

## Diversity of errantivirus (retrovirus) sequences in two cell lines used for baculovirus expression, *Spodoptera frugiperda* and *Trichoplusia ni*

Tila Menzel · George F. Rohrmann

Received: 7 December 2007 / Accepted: 11 March 2008 / Published online: 25 March 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** To determine the diversity of retroviruses (errantiviruses) in cell lines used for baculovirus expression, degenerate primers were designed complementary to conserved regions of lepidopteran errantivirus reverse transcriptase genes. These primers were used to PCR amplify sequences from DNA isolated from *Spodoptera frugiperda* (Sf-9) and *Trichoplusia ni* (Hi-5) cell lines. Cloning, sequencing, and phylogenetic analysis of over 20 PCR products from each cell line demonstrated the presence of diverse populations of retrovirus sequences comprising at least six major lineages.

**Keywords** Insect retrovirus · Errantivirus · *Trichoplusia ni* · *Spodoptera frugiperda* · Baculovirus

### Introduction

Expression of proteins using baculoviruses involves the infection of insect cells with a recombinant virus and the subsequent harvesting of the protein from these cultures. Constructs of the *Autographa californica* multicausid nucleopolyhedrovirus (AcMNPV) are commonly used for this process and two cell lines are normally employed for their replication. These were derived from primary explants of pupal ovarian tissue from moths of the family Noctuidae.

One is from the fall army worm (*Spodoptera frugiperda* (Sf)) [1]. The cell line called Sf9 was derived from the original Sf cell line (Sf21) and is the only Sf line available through the American Type Culture Collection (ATCC) and is widely used for baculovirus expression. Another major cell line used for AcMNPV expression is derived from the cabbage looper, *Trichoplusia ni* (Tn).

Although Tn and Sf cell lines are widely used for baculovirus expression, they have not been well characterized. A predicted feature of these cells is the presence of a variety of endogenous retroviruses called errantiviruses. Although they are likely present in these cell lines, their response to baculovirus infection of their host cell is not clear. One such virus, TED, was originally identified because it integrated into a baculovirus that was replicating in *T. ni* (Tn368) cells [2, 3]. It has been demonstrated to express gag and env genes and is associated with elevated levels of reverse transcriptase activity [4–6]. Although this virus was found in Tn368 cells [7], the *T. ni* cells most commonly used for baculovirus expression are called High-Five™ cells and were derived independently of Tn368 cells [8, 9]. No such viruses have been identified in Sf or Tn High-Five™ cells and no systematic examination of either cell line has been reported.

Errantiviruses are members of the Metaviridae, a major group of LTR-containing retroelements, which are characterized by a reverse transcriptase domain order of protease, reverse transcriptase-RNase H, integrase [10]. The Metaviridae include three genera, the semotiviruses which are a distinct and apparently ancient lineage, the metaviruses, and the errantiviruses. Individual members of all three of these lineages may contain an env-like orf, but the errantiviruses and a lineage of the semotiviruses apparently evolved from recombination events in which their progenitors incorporated an env orf from a

---

The GenBank/EMBL/DDBJ accession number of the sequence reported in this paper are EU082436–EU082444 for sf-17, 18, 20, 31, 37, 58, 67, 70, 311, respectively, and EU082455–EU082455 for Tn-45, 75, 13, 201, 243, 246, 256, 257, 273, 283, respectively.

---

T. Menzel · G. F. Rohrmann (✉)  
Department of Microbiology, Oregon State University,  
Nash Hall Room 220, Corvallis, OR 97331-3804, USA  
e-mail: rohrmann@orst.edu

baculovirus [11]. The Retroviridae are also thought to have been derived from the Metaviridae based on the relatedness of the reverse transcriptase sequences [10]. Although errantiviruses resemble retroviruses, they have not been included within the Retroviridae because they are a distinct lineage, and evidence that they are infectious is indirect. Similar to retroviruses, they have primer binding sites reflecting distinct species of tRNA. In the case of errantiviruses, they are divided into two phylogenetic groups [12] based on a primer binding site of either tRNA<sup>Ser</sup> or tRNA<sup>Lys</sup>.

Five different categories of errantiviruses with 78 complete or partial sequences have been found in the sequenced genome of *Drosophila melanogaster*, and they range from a single full length copy of *gypsy* to 18 full length copies of the element 297 [13]. A sixth category, *zam*, was not found in this sequence indicating the variability of errantivirus distribution between *D. melanogaster* strains. Errantiviruses have also been identified in a number of other insect species including a lepidopteran, *T. ni*. Two categories of errantivirus-like elements (BmRT6 and BmRT7) have been reported in the silkworm genome. These elements have genomes (8–9 kb) as one would predict for retroviruses and are present in over 300 copies each [14].

Errantiviruses have at least two novel relationships with baculoviruses. They appear to have evolved when they obtained their env gene from a baculovirus [15, 16]. This may have caused the conversion of a non-infectious retrotransposon into an element with the potential for infection. Such a recombination event appears to have occurred at least twice [11] and likely occurred when a retroelement integrated into a baculovirus genome during the infection of an insect cell. A lepidopteran retrovirus, TED, was originally found as an integrant in a baculovirus genome [2]. In addition, TED and several other errantiviruses have baculovirus late promoter elements in their LTR. This could lead to their genomes being transcribed at very high levels when they are integrated into baculovirus genomes [3, 17].

In this report we describe a survey of errantivirus sequences in DNA from cultured cells of *S. frugiperda* (Sf) and *T. ni* (High-Five<sup>TM</sup> cells) (Tn).

## Materials and methods

*Spodoptera frugiperda* 9 (Sf-9) cells [1] and *T. ni* 5 cells (High-Five<sup>TM</sup>) [8] were used (see also [18]) for DNA preparation. Two degenerate primers were synthesized based on conserved amino acid sequence domains encoded by the reverse transcriptase genes from lepidopteran errantiviruses. The sequences used were from *T. ni* (TED) [3], *Lymantria dispar* (Lydia) [17], and *Bombyx mori*

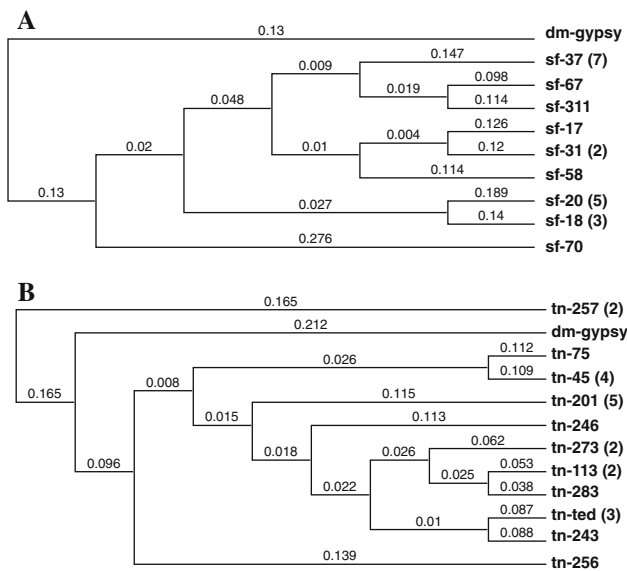
(001107) [14]. The sequences are separated by about 240 nt and encode the amino acid sequences PIWVVPKK (CCI ATH TGG GTI GTN CCN AAR AAR) and MPMGKKN (reverse complement: RTT YTT IAR ICC CAT NGG CAT). The abbreviations are I = inosine; N = any nucleotide; R = A/G; H = A/C/T; Y = T/C. The PCR reactions used platinum taq polymerase (Invitrogen) following the manufacturer's instructions except that the primers were used at 70 μM/ml. The PCR reaction is as follows: 95°, 5 min then 10 cycles of 94°, 1 min; 47°, 1 min; 72°, 1 min followed by 50 cycles of 94°, 20 s; 47°, 45 s; 72°, 45 s and then ending with 72° for 10 min. PCR products were then cloned directly into Topo TA cloning vectors (Invitrogen) or the PCR products were used as a template for a second round of PCR amplification using the same program. Clones containing the correct-sized sequence were identified by PCR using the forward and reverse primers, grown up and purified using Qiagen miniprep columns, and then sequenced. All sequencing was done by the DNA sequence core facility at Oregon State University. PCR amplification was done a number of times on each DNA preparation and usually only a few clones containing errantivirus sequences were identified from each preparation.

Sequences were analyzed using the MacVector 9.5.2 suite of software analysis programs. The sequences complementary to the degenerate sequencing primers were deleted from the sequences and the data used for analysis comprised the sequences located between the two primers.

## Results and discussion

Two degenerate primers were synthesized specific for conserved regions of the reverse transcriptase genes from lepidopteran errantiviruses. The sequences (PIWVVPKK and MPMGKKN) are separated by about 240 nt and are conserved in retrovirus sequences from three Lepidoptera; *T. ni* (TED) [3], *L. dispar* (Lydia) [17] and *B. mori* (001107) [14]. Using a pair of degenerate primers and an extended amplification protocol (see above), PCR products were amplified from DNA from both cell lines (data not shown). Aliquots of the product were cloned, screened, and sequenced, and the data generated was used for phylogenetic analysis.

The alignment of 22 nucleotide sequences from *S. frugiperda* resulted in the identification of nine different viruses that appear to fall into three major groups represented by sf-70, sf-20 and sf-18, and the rest of the sequences. Some of the sequences identified were represented up to seven times in our sample (e.g. sf-37, Fig. 1A). One of the sequences, sf-70 was as distant from the other sequences as was the sequence of the outgroup sequence from a dipteran virus (*D. melanogaster*, *gypsy*) that was included in the analysis.



**Fig. 1** Phylogenetic trees of lepidopteran errantivirus nucleotide sequences. **(A)** Tree derived from *S. frugiperda* sequences. **(B)** Tree derived from *T. ni* sequences. Tree was derived using MacVector software. The method used was Neighbor joining; best tree, tie breaking = systematic; Distance: Uncorrected ‘p’; Gaps distributed proportionately. The numbers following the clone in brackets are the number of times that the identical clone or a closely related sequence was identified in the cloned PCR product population. For example, the sf-37 sequence was identified in 7 clones and sf-20 was found in 5

The alignment of 23 nucleotide sequences from *T. ni* resulted in the identification of 11 different viruses present in three major groups (tn-256, tn-257, and the rest of the sequences) (Fig. 1B). Some of the sequences were represented up to five times in our sample set (tn-201). This included three sequences identical to TED, the first identified lepidopteran retrovirus found in Tn368 cells [2]. One lineage, tn-257, was significantly different and showed the highest sequence similarity to a reverse transcriptase sequence from the genome of *Danio rerio*, the zebra fish. This was also the only sequence that we identified among the Sf and Tn clones that was not completely in frame. It has one stop codon within the reverse transcriptase orf (Fig. 2) otherwise it is 57% identical over 73 amino acids with the zebra fish sequence.



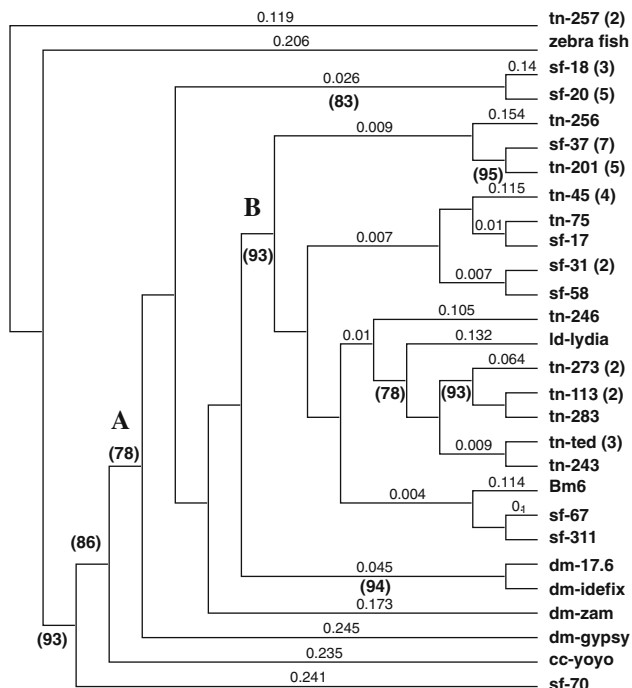
**Fig. 2** Alignment of predicted amino acid sequences from representative nucleotide sequences. The asterisks below the sequence indicate identical amino acids. The • in the tn-257 sequence indicates a stop codon. The sequences do not include the flanking regions from

In order to examine the relationship between the retrovirus sequences from *S. frugiperda* and *T. ni*, we combined the sequence information as a single data set and also included sequences from other errantiviruses including *Drosophila zam*, *idefix*, 17.6, *Ceratitis capitata yoyo*, *L. dispar (lydia)*, *B. mori* (Bm6), and also the sequence from zebra fish. Of the Sf and Tn sequences, the majority (18 out of 20) fall within a node with a bootstrap value of 78 (labeled A in Fig. 3) that includes *Drosophila gypsy*, 17.6, and *idefix*, but not cc-yoyo, tn-257, sf70, or the zebra fish lineages. Most of the lepidopteran lineages are located downstream of a well-supported node labeled B in Fig. 3 (bootstrap value = 93), although most of the subsequent phylogenetic relationships derived from sequences from different species are not well supported. The number and the diversity of sequences in this group suggest that they might be highly active in transposition and this could have led to their amplification and subsequent speciation. The fact that the one active element, TED, is found in this group supports this observation. Tn-257 may represent an inactive lineage as the two clones that we identified have stop codons in their reverse transcriptase sequence (Fig. 2). Sf-20 is represented by five clones in the amplified populations that is second only to sf-37 with seven clones. Therefore, these may represent active lineages or may have been preferentially cloned due to procedure that we used.

From these data it is clear that there is a high degree of diversity of these sequences in the Sf genome. For example, sf-70 is about as distant from the other sequences as the gypsy element (about 50% nucleotide sequence identity). Dm-gypsy, sf-20, tn-256, and TED are all over 50% identical and tn-256 and TED show almost 70% identity. These data are reflected in the amino acid sequence identities where tn-257 and sf-70 show about 46% or less identity with the other sequences and tn-256 and TED are quite similar at 85% identity. An amino acid alignment of these sequences is shown in Fig. 2. Again it shows the diversity of this region with 15 of about 80 amino acids invariant and another 14 that are similar.

The sequences of ~240 nt that we amplified was relatively short compared to the complete sequence of an errantivirus genome (7–8 kb). In order to determine if the

which the primers used for PCR amplification were derived. The asterisks indicate identical amino acids and the dots indicate amino acids with similar properties



**Fig. 3** Phylogenetic tree derived from a combination of *S. frugiperda* and *T. ni* nucleotide sequences. For details see legend Fig. 1. The numbers in brackets indicate confidence/bootstrap values. Bootstrap values were calculated using the following method: Neighbor joining; Bootstrap (1,000 reps); tie breaking = systematic. Distance: Uncorrected 'p'; Gaps distributed proportionally. Nodes with bootstrap values 60% or greater are indicated. The homologous sequences of *zam* (dm-zam), *lydia* (ld-lydia) from *L. dispar*, *B. mori* (Bm6), *dm-idefix*, *dm-17.6*, *cc-yoyo*, and zebra fish are included for comparison. Their accession/reference numbers are AJ000387; AF177773; reference 14, scaffold 001107; aj009736, x01472, u60529 and XM\_001337743, respectively

amplified sequences reflect phylogenies derived from more extensive sequence data, we examined the phylogenetic trees of selected *Drosophila* errantiviruses using sequences homologous to the ~240 nt segment. Sequences for *17.6*, *idefix*, *yoyo*, *zam*, *ted*, and *gypsy* showed a similar pattern of relatedness using these sequences (data not shown) as the complete reverse transcriptase gene [15]. Therefore, the data generated using the ~240 nt sequence likely reflects patterns that would be derived from more complete sets of sequences.

This report demonstrates that two lepidopteran cell lines contain a set of errantiviruses with a diversity similar to that reported for the *D. melanogaster* genome [13]. Because of the evolutionary relationships between baculoviruses and

errantiviruses that have been documented, additional interactions may exist. For example, the initiation of a baculovirus infection could trigger the transposition of the errantiviruses thereby promoting their incorporation into baculovirus genomes. This could be accomplished through direct stimulation of transposition or indirectly through the suppression of a silencing mechanism in the host insect. Although no evidence for errantivirus induction during baculovirus infection has been documented, examination of the full complement of these elements using sensitive technology could yield definitive information.

**Acknowledgement** This research was supported in part with a Tartar research grant.

## References

1. J.L. Vaughn, R.H. Goodwin, G.J. Tompkins, P. McCawley, *In vitro* **13**, 213–217 (1977)
2. D.W. Miller, L.K. Miller, *Nature* (London) **299**, 562–564 (1982)
3. P.D. Friesen, M.S. Nissen, *Mol. Cell. Biol.* **10**, 3067–3077 (1990)
4. K.L. Hajek, P.D. Friesen, *J. Virol.* **72**, 8718–8724 (1998)
5. R.A. Lerch, P.D. Friesen, *J. Virol.* **66**, 1590–1601 (1992)
6. M.S. Ozers, P. Friesen, *Virology* **226**, 252–259 (1996)
7. W.F. Hink, *Nature* **226**, 466–467 (1970)
8. T.R. Davis, T.J. Wickham, K.A. McKenna, R.R. Granados, M.L. Shuler, H.A. Wood, *In vitro Cell. Dev. Biol. Anim.* **29A**, 388–390 (1993)
9. R.R. Granados, L. Guoxun, A.C.G. Derksen, K.A. McKenna, *J. Invertebr. Pathol.* **64**, 260–266 (1994)
10. J.D. Boeke, T. Eickbush, S.B. Sandmeyer, D.F. Voytas, in *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, ed. by C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger, L.A. Ball (Elsevier, San Diego, 2004), pp. 359–367
11. H.S. Malik, S. Henikoff, *PLoS Genet.* **1**, 429–443 (2005)
12. C. Terzian, A. Pélisson, A. Bucheton, *BMC Evol. Biol.* **1**, 3 (2001)
13. J. Kaminker, C. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D. Wheeler, S. Lewis, G. Rubin, M. Ashburner, S. Celniker, *Genome Biol.* **3**(12), 12–23 (2002)
14. J.-S. Xu, Q.-Y. Xia, J. Li, G.-Q. Pan, Z.-Y. Zhou, *Insect Biochem. Mol. Biol.* **35**, 921–929 (2005)
15. H.S. Malik, S. Henikoff, T.H. Eickbush, *Genome Res.* **10**, 1307–1318 (2000)
16. G.F. Rohrmann, P.A. Karplus, *BMC Evol. Biol.* **1**, 1 (2001)
17. T.A. Pfeifer, M. Ring, T.A. Grigliatti, *Insect Mol. Biol.* **9**, 349–356 (2000)
18. D.R. O'Reilly, L.K. Miller, V.A. Luckow, *Baculovirus expression vectors. A laboratory manual*. (Oxford University Press, New York, 1994)
19. A.L. Vanarsdall, K. Okano, G.F. Rohrmann, *J. Virol.* **80**, 1724–1733 (2006)