Insect Molecular Biology (2006) **15**(4), 435–443

piggyBac-like elements in the tobacco budworm, *Heliothis virescens* (Fabricius)

J. Wang,* X. Ren,† T. A. Miller† and Y. Park*

**Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS, USA;* †*Department of Entomology, University of California, Riverside, CA, USA*

Abstract

We identified two different groups of *piggyBac*-like elements (PLE) in the tobacco budworm, *Heliothis virescens*, and named them *HvPLE1* and *HvPLE2*. An intact copy of *HvPLE1* revealed the characteristics of PLE: inverted terminal repeats, inverted subterminal repeats, and an open reading frame encoding transposase, whereas other *HvPLE1* copies and all the *HvPLE2* copies carried disruptive mutations in the region encoding transposase. We also identified none to two bands per genome hybridized to a probe of *Trichoplusia ni piggyBac* in genomic Southern blotting, which are different from *HvPLE1* or *HvPLE2*. Analysis of the sequences of multiple copies of *HvPLE1* and *HvPLE2* suggests that the PLEs are closely related to the *T. ni piggyBac*, of relatively young age, and independently entered the *H. virescens* genome.

Keywords: *piggyBac*-like elements, PLE, transposon, tobacco budworm, vectorette PCR, inverse PCR, transposable element display.

Introduction

Transposable element (TE) is a mobile genetic component in the genome of an organism. Active TEs are self-replicating and mobile in the genome, thus functioning as a mutagen that drives evolutionary processes by causing deleterious or adaptive mutations (reviewed in Kazazian, 2004). The importance of TE in the evolutionary process has been recognized by the ubiquitous presence of TE-derived sequences in the genomes of nearly all eukaryotes and by the horizontal transfer of TE crossing the species barrier. The use of TE for generating transgenic organisms has recently been also recognized as an important tool. Delivering a gene of interest into the genome of an organism to generate a novel phenotype is a promising tool for studying basic science and for various applications of transgenic technology (reviewed in Handler, 2002a; Wimmer, 2003; Robinson *et al*., 2004). The *piggyBac*-based random mutagenesis, gene tagging and enhancer trapping system have also been developed in *Drosophila* and in other arthropods (Lorenzen *et al*., 2002; Horn *et al*., 2003; Thibault *et al*., 2004).

The *piggyBac* element, one of the most widely used TE in transgenesis of insects and mammals (Handler, 2002b; Ding *et al*., 2005), was originally discovered in a cell line of cabbage looper moth, *Trichoplusia ni* (Fraser *et al*., 1983; Cary *et al*., 1989). The *T. ni piggyBac* is 2472 bp in length and contains 13 bp inverted terminal repeats (ITRs) with asymmetrically located 19 bp inverted subterminal repeats. The *piggyBac* transposes via a strict 'cut-and-paste' mechanism, inserting exclusively at 5′-TTAA-3′ target sites and excising precisely, leaving no footprint (Fraser *et al*., 1996). The mobility of *piggyBac* in a broad range of insects indicates that the *piggyBac*-like elements (PLE) might be widely distributed in distantly related insects and other species beyond *T. ni*. For example, a TE that is nearly identical to the *T. ni piggyBac* was found in the Oriental fruit fly (Handler & McCombs, 2000). Fifty divergent PLE sequences recently were identified from genome sequences of a diverse range of organisms, including fungi, plants, insects, crustaceans, urochordates, amphibians, fishes and mammals (Sarkar *et al*., 2003).

We were interested in the distribution of the PLEs in a lepidopteran insect pest, the tobacco budworm, *Heliothis virescens* (Fabricius). We report two distinct PLEs *HvPLE1* and *HvPLE2*, and the presence of a unique band hybridized by *T. ni piggyBac* in the *H. virescens* genome.

Results

Heliothis virescens piggyBac-*like elements* HvPLE1, HvPLE2 *and* HvPLE3

The degenerate primers PLEF266 and PLER351 successfully amplified a 263 bp fragment *HvPLE1*, having 43%

10.1111/j.1365-2583.2006.00653

Received 8 December 2005; accepted after revision 23 February 2006. Correspondence: Yoonseong Park, Department of Entomology, 123 Waters Hall, Kansas State University, Manhattan, KS 66506, USA. Tel.: 785 532 5763; fax: 785 532 6232; e-mail: ypark@ksu.edu. Jianjun Wang: Current address, Department of Plant Protection, Yangzhou University, Yangzhou 225009, China.

identity with *T. ni piggyBac* at the amino acid level in the amplified region. Subsequent inverse PCR yielded a number of partially overlapping clones revealing ITR and 5′- and 3′-flanking sequences. A representative copy of full-length *HvPLE1*, designated as *HvPLE1.1*, eventually was isolated by using the primers located on the flanking region. The *HvPLE1.1* is 2285 bp long with perfect 16 bp ITRs flanking a single open reading frame (ORF) encoding for a transposase of 581 amino acid residues (Fig. 1A). Similar to other PLE, the *HvPLE1.1* contains three C/G residues at the extreme ends of the both ITR (Elick *et al*., 1997) and the tetranucleotide target-site duplication, TTAA (Fig. 1A). Amino acid sequence alignment showed that the translation of *HvPLE1.1* transposase shares 42% and 64% amino acid identity and similarity with *Bombyx mori yabusame-1*, and 32% and 56% amino acid identity and similarity, respectively, with *T. ni piggyBac* (Fig. 2). The PSORT webserver (http://psort.njbb.ac.jp) predicts two bipartite nuclear localization signals (Robbins *et al*., 1991) in DNA binding and the catalytic domain: RRHNIMRETPGPKGRAK, starting at amino acid residue position 99, and RKNKPEIPTCFQP-KRTR, starting at position 364 (Fig. 1A).

Multiple copies of *HvPLE1* were obtained with PCR by using the primer designed in the ITR. A total of four clones from three insects were sequenced and named *HvPLE1.2* (950 bp), *HvPLE1.3* (1083 bp), *HvPLE1.4* (1103 bp), and *HvPLE5* (1108 bp). These copies were > 98% identical to the *HvPLE1.1* sequence, but with large deletions encompassing most of the region encoding the transposase (Fig. 1C).

We found a distinct type of PLE among the clones in the ITR PCR, which we named *HvPLE2*. Nine copies of *HvPLE2* from five different individual insects were isolated, ranging in length from 1361 to 1369 bp. All copies of *HvPLE2* contained multiple stop codons and frameshifts in the ORF. The nucleotide sequences of the nine copies *HvPLE2.1–HvPLE2.9* were > 93% identical to one another (Fig. 1B).

All *HvPLE2* copies obtained contain putative large deletions in their 5′ region. Thus, reliable sequence alignment was possible only for a portion of the transposase-encoding region from 467 to 1136 bp (Figs 1B and 2). We were unable to find the copy of *HvPLE2* comparable with the full-length of other PLEs in repeated ITR PCR with a number of different individual *H. virescens*. The 222 amino acids corresponding to the putative ORF region of *HvPLE2* have 56% identity and 78% similarity with *HvPLE1*, 42% identity and 65% similarity with *Bombyx mori yabusame-1*, and 40% identity and 60% similarity with *T. ni piggyBac* (Fig. 2). As a result, *HvPLE1* and *HvPLE2* were recorded as two distinct groups in phylogenetic analysis (Fig. 3).

In a low stringency genomic Southern hybridization (washing at 50 °C), we observed distinct bands hybridized with the probe of the 1.8 kb *T. ni piggyBac* transposase ORF, implying the presence of a low copy number *PLE* similar to *T. ni piggyBac* (Fig. 4).

Distribution and diversity of piggyBac*-like elements in* Heliothis virescens

To estimate copy numbers and diversity of PLEs among individuals of *H. virescens*, we performed genomic Southern hybridization for *Eco*RI digested DNA by using a number of different probes (Fig. 4). The hybridization pattern with the probe *HvPLE1*-core (Fig. 1C) indicated 0–3 putative full-length copies (bands) per individual ($n = 18$); the hybridization pattern with the probe *HvPLE1*-common and with *HvPLE2* (Fig. 1C) indicated five to 10 copies (bands) per individual (Fig. 4). The repeated hybridizations on the same membrane after stripping were designed to prevent cross-hybridization between *HvPLE1* and *HvPLE2* by high stringency hybridization and wash conditions. A Southern hybridization with the *T. ni piggyBac* probe at low stringency revealed zero to two bands per genome ($n = 14$, Fig. 4 and additional data not shown), indicating the presence of a low copy number that is similar to *piggyBac* of *T. ni* (Fig. 4). We were unable to clone the PLE closer to the *T. ni piggyBac* than *HvPLEs* in repeated degenerate PCRs with various stringencies, probably due to low number of the copy close to *T. ni piggyBac* being masked by other high copy number PLEs, such as *HvPLE1* and *HvPLE2*.

Because the Southern blotting revealed considerable degrees of diversity in the banding patterns ($n = 18$) in both *HvPLE1* and *HvPLE2*, we examined 10 additional individuals by using Vectorette PCR for insertion site variation. As was shown in the genomic Southern blotting, large variations in the band patterns for 5′- and 3′-sides in both *HvPLE1* and *HvPLE2* were observed in the Vectorette PCR (Fig. 5).

Insertion sites and inverted terminal repeats

We recovered a total of eight 5′- and 12 3′-flanking sequences of *HvPLE1* elements from four individual insects by inverse PCR and Vectorette PCR (Fig. 6). BLAST search of the flanking sequences of the *HvPLEs* found no significant match to known genes in the database, indicating that the insertions are likely located in non-coding

Figure 1. Nucleotide sequence and conceptual translations of putative transposase for the full-length *piggyBac*-like element *HvPLE 1.1* (A), nucleotide consensus sequence and putative translation of *HvPLE2* (B), and their structures (C). The putative 16 bp inverted terminal repeats (ITRs) are boxed; the inverted subterminal repeats are underlined. The upstream and downstream flanking sequences are in lower case. The potential duplicated tetranucleotide targets TTAA are in bold. Two regions of bipartite nuclear localization signal (NLS) are underlined (Robbins *et al*., 1991). The conserved regions corresponding to the degenerate primers used in this study are boxed on the *HvPLE1*. The indels in *HvPLE2* are indicated by hash marks (#) and the stop codons by asterisks (*). Structures of *HvPLE1* and *HvPLE2* are depicted with inverted terminal repeats (arrowheads) and transposase open reading frames (box). Lines indicate the clones obtained for entire sequences with the deletions (brackets). Inverted boxes are the regions used for probes in Southern hybridizations.

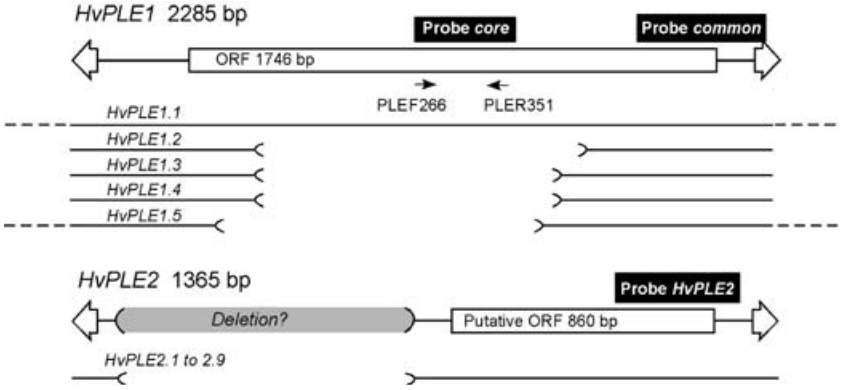
A

```
   1  gttgaggctttttaaCCCTTAATTACTCGCGTGGGGTATATTTAACCCCAcagcatatttgaagcgttgtatctc
  76  CACCATTTTTAAAGGTACCCTCATAGCGTTCTTAGTAGCTGCCTGAATACCTCCCCCCGCTGCAAATCAGTCGCG
 151  GGCGCCGCTCTGTGTCGTGTGGTTGTGTGAGGAGATCGCGTTTGGGGTAAATGTTGCCCCACGAGAGTGTTAGTG
 226  CGACTTTTTTTTGGACGAGTGAGTTTTTTTTATATTTGCTTTTTATGTGCACTACAAATACTAAACTGTTAGTTTT
 301  TTTAATTATCCTTGTTTTCTGTTTTTTTTTAGGTTAGAATGTCTAAAATGGCAAGTAAGCGTCTTTCTAATACGC
                                                  M  S  K  M  A  S  K  R  L  S  N  T
 376  AGATTGTTGATGTATTAGAAGACGATGAAGAATGCATCATAGATAGTCCCGATGAAGATGAGGTAGATGCTGAAG
      Q  I  V  D  V  L  E  D  D  E  E  C  I  I  D  S  P  D  E  D  E  V  D  A  E
 451  TCATCCAGAGTGACCACAACTCAGAATCTCAAGAAAGTGCCAATGAAATCGATTCAGAATGGAGTGATACCGATA
      V  I  Q  S  D  H  N  S  E  S  Q  E  S  A  N  E  I  D  S  E  W  S  D  T  D
 526  ATGAACCGCTGAGTAGACTCGCAAGTGAGGATTCTGATAACTTTTATTTTTCAAAAAATAAATGCACAAAGTGGG
      N  E  P  L  S  R  L  A  S  E  D  S  D  N  F  Y  F  S  K  N  K  C  T  K  W
 601  CAAAAGAGCCGCCGCGTACTAGCGTTCGAGTTCGTAGACATAATATTATGAGGGAAACACCAGGACCTAAGGGAA
      A  K  E  P  P  R  T  S  V  R  V  R  R  H  N  I  M  R  E  T  P  G  P  K  G
 676  GAGCCAAAGAGATACAAACAATCTCAGAAGCATTCTTTTGTATGTTTTCAATGGACACAGTCAATTTAGTACTGC
      R  A  K  E  I  Q  T  I  S  E  A  F  F  C  M  F  S  M  D  T  V  N  L  V  L
 751  AACAGACCAACGATTACATAAAATCAATTCAAGAAAAGTTTCAACGTGAACGTGATTGCAAAGTCCTAGAATACG
      Q  Q  T  N  D  Y  I  K  S  I  Q  E  K  F  Q  R  E  R  D  C  K  V  L  E  Y
 826  AAGAACTTTTGGCTTACCTAGGTCTGCTTTACATGTCTGGTGTGCTACGTTCATCTCACTTGAACTTCAAAGATC
      E  E  L  L  A  Y  L  G  L  L  Y  M  S  G  V  L  R  S  S  H  L  N  F  K  D
 901  TGTGGGCTACCGATGGGACGGGAATCGAGTTTTTTCAAAACACAATGTCATTCAATAGATTTCTTTTTATATCAA
      L  W  A  T  D  G  T  G  I  E  F  F  Q  N  T  M  S  F  N  R  F  L  F  I  S
 976  GATGTGTTCGTTTCGATGACAAAAATACGAAGTCTGAAAGACTAAAGACAGACAAACTAGCAGCAGTCAGAGAGT
      R  C  V  R  F  D  D  K  N  T  K  S  E  R  L  K  T  D  K  L  A  A  V  R  E
1051  TCACTGATTTGATGAATAATAATTTCATCAATAATTACTGTGCGTCTGAAAATGTTACCCTAGATGAACAATTAC
      F  T  D  L  M  N  N  N  F  I  N  N  Y  C  A  S  E  N  V  T  L  D  E  Q  L
1126  CTGCCTTTCGAGGAAGATTTTCTGGTGTTGTTTATATGCCCAACAAACCTACCAAGTACGGCATAAAGCATTATG
      P  A  F  R  G  R  F  S  G  V  V  Y  M  P  N  K  P  T  K  Y  G  I  K  H  Y
1201  CACTCGTAGACTCTGCAACATTTTATTTATTGAAATTTGAGATTTATGCTGGAGTACAACCTGAGGGTCCGTATC
      A  L  V  D  S  A  T  F  Y  L  L  K  F  E  I  Y  A  G  V  Q  P  E  G  P  Y
1276  GAATGCCTAATGATACAGTAAGTTTGGTAAAGCGTATGACAGAACCCATTTGGGGTACTGGTCGAAATGTTACGA
      R  M  P  N  D  T  V  S  L  V  K  R  M  T  E  P  I  W  G  T  G  R  N  V  T
1351  TGGACAACTGGTTTACTAGTGTCCCTCTTGCAAATATACTTCTAAAAGATCACCAGCTTACTATGGTTGGAACTA
      M  D  N  W  F  T  S  V  P  L  A  N  I  L  L  K  D  H  Q  L  T  M  V  G  T
1426  TTCGCAAAAATAAACCGGAAATTCCAACTTGCTTCCAGCCTAAACGAACACGAACTGAGCATTCCAGTCTGTTTG
      I  R  K  N  K  P  E  I  P  T  C  F  Q  P  K  R  T  R  T  E  H  S  S  L  F
1501  GTTTCCAAGAAGACGTGACACTTTGCTCTTACGTTCCAAAAAAATCTAAAGCTGTACTGCTGATTTCGTCAATGC
      G  F  Q  E  D  V  T  L  C  S  Y  V  P  K  K  S  K  A  V  L  L  I  S  S  M
1576  ATAATGACAATAATATAGTCGAAAGTGAAAAGAAGAAACCAGAAATTATTTTGTACTACAATAGTACGAAAGGCG
      H  N  D  N  N  I  V  E  S  E  K  K  K  P  E  I  I  L  Y  Y  N  S  T  K  G
1651  GTGTAGATACTAACGACCAAATGTGTGCTAACTATAACGTGGGCAGACGTACAAAGAGATGGCCTATGGTTATAT
      G  V  D  T  N  D  Q  M  C  A  N  Y  N  V  G  R  R  T  K  R  W  P  M  V  I
1726  TTTATCACCTTCTCAACGTTGCTGGTATCAATGCTTATGTAATTTTTAAAAATAAAATTGATCACGGAATATCAA
      F  Y  H  L  L  N  V  A  G  I  N  A  Y  V  I  F  K  N  K  I  D  H  G  I  S
1801  GAAGGGAGTTTTTAAAGCATCTTGCAGTTGATTTGGTAAAGGTTCACCAACAAACTCGATCAAATATACCACAAC
      R  R  E  F  L  K  H  L  A  V  D  L  V  K  V  H  Q  Q  T  R  S  N  I  P  Q
1876  TGCCTAGGGCCGTACAAAAACGGCTGAAAAGAAATGCTGAAGTTCAGGATCCTGGCTCCACTTCCAGAGGTGGTC
      L  P  R  A  V  Q  K  R  L  K  R  N  A  E  V  Q  D  P  G  S  T  S  R  G  G
1951  CATCGACAAGTTATAAAAGGTGCCATATCTGCCCAAGGTCAAAGGATAAAAAAATAAGATTTATGTGTGCCAAAT
      P  S  T  S  Y  K  R  C  H  I  C  P  R  S  K  D  K  K  I  R  F  M  C  A  K
2026  GCCACCATCACATCTGTCACGACCATTCCACTATGATATGTGATAAATGTATTGATTAAAAAGTTCCTACATTTA
      C  H  H  H  I  C  H  D  H  S  T  M  I  C  D  K  C  I  D  *
2101  AAATTTCCTATATTTAAAAGTTTATAATAATGTTGATTATTGATAATATTAAAATTTAGAGCAATGTTAGCTTTT
2176  TCTAATTTTCATAATAAAAATAAAAGTATGTCATTACAACTATGTTTTATTGTACCTTAAAAATTGGGGTTATAT
2251  ATACCCCACGCGAGCGTAAGTGTAACTATATGGACGCGAGTAATTAAGGGttaagttttataaacctta
```

B

```
   1  CCCTTAATTACTCGCGCGGGGTCATTCTGACCCCAGAGCGCATTTAAAAGCGTATAACTATTAGAGTTTTTAAGT
  76  TATAAACAAAACAGTCCAGGTAGCTTCCTAACTTGTCGGTACCTTTAACGTCTGTTGCCGGCAAGCAGCGCCGGA
 151  GAGTGGACTCGTCTCTGTGATAGTGCTGGGGTCGATCTGACCCCACGCGAGCATAAACGTAAGTTTTTTGAAACG
 226  ATTTCTTATTTCAAAACACATTTTTTTAGTTTTATTTAAATGTCTATGGATGTATTACTATCACCTATTAGTCTT
                                                  M  Y  Y  Y  H  L  L  V  L
 301  GTATACTGTTATTCTTCAATTCGTGTGCAATAACAATAAAATATTTGTTTTACATTATTACCAGTTCGATATTAT
        Y  T  V  I  L  Q  F  V  C  N  N  N  K  I  F  V  L  H  Y  Y  Q  F  D  I  I
 376  TACATTTATGTTCACTTTTACTTACAGACAATGAGTCGGCTCAAAGATAATGATATTCTTCAATTATTAGAAGAC
        T  F  M  F  T  F  T  Y  R  Q  *  V  G  S  K  I  M  I  F  F  N  Y  *  K  T
 451  GAGGAAAGTAACAACTCTGGCAATTATGGGTACACTTCGTAAGAACAAAAAAGAAATCCCTTCAAATTTTCTACC
        R  K  V  T  T  L  A  I  M  G  T  L  R  K  N  K  K  E  I  P  S  N  F  L  P
 526  AAACAGACAACGCCAAGAACAAAGTGCACTATTTGGCTTCAAAGAAAATATGACGCTCTGTTCATATGTGACCAA
        N  R  Q  R  Q  E  Q  S  A  L  F  G  F  K  E  N  M  T  L  C  S  Y  V  T  K
 601  AAAATCAAAGGCTGTTCTCATGCTATCAACTACGCACCACGATAATGCCGTAGTAGAATCGGCTAAGAATAAGCC
        K  S  K  A  V  L  M  L  S  T  T  H  H  D  N  A  V  V  E  S  A  K  N  K  P
 676  AGAAATTATTTTATTCTACAATAAAACCAAAGGTGGCGTCGACACCAATGATCAGATGTGTTCCACATACAACGT
        E  I  I  L  F  Y  N  K  T  K  G  G  V  D  T  N  D  Q  M  C  S  T  Y  N  V
 751  TGGTCGCAGGACAAAGCGCTGGCCTCTGGTTATATTTTATCACTTGTTCAATGTGGATACCTATATATTCTTTGA
        G  R  R  T  K  R  W  P  L  V  I  F  Y  H  L  F  N  V  D  T  #  I  Y  S  L
 826  TAGTTTATAAGCACAACACACAGTCGAAAATTATTCGCCGAGACTAGTTGAAGCTATTGGCGGTATTTTTGACCA
        I  V  Y  K  H  N  T  Q  S  K  I  I  R  R  D  *  L  K  L  L  A  V  F  L  T
 901  ACCCCTTTCGCCAAACAAGAGCGCAAATACCGCCACTACCAAGAGTTCTAAAAAGGCGCCTACAAAAAAATTCAG
        N  P  F  R  Q  T  R  A  Q  I  P  P  L  P  R  V  L  K  R  R  L  Q  K  N  S
 976  ATCTTGCAACGCCTGGTCCAAGTGGAGTTTCCAACGTTTCCAAAAAACGCTTCCGTTGTGAAATTTGCCCAAGGT
        D  L  A  T  P  G  P  S  G  V  S  N  V  S  K  K  R  F  R  C  E  I  C  P  R
1051  CTAAGGACCGAAAAACGTCGGATGTGTGCATGAAATGCAATAAACACAATTGCAGCGAACACATAAAAAAAATT
        S  K  D  R  K  T  S  D  V  C  M  K  C  N  K  H  N  C  S  E  H  I  K  K  N
1126  GTGATTATTGTTAGGAGTTTTAAGTATATTTTTGGAAGACTTTTTTTTGTTTCAGAAAAGCTATTTAGCAAAGTT
        C  D  Y  C  *
1201  TAATTATCCTGATCTCATTAATATTGTTCCTACCTAAAGTTTAATAAAAGTTGATTTTCATAAAATATTTTTTTT
1276  CTTGAATTTTCTGCCCATTTGTGTTTGGGGTCACTGTGACCCCATGCGAGTATCAATGCAAGTTTTTTAGCGCGA
1351  GTAATTAAGGG
```

C



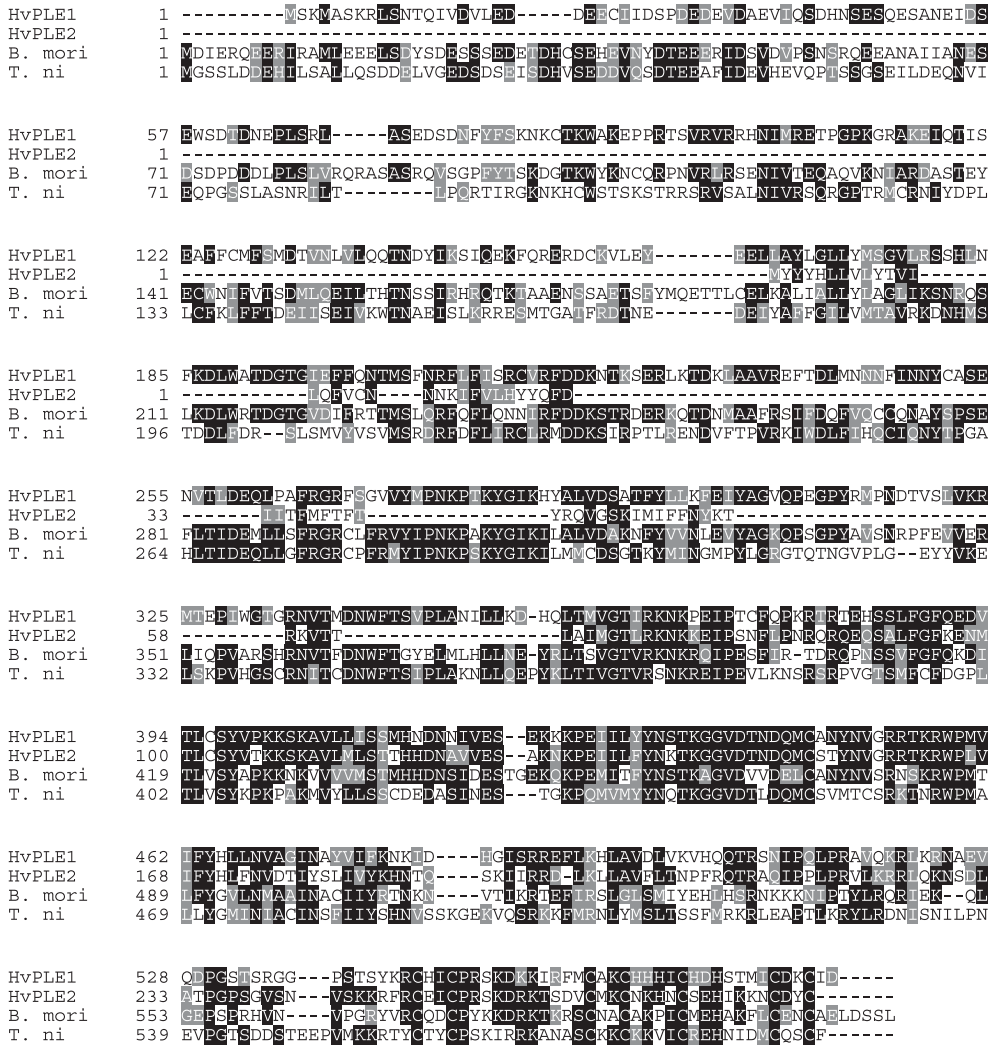
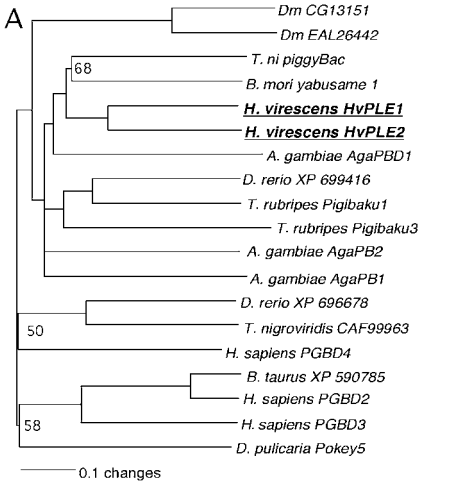


Figure 2. Multiple transposase sequence alignment of PLE-related to the *HvPLEs*. The *yabusame-1* from *B. mori* (GenBank accession number BAD11135) and the *piggyBac* from *Tricoplusia ni* (AAA87375) are aligned together with *HvPLEs*. The *HvPLE2* transposase is the reconstructed sequence based on a consensus of a total nine clones, and ignored one frameshift. Inverted boxes are identical and grey boxes are similar sequences in 50% majority.



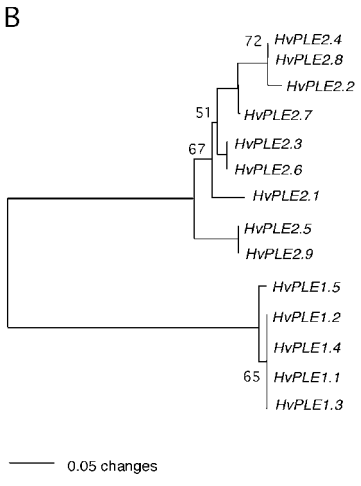


Figure 3. Phylogenetic relationship among *piggyBac-like elements* in amino acid sequence for the transposases (A) and among the nucleotide sequences for *HvPLE1* and *HvPLE2* copies. Unrooted tree were generated in PAUP 4.01b or in MEGA3 with neighbour-joining tree with 1000 bootstrapping. Only the bootstrapping value lower than 80% are written on the branch. Amino acid sequences without GenBank accession numbers are obtained from Sarkar *et al.* (2003). The *HvPLE* sequences are deposited in GenBank with accession numbers: HvPLE1.1 (DQ407726), HvPLE1.2 (DQ407727), HvPLE1.3 (DQ407728), HvPLE1.4 (DQ407729), HvPLE1.5 (DQ407730), HvPLE2.1 (DQ407731), HvPLE2.2 (DQ407732), HvPLE2.3 (DQ407733), HvPLE2.4 (DQ407734), HvPLE2.5 (DQ407735), HvPLE2.6 (DQ407736), HvPLE2.7 (DQ407737), HvPLE2.8 (DQ407738), HvPLE2.9 (DQ407739).

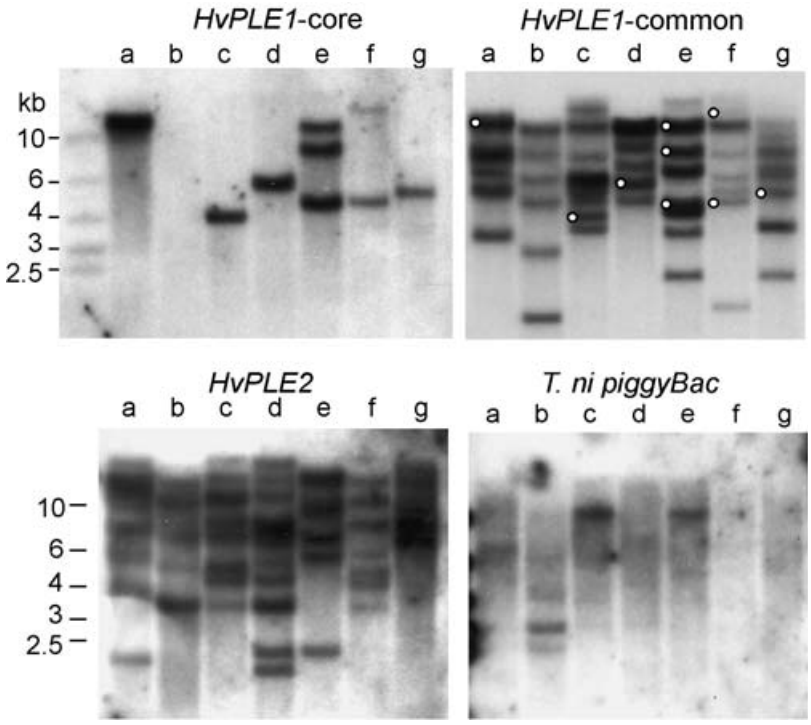


Figure 4. Genomic Southern hybridizations of seven different individuals of *H. virescens* for *HvPLE1*, *HvPLE2* and *T. ni piggyBac*. The blot for *Eco*RI digested genomic DNA was repeatedly used for hybridizations with four different probes. The common bands appeared in both blots for *HvPLE1*-core (putative full length) and for *HvPLE1*-common (putatively, including deletion copies) are marked with small empty circles. Note that the *T. ni piggyBac* probe (1.8 kb) was hybridized in low stringency, whereas all other probes were hybridized in high stringency.

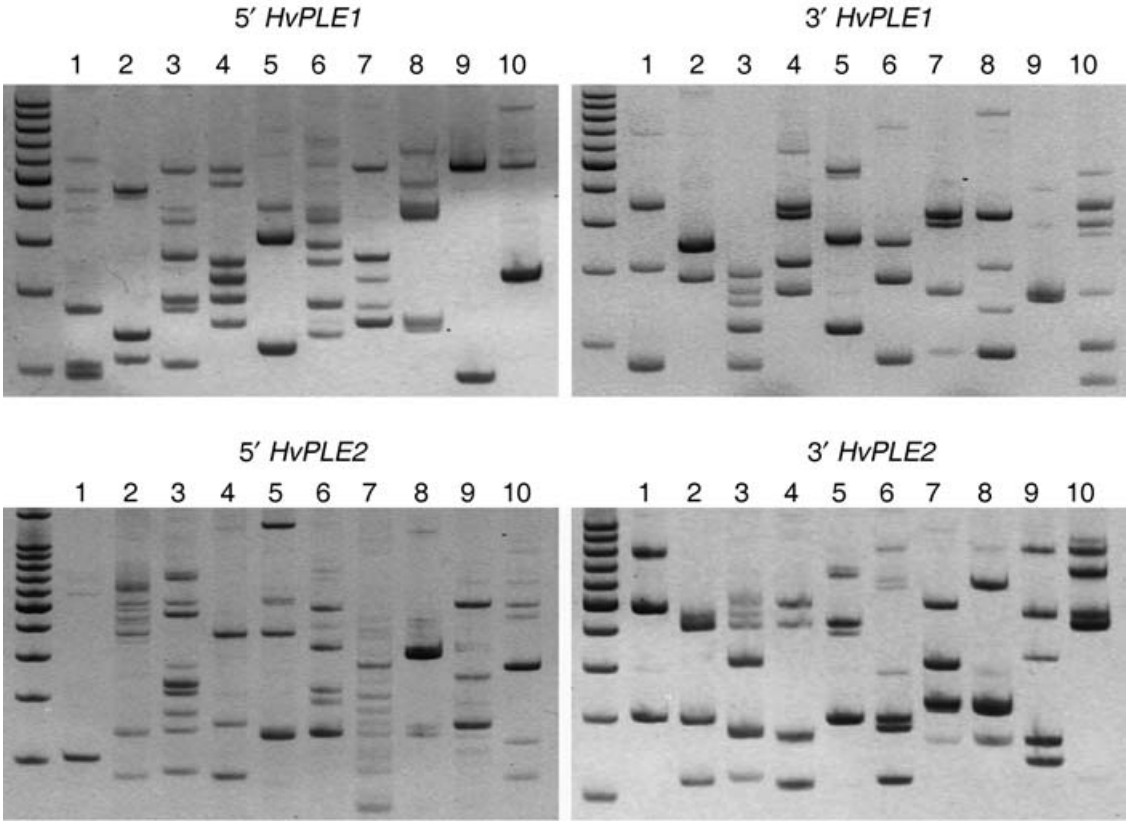


Figure 5. A representative picture showing Vectorette PCR of randomly selected 10 *H. virescens* individuals. The Vectorette PCR products for *HvPLE1* and *HvPLE2* for both 5′ and 3′ sides are displayed on 8% polyacrylamide gel. The picture colour is inverted and adjusted for maximum visibility.

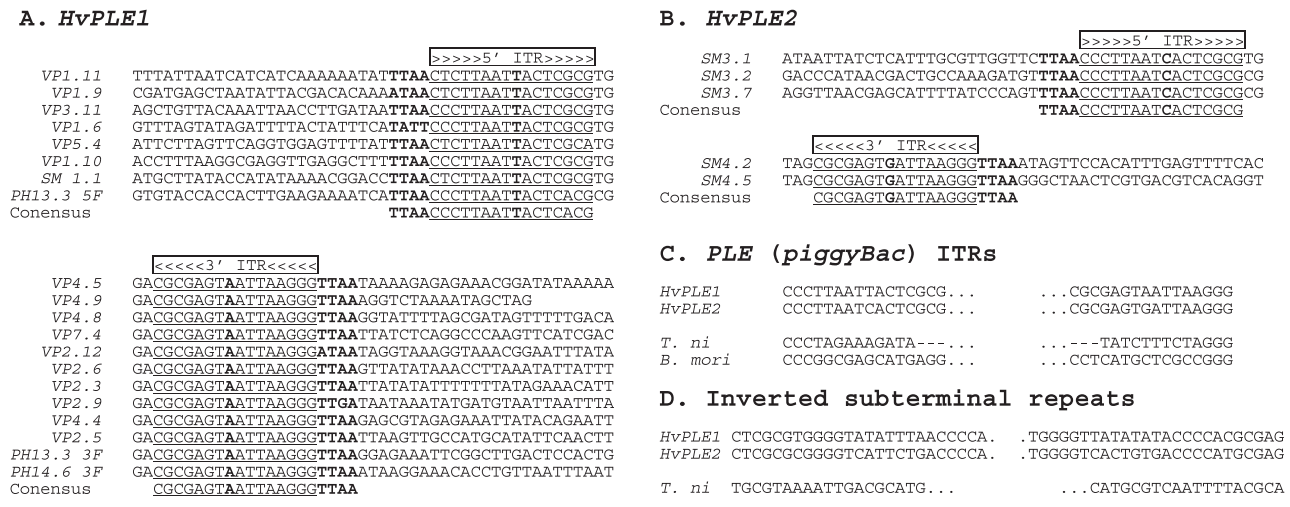


Figure 6. Nucleotide sequences of the inverted terminal repeats (ITRs) and the flanking genomic sequences in *HvPLE1* (A), *HvPLE2* (B), and the ITR sequences for *HvPLE*s, *T. ni piggyBac* and *B. mori yabusame-1* (C), and sequence comparisons of inverted subterminal repeats in *HvPLE*s and *T. ni piggyBac* (D).

regions. Most of the insertions occurred at target site TTAA, which is a characteristic of all members of TTAA-specific family DNA transposons (Fraser *et al*., 1996); some exceptions existed, such as ATAA and TATT in 5′, and ATAA and TTGA in 3′ insertion sites of the *HvPLE1* (Fig. 6A). Three 5′- and two 3′-flanking sequences obtained from the *HvPLE2* insertions were found to be strictly conserved for the TTAA target site duplications (Fig. 6B).

The 16 bp perfect ITRs are well conserved in both *HvPLE1* and *HvPLE2*, with some minor exceptions (Figs 1 and 6). A single nucleotide difference was found between the ITRs of *HvPLE1* and *HvPLE2*: T to C in 5′ ITR and A to G in 3′ ITR. It is interesting that the different bases are located in the reverse complementing positions of the inverted sequences in each type of element (Fig. 6C). A 25 bp inverted subterminal repeat was identified in both *HvPLE1* and *HvPLE2* with similar structures and sequences each other (Figs 1 and 6D). The 5′ inverted subterminal repeat partially overlaps with the 16 bp 5′ ITR, while the 3′ repeat is located in 20 bp upstream of the 16 bp 3′ ITR.

Discussion

A large diversity of PLE in eukaryotes has recently been documented in a computational analysis of genomic sequence data (Sarkar *et al*., 2003). Because the computational analysis uses the publicly available databases, knowledge on the diversity of PLEs in agriculturally important insect pests is limited due to relative paucity in the sequence information available. Therefore, we used degenerate PCR and genomic Southern hybridization to explore PLEs in tobacco budworm, an important lepidopteran pest of cotton and other crops in the United States.

Our approach, particularly the universal degenerate primers designed for the PLEs, seems to be a valuable tool to study the diversity of PLEs in other organisms. The same degenerate primers have been successfully used to isolate PLEs in the *Pectinophora gossypiella*, and *Helicoverpa zea* (Wang and Park, unpublished data), suggesting that the degenerate primers may be extended to identify PLEs in other divergent species, particularly in lepidopterans.

We identified two different PLEs in *H. virescens* with moderate to low copy numbers for each. Sequence analysis revealed that *HvPLE1* and *HvPLE2* are closely related to each other and occur in the clade with *T. ni piggyBac* in the phylogenetic analysis (Fig. 3). One of the clones for *HvPLE1*, *HvPLE1.1*, revealed an intact full-length element. We also found a low copy number fragment that cross-hybridized to the *T. ni piggyBac* probe.

Invasions of this species (or populations) by *HvPLE1* and by *HvPLE2* had likely occurred in separate events. The *HvPLE2*, with greater sequence variation is suggestive of earlier invasion of *HvPLE2* followed by *HvPLE1*. The average proportions of nucleotide differences (*p* distance) were 0.021 ± 0.005 (SE in 1000 bootstrap replications in MEGA3) among *HvPLE1* and 0.081 ± 0.02 among *HvPLE2* sequences in the region encoding transposase (Fig. 2 and Supplementary data 1). Thus, the age of *HvPLE2* is ≈ 4 times older than that of *HvPLE1*, assuming constant running of the molecular clock.

It is interesting that a striking similarity between *HvPLE1* and *HvPLE2* is found in the ITRs (15/16 nucleotides), whereas the ITRs in other closely related PLEs or *piggyBac* are more divergent (Fig. 6C). The conserved structure and sequence is extended to the subterminal repeat that may

be important in the interaction between transposase and the TE. Highly conserved ITR, the substrate of transposase, suggests a possible coevolution between *HvPLE1* and *HvPLE2* by transmobilization between these two elements.

Current status of mobility of a TE may be indicated simply by presence of the copy with intact transposase and its substrate ITR. The presence of an intact copy for *HvPLE1* was shown by cloning and sequencing and by Southern hybridization with the probe *HvPLE1*-core for the putative full-length copies (Fig. 4). The presence of an intact copy could be the consequence of purifying selection in which the intact copies are selectively favoured. Alternatively, the invasion of the TE in this species may have occurred recently, thus the intact copy could be remaining by chance, while random mutations are being accumulated including the mutation causing amino acid substitution(s) inactivating the enzyme. We were unable to test the null hypothesis of purifying selection (Kumar *et al*., 2004) on the *HvPLE1* copies due to the small number of informative sites in the sequences encoding transposase. These observations favour the scenario of recent invasion rather than purifying selection for the presence of an intact copy of *HvPLE1*, but do not rule out the possibility of purifying selection upon *HvPLE1*.

For *HvPLE2*, however, we found significant levels of purifying selection. The null hypothesis of neutral evolution was tested by using the *Z*-test that is based on the ratio between synonymous and non-synonymous change for the Nei–Gojobori model (Kumar *et al*., 2004). Pair-wise comparisons among the *HvPLE2* sequences (10 sequences, total 35 pairs) found that 29 pairs deviated significantly ($P < 0.05$) from the neutrality for purifying selection (high non-synonymous mutation rate). We conclude that the intact coding in the sequence encoding transposase has been selectively favoured in the evolution of *HvPLE2* copies, though we were not able to identify the copy with the full-length transposase similar to those of other PLEs. This is contrasted to the case of *mariner*-like element (MLE) evolution for which neutral evolution of the copies within a host species was suggested (Lampe *et al*., 2003; Witherspoon & Robertson, 2003) and as it was also found in our previous study (Wang *et al*., 2005).

Another indicator for the history of transposition of a TE may be diversity of the insertion sites. Inactive copies of TE will be fixed or lost in the population over time if they are neutral. Therefore, the insertion sites for TEs will become more homogeneous (Deceliere *et al*., 2005) within the population over time. For example, our earlier study with inactivated copies of MLE in the pink bollworm *PgMLE1* found that the insertion sites are highly homogeneous within and among populations (Wang *et al*., 2005). This result for the *PgMLE1* copies was interpreted as inheritance of ancestral insertions when the insect host invaded the New World. In contrast, we found large variation in the insertion sites in all three *HvPLEs* in Southern blotting (Fig. 4 and other data not shown, $n = 18$) and also in TE displays (Fig. 5, $n = 10$).

Three possible reasons are proposed to explain the different distribution patterns between *HvPLEs* and *PgMLE1*. First, the large insertion site variation among individuals in the *HvPLEs* could be simply due to the recent invasion of *HvPLEs* (i.e. before the homogenization has taken place in the population). Next, the different patterns may be contributed by different degrees of selection pressure on the two different TEs. A PLE is known to leave no footprint (i.e. *T. ni piggyBac*; Fraser *et al*., 1996; Thibault *et al*., 1999), whereas MLE leaves a typical excision footprint of 5 bp in the host genome (Plasterk *et al*., 1999), which may be detrimental to the host, increasing the genetic load upon mobilization, particularly if it happens in a coding region. Thus, the fitness costs of MLE are determined not only by its copy number (Charlesworth & Langley, 1989), but also by mobile events adding accumulated cost by its footprints. Therefore, selection against the mobility of PLE is considered to be lower than that in MLE in this aspect, and may have contributed to the large insertion site variation in *HvPLE*s. Lastly, another important difference between PLE in *H. virescens* and MLE in pink bollworm is the population structure of insect host. *The H. virescens* is generally known for its large effective population size (Taylor *et al*., 1996). The pink bollworm expanded its habitat recently, with the expanding cultivation of cotton and seed trade. Therefore, the *PgMLE1* may have been under a strong recent bottleneck effect, homogenizing insertion site variations, whereas *HvPLEs* have likely been under genetic drift for a long time in a large population.

Experimental methods

We amplified a fragment of *HvPLE1* by a degenerate PCR with primers based on a highly conserved region in the multiple PLE sequence alignment (Sarkar *et al*., 2003). The nucleotide sequence obtained from the partial fragment was used for cloning a full-length copy by inverse PCR, which revealed ITR. One primer anchored in the ITR was used to amplify multiple copies of *HvPLE1*. In addition to the multiple copies of *HvPLE1*, another distantly related element *HvPLE2* serendipitously was identified in the ITR PCR. Multiple copies of each *HvPLE1* and *HvPLE2*, and their flanking sequences, were obtained by Vectorette PCR and ITR PCR. The diversity in insertion sites was examined by using genomic Southern hybridization and Vectorette PCR.

Isolation of genomic DNA, cloning and sequence analysis

The *H. virescens* strain YDK was obtained from Dr Fred Gould (North Carolina State University), which was originally established from a collection of tobacco budworm eggs from three adjacent counties in North Carolina in 1988 (Gould *et al*., 1995). The samples used in this study are randomly chosen individuals in the same generation. Genomic DNA for PCR and Southern

hybridization was extracted from a single larva or adult by a phenol–chloroform extraction method as previously described (Wang *et al*., 2005). The PCR products were cloned into pGEM-T-Easy vector (Promega, Madison, WI, USA) and sequenced. The DNA sequences were compared with non-redundant databases by using the NCBI server with tblastx and tblastn (www.ncbi.nlm.gov/cgibin/BLAST). Sequence alignment was performed with CLUSTAL X (Thompson *et al*., 1994). The aligned sequences were used for construction of phylogenetic tree in PAUP 4b.2 (Swofford, 2000) or in MEGA3 (Kumar *et al*., 2004).

Polymerase chain reaction strategies

A pair of degenerate primers, PLEF266: 5′-ACNRTNGAYGAR-CARYT-3′ and PLER351: 5′-SWNGTRAAVMARTTRTC-3′, were designed in the highly conserved regions 266–271 and 346–351 in *T ni. piggyBac* transposase after multiple sequence alignment of 11 PLEs (Sarkar *et al*., 2003). In general, PCR was performed in a 25 μl reaction volume containing 20–50 ng gDNA, 0.8 μM of each degenerate primer, 0.2 mM of each dNTP, 2 mM of $MgCl_2$, and 1 μl (1 unit) of REDTaq Genomic DNA Polymerase (Sigma, St Louis, MO, USA) in buffer supplied by the manufacturer (100 mM Tris–HCl, pH 8.3, 500 mM KCl, and 0.1% gelatin). Touchdown PCR protocol was used for the degenerate PCR, which consisted of one cycle at 94 °C for 5 min, 10 cycles at 94 °C for 30 s, 50–41 °C (decreasing by 1 °C/cycle) for 30 s, and 72 °C for 1 min, followed by 25 cycles at 94 °C for 30 s, 40 °C for 30 s, 72 °C for 1 min, and a final extension at 72 °C for 10 min.

Inverse PCR was followed with primers designed in the sequence obtained from the degenerate PCR. About 2 μg genomic DNA was digested with either *Eco*RI, *Xho*I, *Pst*I, *Mlu*I or *Hin*dIII (Promega) for 10–16 h at 37 °C in a 20 μl total volume. The enzyme was inactivated at 70 °C for 20 min. The digested DNA was diluted to final 4–6 ng/μl and ligated at 4 °C for 16 h, followed by precipitation and resuspension in 20 μl of water. The following two primer pairs were used for the nested inverse PCR:

External primer pair

Forward primer HPEF: 5′-CCGTATCGAATGCCTAATGA-3′

Reverse primer HPER: 5′-CAGGTTGTACTCCAGCATAA-3′

Internal primer pair

Forward primer HPIF: 5′-TGGTAAAGCGTATGACAGAA-3′

Reverse primer HPIR: 5′-TACTTGGTAGGTTTGTTGGG-3′

The full-length PLE copy *HvPLE1.1* was successfully amplified from the same genomic DNA that was used for inverse PCR. The primers used for the *HvPLE1.1* were located in the 5′ and 3′ of flanking sequences extended to the part of ITR: HvFP: 5′-GTTGAGGCTTTTTAACCCTT-3′ and HvRP: 5′-TAAGGTTTATAA AACTTAACCC-3′. Multiple copies of *HvPLE1* or *HvPLE2* were obtained by PCR using a single primer designed on the 16 bp ITR including TTAA target site duplication (5′-TTAACCCTTAAT-TACTCGCG-3′).

Genomic Southern hybridization and Vectorette polymerase chain reaction for transposable element display

Genomic DNAs from individual TBW were digested with *Eco*RI, electrophoresed on a 0.8% agarose gel, transferred to positively charged Hybond-N^+ nylon membrane (Amersham, Piscataway, NJ, USA) by using a vacuum blotter, and fixed by ultraviolet light in a UVP HL-2000 Hybrilinker. The probes were labelled with [^{32}P]dCTP by the random priming method by using the Megaprime™. DNA labelling system (Amersham). Hybridizations were performed at 65 °C in Rapid-Hyb Buffer (Amersham), membranes were washed at high stringency conditions, and the probe with *T. ni piggyBac* was washed at low stringency (50 °C). The membrane was used for hybridization with four different probes: *HvPLE1*-core that only hybridizes to putative full-length *HvPLE1*; *HvPLE1*-common that detects the deleted copies of *HvPLE1*; *HvPLE2*; and the 1.8 kb *T. ni piggyBac* ORF region (Fig. 1).

We employed a modified-Vectorette PCR (Riley *et al*., 1990; Wang *et al*., 2005) to obtain the flanking sequence of *HvPLE1* and *HvPLE2* and to examine the diversities of insertion sites. The Vectorette construct, genomic DNA digestion/ligation and PCR were performed according to Wang *et al*. (2005) with the following primers:

PLE1.R1: 5′-CTACTAAGAACGCTATGAGG-3′ (*HvPLE1* 5′ external)
PLE1.R2: 5′-TGGAGATACAACGCTTCAAATA-3′ (5′ internal)
PLE1.F1: 5′-TGTCACGACCATTCCACTAT-3′ (3′ external)
PLE1.F2: 5′-CCATGCGAGCGTAAGTGTAA-3′ (3′ internal)
PLE2.R1: 5′-AACTCTAATAGTTCTACGCTTT-3′ (*HvPLE2* 5′ external)
PLE2.R2: 5′-TAAATGCGCTCTGGGGTCAG-3′ (5′ internal)
PLE2.F1: 5′-ATTTGCCCAAGGTCTAAGGA-3′ (3′ external)
PLE2.F2: 5′-GAATTTTCTGCCCATTTGTG-3′ (3′ internal)

Acknowledgements

We thank Dr Fred Gould for providing insects and Dr Srini Kambhampati for reviewing an earlier version of this manuscript. This paper is contribution no. 06-89-5 from the Kansas Agricultural Experiment Station, Manhattan, Kansas.

References

Cary, L.C., Goebel, M., Corsaro, B.G., Wang, H.G., Rosen, E. and Fraser, M.J. (1989) Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**: 156–169.

Charlesworth, B. and Langley, C.H. (1989) The population genetics of *Drosophila* transposable elements. *Annu Rev Genet* **23**: 251–287.

Deceliere, G., Charles, S. and Biemont, C. (2005) The dynamics of transposable elements in structured populations. *Genetics* **169**: 467–474.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y. and Xu, T. (2005) Efficient transposition of the *piggyBac* (PB) transposon in mammalian cells and mice. *Cell* **122**: 473–483.

Elick, T.A., Lobo, N. and Fraser, M.J. Jr (1997) Analysis of the cis-acting DNA elements required for *piggyBac* transposable element excision. *Mol Gen Genet* **255**: 605–610.

Fraser, M.J., Smith, G.E. and Summers, M.D. (1983) Acquisition of host cell DNA sequences by Baculoviruses: relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis virus. *J Virol* **47**: 287–300.

Fraser, M.J., Ciszczon, T., Elick, T. and Bauser, C. (1996) Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol Biol* **5**: 141–151.

Gould, F., Anderson, A., Reynolds, A., Bumgarner, L. and Moar, W.

(1995) Selection and genetic analysis of a *Heliothis virescens* (Lepidoptera: Noctuidae) strain with high levels of resistance to *Bacillus thuringiensis* toxins. *J Ecol Entomol* **88**: 1545–1559.

Handler, A.M. (2002a) Prospects for using genetic transformation for improved SIT and new biocontrol methods. *Genetica* **116**: 137–149.

Handler, A.M. (2002b) Use of the piggyBac transposon for germ-line transformation of insects. *Insect Biochem Mol Biol* **32**: 1211–1220.

Handler, A.M. and McCombs, S.D. (2000) The piggyBac transposon mediates germ-line transformation in the Oriental fruit fly and closely related elements exist in its genome. *Insect Mol Biol* **9**: 605–612.

Horn, C., Offen, N., Nystedt, S., Hacker, U. and Wimmer, E.A. (2003) piggyBac-based insertional mutagenesis and enhancer detection as a tool for functional insect genomics. *Genetics* **163**: 647–661.

Kazazian, H.H. Jr (2004) Mobile elements: drivers of genome evolution. *Science* **303**: 1626–1632.

Kumar, S., Tamura, K. and Ney, M. (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings Bioinformatics* **5**: 150–163.

Lampe, D.J., Witherspoon, D.J., Soto-Adames, F.N. and Robertson, H.M. (2003) Recent horizontal transfer of mellifera subfamily mariner transposons into insect lineages representing four different orders shows that selection acts only during horizontal transfer. *Mol Biol Evol* **20**: 554–562.

Lorenzen, M.D., Brown, S.J., Denell, R.E. and Beeman, R.W. (2002) Transgene expression from the *Tribolium castaneum* Polyubiquitin promoter. *Insect Mol Biol* **11**: 399–407.

Plasterk, R.H., Izsvak, Z. and Ivics, Z. (1999) Resident aliens: the Tc1/mariner superfamily of transposable elements. *Trends Genet* **15**: 326–332.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S. *et al.* (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res* **18**: 2887–2890.

Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**: 615–623.

Robinson, A.S., Franz, G. and Atkinson, P.W. (2004) Insect transgenesis and its potential role in agriculture and human health. *Insect Biochem Mol Biol* **34**: 113–120.

Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M. and Collins, F.H. (2003) Molecular evolutionary analysis of the widespread *piggyBac* transposon family and related 'domesticated' sequences. *Mol Genet Genomics* **270**: 173–180.

Swofford, D.L. (2000) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Taylor, M.F.J., Park, Y. and Shen, Y. (1996) Molecular population genetics of sodium channel and juvenile hormone esterase markers in relation to pyrethroid resistance in *Heliothis virescens* (Lepidoptera: Noctuidae). *Ann Entomol Soc Am* **89**: 728–738.

Thibault, S.T., Luu, H.T., Vann, N. and Miller, T.A. (1999) Precise excision and transposition of *piggyBac* in pink bollworm embryos. *Insect Mol Biol* **8**: 119–123.

Thibault, S.T., Singer, M.A., Miyazaki, W.Y., Milash, B., Dompe, N.A., Singh, C.M. *et al.* (2004) A complementary transposon tool kit for *Drosophila melanogaster* using P and *piggyBac*. *Nat Genet* **36**: 283–287.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Wang, J., Staten, R., Miller, T.A. and Park, Y. (2005) Inactivated mariner-like elements (MLE) in pink bollworm, *Pectinophora gossypiella*. *Insect Mol Biol* **14**: 547–553.

Wimmer, E.A. (2003) Innovations: applications of insect transgenesis. *Nat Rev Genet* **4**: 225–232.

Witherspoon, D.J. and Robertson, H.M. (2003) Neutral evolution of ten types of mariner transposons in the genomes of *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J Mol Evol* **56**: 751–769.

Supplementary data

Sequence alignment of *HvPLE*s in MEGA3 format.