSR which exhibit similarity with the genome of *Trichomalopsis* (unpublished).

Analysis of the phylogenetic relationships among the elements were complicated by the presence of multiple lineages within single species. In analyses of element phylogeny to infer the pattern of historical maintenance, presence of multiple lineages must be considered (Cummings 1994). If elements are sparsely sampled from individual genomes, an element phylogeny may be obtained that is inconsistent with host phylogeny. Thorough sampling of elements increases the likelihood of identifying all members of orthologous lineages, and these orthologous elements should accurately reflect the host phylogeny when maintained vertically. Presence of multiple divergent element lineages within a single genome is common for retroelements (Burke et al. 1993; VanderWeil, Voytas, and Wendel 1993; Lathe et al. 1995; Springer et al. 1995; Casavant, Sherman, and Wichman 1996). The common explanation for multiple lineages is ancient divergence among independent lineages. Alternatively, multiple lineages in a single genome may represent the acquisition of a divergent lineage through horizontal transfer. The paralogous lineages in the *NATE* phylogeny are consistent with vertical maintenance of a multicopy element. All of the major paralogous lineages are contained within the entire element phylogeny and represented by more than one group of wasps. Furthermore, there is no evidence of horizontal transfer among these species because all the paralogous lineages diverged prior to the basal splitting of the genera.

Extinction of an element lineage occurs when all elements in a lineage lose the ability to replicate (stochastic loss). Although it could not be substantiated with independent data, a possible example of stochastic loss has occurred in the ancestor of the three *Nasonia* species. Based on sequences of the ITS and 28s D2 regions of rDNA (Campbell, Steffen-Campbell, and Werren 1993) and mtDNA (unpublished), *T. dubius* is closely related to the *Nasonia* species. Also, preliminary sequencing data indicate that the *Muscidifurax* species represent a divergent group separated from a clade containing *Trichomalopsis* sp., *Nasonia* sp., and *Urolepis* sp. (unpublished). Given these relationships among the genera, it would appear that an entire lineage of elements has been lost in the common ancestor of the *Nasonia* species. Elements should have been detected in *Nasonia* that were contained in the clade represented by PSR, *T. americanus, T. dubius,* and *U. rufipes* (fig. 2). The absence of elements was not due to experimental limitations, because an extensive search in the *Nasonia* genome using PCR and restriction digests of the product failed to identify this lineage of elements and the presence of these elements was not revealed by Southern hybridization (unpublished). Because retroelements exhibit the propensity for maintaining multiple element lineages within a single genome, and with the added factor of lineage extinction, this may preclude the detection of all members of an orthologous group. As more disjoined host taxa are examined, and as the similarity

among elements decreases, the probability of obtaining an element phylogeny which is incongruent with host phylogeny increases. This may explain the incongruent phylogenies that are observed in comparisons among extremely divergent retroelements (Xiong and Eickbush 1990; Flavell 1992; Springer and Britten 1993; Eickbush 1994).

The evolutionary history of *NATE* in this group of parasitoids was very similar to the pattern exhibited by the *SURL* element in echinoids (Springer et al. 1995). Both of these elements are members of the *gypsy/Ty3* group of LTR retrotransposons (Springer and Britten 1993; McAllister 1995), a group which is closely allied with retroviruses (Xiong and Eickbush 1990). The PCR primers used to amplify *NATE* from these wasp species were also used in attempts to amplify homologous elements from potential "sources" or "sinks" for horizontal transfer. No elements were amplified from genomic DNA of flies (*Protocalliphora* sp., *Calliphora* sp., or *Sarcophaga bullata*) used as hosts by *Nasonia* (unpublished). It has been shown that a cytoplasmic microorganism (*Wolbachia*) has probably been horizontally transferred between *Nasonia* and *Protocalliphora* (Werren, Zhang, and Guo 1995), so if *NATE* is prone to horizontal transmission, this is an ecological association where it could occur. With the relationship between LTR retrotransposons of nonvertebrates and retroviruses of vertebrates, it is interesting that the two elements studied thus far exhibit patterns of vertical transmission.

Sequences of RT are commonly used in the comparisons among distantly related elements, and a measure of the constraints on this protein may help evaluate these comparisons. The region of *NATE* that was analyzed in this study provides a measure of the evolutionary constraints on RT. Evolutionary constraint on the RT amino acid sequence was inferred by examining the relationship between silent (synonymous) and replacement (nonsynonymous) substitutions in pairwise comparisons between elements. In the absence of purifying selection, the substitution rate at synonymous (K_s) and nonsynonymous (K_a) sites should be equal, but as the intensity of purifying selection increases, a corresponding decrease in K_a should also be observed. Examining the ratio K_s/K_a provides a measure of purifying selection. In comparisons among *NATE,* an interesting pattern was observed. Closely related elements exhibited low values for K_s/K_a, but as more distantly related elements were compared the value of the ratio increased greatly (fig. 4). This pattern has also been noted in other studies of retroelements (Springer, Davidson, and Britten 1991; Lathe et al. 1995; Springer et al. 1995). In a statistical test of the *NATE* data, the patterns of silent and replacement substitutions were significantly different among different levels of divergence. A large number of replacement substitutions were observed in terminal branches (autapomorphic changes) within a phylogenetic framework, and replacement substitutions decreased in more ancient branches of the phylogenetic tree.

Two hypotheses have been proposed to account for the decrease in replacement substitutions as increasingly divergent elements are compared (Springer, Davidson, and Britten 1991; Springer et al. 1995), but neither hypothesis has been directly tested. The first hypothesis is that weakly selected amino acid sites rapidly saturate; thus, nonsynonymous divergence reaches a maximum as synonymous divergence increases. A prediction of this model is that an excess of nonsynonymous sites should exhibit independent (convergent) nucleotide substitutions resulting in the same amino acid replacements. Statistical analysis of the *NATE* data did not support this prediction. Only 8 of the 526 nonsynonymous sites exhibited convergent substitutions leading to identical replacements, which was lower than the random expectation given the observed number of nonsynonymous substitutions.

An alternative hypothesis for explaining this pattern is the "pseudogene effect." Standard coding genes are subject to purifying selection which eliminates replacement substitutions resulting in reduced protein function. In contrast, following the insertion of a retrotransposon into the genome, nucleotide substitutions accumulate in the element without being subjected to purifying selection. In effect, each element acts as a "pseudogene" free to accumulate both synonymous and nonsynonymous substitutions; thus, each terminal element lineage should evolve with equal per-site rates of synonymous and nonsynonymous substitution (Hardies et al. 1986; Springer, Davidson, and Britten 1991; Springer et al. 1995). The pseudogene effect arises from comparisons among elements to measure K_s and K_a, because the value of these distance measures are affected by two components. The pseudogene component is the equal per-site accumulation of synonymous or nonsynonymous substitutions in individual elements since their insertion, and the lineage component is the accumulation of nucleotide substitutions under selective constraints revealed by the replicating ancestral lineages connecting the extant elements. Comparisons among closely related elements are primarily influenced by the terminal pseudogene component, with negligible input from the connecting lineage. Correspondingly, the ratio of K_s/K_a is close to unity. On the other hand, comparisons among distantly related elements are primarily influenced by selective constraints on the lineages connecting the elements, with negligible input from the pseudogene component. Distant comparisons should reflect the selective constraints on the amino acid sequence. This is exactly the pattern that was observed in the *NATE* sequences.

Comparisons among divergent elements indicated that there are strong evolutionary constraints on the RT amino acid sequence of *NATE.* The mean K_s/K_a ratio for the 12 most distant comparisons was 28.4 (range = 17.0–45.9). Examining this measure of constraint on RT compared to other protein sequences (Li, Wu, and Luo 1985) reveals that RT is a slowly evolving protein. For example, histone H2A exhibits a K_s/K_a ratio of 26 in comparisons among mammalian orders (Li, Wu, and Luo 1985). Alternatively, *Adh* exhibits an average K_s/K_a ratio of 9.7 in nine pairwise comparisons among species of *Drosophila* (Lathe et al. 1995). The fact that RT exhibits such strong selective pressures for maintaining its amino acid sequence should not be surprising, since this

is the polymerase used by retroelements for replication. Other studies (Hardies et al. 1986; Springer, Davidson, and Britten 1991; Lathe et al. 1995; Springer et al. 1995) have not reported such high restrictions on replacement substitutions in the RT sequence; however, this is the first study comparing a large number of elements with variable degrees of relatedness over the majority of the RT domain.

Evidence for strong evolutionary constraints on RT has bearing on reports that elements isolated from distantly related taxa exhibit high similarity in their amino acid sequences, thus supporting horizontal transmission. Flavell (1992) reported that *copia*-like retroelements in the barley and potato genomes exhibited an amino acid identity of 57% in a segment of the RT domain, whereas monocots and dicots diverged approximately 200 MYA. This level of conservation was used as support for horizontal transfer because the retroelements are unnecessary for host survival and are subjected to higher (nucleotide) mutation rates due to their error-prone replication (Flavell 1992). The argument fails on both grounds. Elements must replicate for survival through 200 Myr of host evolution; thus, there has been selection for the ability to replicate. A high nucleotide substitution rate only influences the pseudogene component of retroelement evolution and it does not necessarily invoke high rates of nonsynonymous substitution in the lineage component. All nucleotide substitutions are subjected to purifying selection during the long-term maintenance of an actively replicating element lineage.

An additional factor that may compound the pseudogene effect is the apparent ability of retroelements with deleterious substitutions in their own reading frame to "parasitize" the proteins encoded by other elements in the genome. In studying the mechanism of retrotransposon replication, helper elements have been used to encode proteins required for replication of RNA from donor elements containing the required *cis*-acting sequences (Kirchner, Sandmeyer, and Forrest 1992). This indicates that replication ability is not dependent on encoding proteins, but requires certain *cis*-acting sequences. Parasitization of replication machinery within the genome may explain the observed distribution of excess replacement substitutions observed among *NATE*. When the K_s/K_a ratio was plotted relative to evolutionary distance, an exponential increase was observed. The spread of low K_s/K_a values over such a large time frame may indicate that elements with deleterious substitutions replicate using functional proteins encoded by other elements. Additionally, three element pairs exhibited shared deleterious changes in the reading frame, indicating replicative persistence.

This study of *NATE* provides an additional example of vertical maintenance of a retrotransposon in a group of closely related species. By examining the pattern of nucleotide substitution in this element, strong constraints on replacement substitutions in RT were evident. If these constraints on RT are universal among retroelements, evidence for horizontal transfer of these elements must be reevaluated. The ubiquitous distribution of these elements may be due to their presence prior to the divergence of eukaryotes. Alternatively, retroelements may undergo infrequent horizontal transmission, but no examples of recent transfer have been identified. Furthermore, this study emphasizes the need to understand how these elements are so successful at persistence in genomes over long periods of time.

Acknowledgments

The authors would like to thank T. Eickbush, U. Bergthorsson, R. Weston, and two anonymous reviewers for helpful discussions and for comments on the manuscript. Provision of wasp strains by S. Long and J. Petersen is appreciated. This project was supported by a grant from the National Science Foundation to J. Werren. This research is part of the Ph.D. dissertation submitted by B. McAllister, who was partially supported by an Ernst Caspari Fellowship.

LITERATURE CITED

BEUKEBOOM, L. W. 1992. Genetics and population biology of the paternal sex ratio (PSR) chromosome. Ph.D. dissertation, University of Rochester, Rochester, N.Y.

BURKE, W. D., D. G. EICKBUSH, Y. XIONG, J. JAKUBCZAK, and T. H. EICKBUSH. 1993. Sequence relationships of retrotransposable elements R1 and R2 within and between divergent insect species. Mol. Biol. Evol. **10**:163–185.

BURKE, W. D., F. MÜLLER, and T. H. EICKBUSH. 1995. R4, a non-LTR retrotransposon specific to the large subunit rDNA genes of nematodes. Nucleic Acids Res. **23**:4628–4634.

BURKS, B. D. 1979. Family Pteromalidae. Pp. 768–835 *in* K. V. KROMBEIN, P. D. HURD JR., D. R. SMITH, and B. D. BURKS, eds. Catalog of Hymenoptera in America north of Mexico. Smithsonian Institution Press, Washington, D.C.

CABOT, E. L., and A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences using ESEE. Comput. Appl. Biosci. **5**:233–234.

CAMPBELL, B. C., J. D. STEFFEN-CAMPBELL, and J. H. WERREN. 1993. Phylogeny of the *Nasonia* species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer (ITS2) and 28S rDNA sequences. Insect Mol. Biol. **2**:225–237.

CAPY, P., D. ANXOLABÉHÈRE, and T. LANGIN. 1994. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. **10**:7–12.

CASAVANT, N. C., A. N. SHERMAN, and H. A. WICHMAN. 1996. Two persistent LINE-1 lineages in *Peromyscus* have unequal rates of evolution. Genetics **142**:1289–1298.

CHARLESWORTH, B., P. SNIEGOWSKI, and W. STEPHAN. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature **371**:215–220.

CLARK, J. B., W. P. MADDISON, and M. G. KIDWELL. 1994. Phylogenetic analysis supports horizontal transfer of *P* transposable elements. Mol. Biol. Evol. **11**:40–50.

CUMMINGS, M. P. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol. Evol. **9**:141–145.

DARLING, D. C., and J. H. WERREN. 1990. Biosystematics of *Nasonia* (Hymenoptera: Pteromalidae): two new species reared from birds' nests in North America. Ann. Entomol. Soc. Am. **83**:352–370.

DOOLITTLE, R. F., D.-F. FENG, M. S. JOHNSON, and M. A. MCCLURE. 1989. Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64**:1–30.

EICKBUSH, D. G., and T. H. EICKBUSH. 1995. Vertical transmission of the retrotransposable elements *R1* and *R2* during the evolution of the *Drosophila melanogaster* species subgroup. Genetics **139**:671–684.

EICKBUSH, T. H. 1994. Origin and evolutionary relationships of retroelements. Pp. 121–157 *in* S. S. MORSE, ed. The evolutionary biology of viruses. Raven, New York.

EVGEN'EV, M. B., V. G. CORCES, and D.-H. LANKENAU. 1992. Ulysses transposable element of *Drosophila* shows high structural similarities to functional domains of retroviruses. J. Mol. Biol. **225**:917–924.

FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FLAVELL, A. J. 1992. *Ty1-copia* group retrotransposons and the evolution of retroelements in the eukaryotes. Genetica **86**: 203–214.

———. 1995. Retroelements, reverse transcriptase and evolution. Comp. Biochem. Physiol. **110B**:3–15.

HARDIES, S. C., S. L. MARTIN, C. F. VOLIVA, C. A. HUTCHISON III, and M. H. EDGELL. 1986. An analysis of replacement and synonymous changes in the rodent L1 repeat family. Mol. Biol. Evol. **3**:109–125.

HOEBEKE, E. R., and D. A. RUTZ. 1988. *Trichomalopsis dubius* (Ashmead) and *Dibrachys cavus* (Walker): newly discovered pupal parasitoids (Hymenoptera: Pteromalidae) of house flies and stable flies associated with livestock manure. Ann. Entomol. Soc. Am. **81**:493–497.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

KIDWELL, M. G. 1992. Horizontal transfer. Curr. Biol. **2**:868–873.

KIM, A., C. TERZIAN, P. SANTAMARIA, A. PÉLISSON, N. PRUD'HOMME, and A. BUCHETON. 1994. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91**:1285–1289.

KIRCHNER, J., S. B. SANDMEYER, and D. B. FORREST. 1992. Transposition of a Ty3 *GAG3-POL3* fusion mutant is limited by availability of capsid protein. J. Virol. **66**:6081–6092.

KOGAN, M., and E. F. LEGNER. 1970. A biosystematic revision of the genus *Muscidifurax* (Hymenoptera: Pteromalidae) with descriptions of four new species. Can. Entomol. **102**: 1268–1290.

KUMAR, S., K. TAMURA, and M. NEI. 1993. MEGA: molecular evolutionary genetics analysis. Version 1.01. The Pennsylvania State University, University Park, Pa.

LATHE, W. C. III, W. D. BURKE, D. G. EICKBUSH, and T. H. EICKBUSH. 1995. Evolutionary stability of the *R1* retrotransposable element in the genus *Drosophila*. Mol. Biol. Evol. **12**:1094–1105.

LI, W.-H., C.-I. WU, and C.-C. LUO. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.

MCALLISTER, B. F. 1995. Isolation and characterization of a retroelement from a B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis*. Insect Mol. Biol. **4**:253–262.

MCDONALD, J. H., and M. KREITMAN. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351**:652–654.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

REED, K. M., L. W. BEUKEBOOM, D. G. EICKBUSH, and J. H. WERREN. 1994. Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis*: association of palindromes with recombination. J. Mol. Evol. **38**:352–362.

ROBERTSON, H. M. 1993. The *mariner* transposable element is widespread in insects. Nature **362**:241–245.

ROBERTSON, H. M., and E. G. MACLEOD. 1993. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Mol. Biol. **2**:125–139.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SMYTH, D. R., P. KALITSIS, J. L. JOSEPH, and J. W. SENTRY. 1989. Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the gypsy group of *Drosophila*. Proc. Natl. Acad. Sci. USA **86**:5015–5019.

SPRINGER, M. S., and R. J. BRITTEN. 1993. Phylogenetic relationships of reverse transcriptase and RNase H sequences and aspects of genome structure in the gypsy group of retrotransposons. Mol. Biol. Evol. **10**:1370–1379.

SPRINGER, M. S., E. H. DAVIDSON, and R. J. BRITTEN. 1991. Retroviral-like element in a marine invertebrate. Proc. Natl. Acad. Sci. USA **88**:8401–8404.

SPRINGER, M. S., N. A. TUSNEEM, E. H. DAVIDSON, and R. J. BRITTEN. 1995. Phylogeny, rates of evolution, and patterns of codon usage among sea urchin retroviral-like elements, with implications for the recognition of horizontal transfer. Mol. Biol. Evol. **12**:219–230.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Smithsonian Institution, Washington, D.C.

VANDERWIEL, P. L., D. F. VOYTAS, and J. F. WENDEL. 1993. *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J. Mol. Evol. **36**:429–447.

VARMUS, H., and P. BROWN. 1989. Retroviruses. Pp. 53–108 *in* D. E. BURG and M. M. HOWE, eds. Mobile DNA. American Society for Microbiology, Washington, D.C.

WERREN, J. H. 1991. The paternal-sex-ratio chromosome of *Nasonia*. Am. Nat. **137**:392–402.

WERREN, J. H., W. ZHANG, and L. R. GUO. 1995. Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. Proc. R. Soc. Lond. B **261**:55–63.

XIONG, Y., and T. H. EICKBUSH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9**:3353–3362.

DANIEL L. HARTL, reviewing editor

Accepted September 24, 1996