Insect Molecular Biology (2008) **17**(4), 387–393

Highly similar *piggyBac* elements in *Bactrocera* that share a common lineage with elements in noctuid moths

A. M. Handler*, G. J. Zimowska* and K. F. Armstrong†

**Center for Medical, Agricultural, and Veterinary Entomology, Agricultural Research Service, U.S. Department of Agriculture, Gainesville, FL, USA; and* †*Bio-Protection and Ecology Division, Lincoln University, Christchurch, New Zealand*

Abstract

The *piggyBac* IFP2 transposable element, originally discovered in a *Trichoplusia ni* cell line, also exists as nearly identical elements in other noctuid lepidopterans, and in several species of the tephritid genus *Bactrocera*. To further define the distribution of *piggyBacs* in *Bactrocera*, and compare their relationship to sequences found in Lepidoptera, a survey by PCR amplification was performed in a range of *Bactrocera* species. Highly similar *piggyBac* sequences were found in all *B. dorsalis* complex species tested, as well as in species in the *B. zonata* and *B. frauenfeldi* complexes. All nucleotide sequences had > 94% identity to corresponding sequences in the *T. ni* IFP2 element, and > 88% identity among the sequences. Conserved primers did not amplify any distantly related sequences that have been found by computational searches in a wider range of insect and non-insect species. Notably, 55 nucleotide substitutions relative to IFP2 were common to all the *Bactrocera* sequences, 44 of which exist in *piggyBacs* previously sequenced from moths, with 17 resulting in amino acid substitutions. These *piggyBac* elements, that apparently traversed orders by horizontal transfer, probably arose from a lineage separate from IFP2 and the other known elements in *T. ni*. Implications for the presence of nearly identical *piggyBac*s, in widely distributed insects, to the applied use of *piggyBac* vectors are discussed.

Received 20 September 2007; accepted after revision 24 March 2008. Correspondence: A. M. Handler, Center for Medical, Agricultural, and Veterinary Entomology, Agricultural Research Service, U.S. Department of Agriculture, 1700 SW 23rd Drive, Gainesville, FL, 32608, USA. Tel.: +1 352 374 5793; fax: +1 352 374 5794; e-mail: al.handler@ars.usda.gov

Keywords: *piggyBac*, Bactrocera, *Bactrocera dorsalis* complex, transposable elements, horizontal transfer, noctuid *piggyBac*s.

Introduction

The *piggyBac* transposon (IFP2) was originally discovered by Fraser *et al*. (1983) in the *Trichoplusia ni* TN-368 cell line genome as a repetitive element that was isolated after its transposition into the AcNPV FP locus (Cary *et al*., 1989; Elick *et al*., 1995). Subsequently, the element was developed into a binary vector : helper transformation system (Handler *et al*., 1998) that has successfully transformed the germ-line of insects within four separate orders (see Handler, 2002; Handler & O'Brochta, 2005). Consistent with this wide range of function, *piggyBac*-related elements and sequences have been found in a variety of insect species and other organisms. Many of these elements are highly diverged (Sarkar *et al*., 2003), although more conserved *piggyBac*-related elements have been identified in several Lepidoptera (Wang *et al*., 2006; Xu *et al*., 2006; Sun *et al*., 2008).

In an early transformation experiment in the oriental fruit fly, *Bactrocera dorsalis sensu stricto*, Southern hybridization studies of transformed and nontransformed lines also revealed the existence of 10 or more *piggyBac*-like sequences, but these had nearly identical nucleotide sequences to the IFP2 *piggyBac* (Handler & McCombs, 2000). PCR-amplified sequences from two wild strains and the *white eye* host strain spanned 1.5 kb of the transposase coding region and were 95 to 98% identical to *piggyBac*, with 92% identity to the conceptual translation of these sequences. An initial survey found similarly conserved *piggyBac* sequences in several other *Bactrocera* species (Handler, 2002), which has recently been expanded upon (Bonizzoni *et al*., 2007). These findings gave support to the notion that *piggyBac* had undergone recent horizontal transmission between phylogenetically and geographically distinct species. This led to a further examination of *piggyBac* in Lepidoptera by hybridization and PCR analysis that showed that highly similar *piggyBac* sequences exist in several noctuid species including *Helicoverpa armigera*

and *Helicoverpa zea* (> 94% nucleotide identity) with more diverged sequences in *Spodoptera frugiperda* (77% nucleotide identity) (Zimowska & Handler, 2006). Of five independent sequences isolated from *T. ni* organismal genomic DNA, none were identical to IFP2, although all had uninterrupted reading frames. Of particular interest was the finding of three nucleotide substitutions found in all the noctuid sequences including the non-IFP2 sequences from *T. ni*, that were also evident in the *B. dorsalis s.s.* sequences.

To further define the distribution of highly similar *piggyBac* sequences among geographically dispersed *B. dorsalis* strains and *Bactrocera* species, and determine their relationship to the noctuid elements, a PCR survey for *piggyBac* was initiated. Hybridization studies previously indicated that *piggyBac* did not exist in several strains of the melon fly, *B. cucurbitae*, or the Mediterranean fruit fly, *Ceratitis capitata* (Handler & McCombs, 2000). However, *B. dorsalis s.s.* is a member of a large species complex that includes 60 or more sibling species (see White & Elson-Harris, 1992; Clarke *et al*., 2005). The use of differing or inconclusive morphological characteristics has made some of these species distinctions imprecise, and mating compatibility does exist in some cases. More recent use of molecular techniques including mitochondrial DNA sequence analysis has helped to clarify some relationships, but relatively few species within the complex have been tested (Muraji & Nakahara, 2001; Smith *et al*., 2003). For the purposes of our analysis, conserved primer sequences were used in several *B. dorsalis* complex species in the *Bactrocera* subgenus, as well as several species thought to be outside the complex. Consistent with earlier results, sequences nearly identical to the IFP2 *piggyBac* were identified in all *B. dorsalis* complex species, in addition to members of the closely related *B. zonata* and *B. frauenfeldi* complexes. Comparison of the *Bactrocera piggyBac* sequences to those in noctuids suggests that these elements share a common lineage resulting from a horizontal transfer, that is separate from the known elements in *T. ni*, including IFP2.

Results

PCR amplification of piggyBac *sequences*

Total genomic DNA extracted from the *Bactrocera* species listed in Table 1 was assayed for the presence of *piggyBac* sequences using several sets of conserved *piggyBac* primers for PCR amplification. The nucleotide positions for nondegenerate primers based on the original IFP2 *piggyBac* sequence are given in the Experimental procedures. Primer pairs were designed to generate products that corresponded to the first 2443 bp of the 2474 bp IFP2 element, including the 5′ terminus and the entire transposase coding region. PCR products were not observed with any of the three primer sets in *B. redunca*, *B. umbrosa*, *B. cucurbitae* or *Dacus porina*. Southern hybridization analysis could clarify whether more divergent *piggyBacs* exist in these species, but the limited amount of tissue available for many of these species (one fly or less) limited this analysis, except for *B. cucurbitae* that was tested previously (Handler & McCombs, 2000).

Nucleotide sequence comparisons

For purposes of initial sequence comparisons, focus was placed on nucleotide sequences generated from primer set C, 643F-2443R, spanning most of the coding region and the proximal 3′ terminal region, and the amino acid conceptual translation of these sequences. A direct comparison to IFP2 showed that nucleotide sequence pair distances of this region generated by a multiple sequence alignment (CLUSTALW) range from 94 to 97% identity, indicating a very high level of sequence similarity to the functional element (see Supplementary Material Table S1). Among the different species, most sequence identities are greater than 95%, with several greater than 99%, in particular sequences from *B. cacuminata* compared to *B. dorsalis s.s.* or *B. papayae*. The lowest identities are about 88% among sequences from *B. endiandre* (Bend-pB1C2) and *B. correcta* (Bcor-pB1C1) and *B. carambolae* (Bcar-pB1C1). Relative to IFP2 nucleotides 1–2443, there are 55 substitutions common to all the *Bactrocera* sequences (see Supplementary Material Fig. S1A and S1B). Other common indels include a 4-bp insertion at IFP2 nt 291 and a 3-bp deletion at nts 523–525 (Supplementary Material Fig. S1A), and a deletion at nt 924 (Supplementary Material Fig. S1B).

The phylogenetic relationship of the Bactrocera nucleotide sequences generated by the 643F-2443R primer set is illustrated by a maximum likelihood tree using the *T. ni* IFP2 and Tn-pBac5 sequences as an outgroup (Fig. 1). Few distinct clades are apparent, with strongest bootstrap support for a subgroup of sequences from *B. cacuminata*, *B. dorsalis s.s.*, and *B. papayae*, consistent with their sequence pair distances. Of these species, only a sequence from *B. papayae* (Bpap-pB1C2) exists in another subgroup with a *B. carambolae* sequence (Bcar-pB1C1). Beyond another subgroup including Bcar-pB1C2 and Bend-pB1C1, all other sequences from *B. albistrigata*, *B. carambolae*, *B. correcta*, *B. endiandre*, and *B. zonata* are either not grouped or subgrouped within the same species. A maximum likelihood tree of the conceptual translation of these sequences (data not shown) shows a generally consistent branch pattern.

Amino acid sequence comparisons

Consistent with the nucleotide comparisons, analysis of a multiple amino acid sequence alignment (CLUSTALW) of the conceptual translation of the *Bactrocera* sequences (corresponding to IFP2 nts 643–2110) also yielded high

Table 1. Summary of *piggyBac* sequences amplified from *Bactrocera* species

Host species	Sample origin	Sequence designation	Primer set*	% identity to IFP2†	GenBank accession nos.
***B. dorsalis* complex**					
B. cacuminata	Australia	Bcac-pB1C1	C	95.5	EU139547
		Bcac-pB1C2	C	95.5	EU139548
		Bcac-pB1A1	A	–	EU139843
B. carambolae	Surinam	Bcar-pB1C1	C	95.5	EU139549
		Bcar-pB1C2	C	94.3	EU139550
		Bcar-pB1C3	C	95.1	EU139551
		Bcar-pB1A1	A	–	EU139844
B. dorsalis	Taiwan	Bdor-pB1C1	C	95.5	EU139554
		Bdor-pB1C2	C	95.3	EU139555
		Bdor-pB1A1	A	–	EU139845
B. endiandrae	Australia	Bend-pB1C1	C	95.5	EU139556
B. minuta	Vanuatu	Bmin-pB1C2	C	94.8	EU139557
		Bmin-pB1C3	C	96.5	EU139558
B. papayae	Australia	Bpap-pB1C1	C	94.7	EU139559
		Bpap-pB1C2	C	96.1	EU139560
		Bpap-pB1C3	C	94.0	EU139561
		Bpap-pB1A1	A	–	EU139846
	Malaysia	Bpap-pB2C1	C	95.3	EU139562
		Bpap-pB2C2	C	95.3	EU139563
		Bpap-pB2C3	C	95.5	EU139564
		Bpap-pB2A1	A	–	EU139847
***B. zonata* complex**					
B. correcta	Vietnam	Bcor-pB1C1	C	91.2	EU139552
		Bcor-pB1C2	C	94.3	EU139553
B. zonata	La Reunion	Bzon-pB1C1	C	94.0	EU139565
***B. frauenfeldi* complex**					
B. albistrigata	Indonesia	Balb-pB1C1	C	96.3	EU139544
		Balb-pB1C2	C	95.5	EU139545
		Balb-pB1C3	C	95.9	EU139546
B. redunca	Vanuatu		ns		
B. umbrosa	Vanuatu		ns		
B. cucurbitae	Hawaii		ns		
Dacus porina	Australia		ns		

*primer sets: A, 1F-1293R; C, 643F-2443R; ns, no sequences generated with any primer set.
†per cent amino acid sequence identity from the conceptual translation of primer set C sequences.

levels of identity to IFP2 ranging from 91% in *B. correcta* (Bcor-pB1C1) to 97% in *B. minuta* (Bmin-pB1C3) (Table 1), with a similar range of identities of 90 to 99% among sequences from the different *Bactrocera* species (see Supplementary Material Table S1). Of the 22 *Bactrocera* sequences amplified, only the Bmin-pB1C3 sequence translation is uninterrupted, indicating that it could encode functional transposase. In addition to unique nonconserved and moderately conserved amino acid substitutions relative to IFP2, 34 of the 55 common nucleotide substitutions are within the putative transposase coding region with 17 being synonymous and 17 nonsynonymous resulting in amino acid substitutions (Table 2). An alignment of partial translations (~489 of 594 amino acids) shows 13 of the highly conserved amino acid substitutions, only one of which exists in Tni-pBac5 at position 448 (at IFP2 nt 1986, Table 2; see Supplementary Material Fig. S2). Two of the sequences do not have a substitution at one position. A three to four-amino-acid deletion is observed in seven sequences from four different species at positions 264–267, and a unique 30-amino-acid deletion at positions 360–389 exists in Bzon-pB1C1.

Common piggyBac *sequence variations in* Bactrocera *and Lepidoptera*

The analysis of *piggyBac* nucleotide sequences from several lepidopterans showed the existence of three common nucleotide substitutions from the IFP2 element present in all the sequences (including five other *T. ni* sequences), that also exist in previously identified sequences from *B. dorsalis s.s.* (Zimowska & Handler, 2006). The current analysis of *Bactrocera piggyBac* sequences relative to the first 2443 nts in IFP2, and known sequences from noctuid species, shows an additional 41 nucleotide substitutions in common with those found in all the *Bactrocera*, as well as two common indels described above. These variations are present in at least one or more noctuid *piggyBac* sequences except for those from *T. ni* (see Supplementary Material Fig. S1A and S1B; only Fig. S1A shows the *Helicoverpa* sequences). Several of these substitutions are not

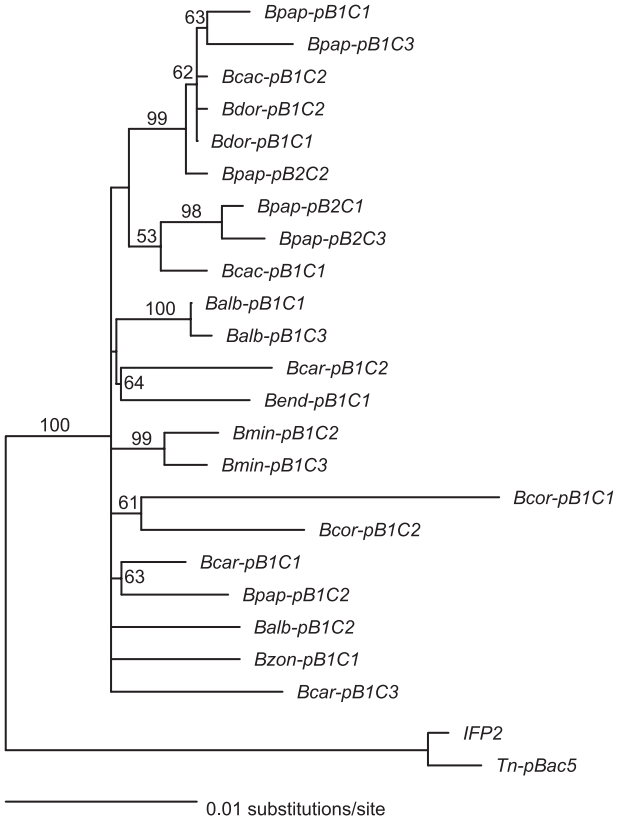

Figure 1. Phylogenetic relationships of nucleotide sequences corresponding to IFP2 nts 643–2443 amplified from the genomic DNA of indicated *Bactrocera* species and *Trichoplusia ni* using the *piggyBac* primer set 643F-2443R (see Table 1 for sequence designations and GenBank accession numbers). Phylogenetic relationships were constructed using the maximum likelihood method (–ln = 4468.3037) with the HKY + G model, placing IFP2 and Tn-pBac5 sequences as the outgroup (accession numbers given in Fig. 2). Bootstrap values (> 50%) are shown at the nodes based on 1000 replications of the data set.

observed or differ in the *S. frugiperda* sequence (Sf-pBac1), which is most highly degenerate, and some regions in the *Helicoverpa* sequences could not be compared because of large deletions in the known elements. Of the observed substitutions, 29 of the 34 found in the *Bactrocera* coding regions also exist in the noctuid moths. Four of the nucleotides were not sequenced in *Helicoverpa*, and one corresponding to IFP2 nt 2095 was unchanged. Relative to IFP2, these result in the noctuid species sharing 16 of the 17 nonsynonymous substitutions observed in the *Bactrocera* putative transposase coding region (Table 2), and 12 of the 17 *Bactrocera* silent synonymous substitutions (13 amino acid substitutions shown in Supplementary Material Fig. S2).

The phylogenetic relationship between the *Bactrocera* elements and those from noctuid moths is shown in a maximum likelihood tree for nucleotide sequences corresponding to nts 643–1293 in IFP2 (Fig. 2). As a result of

Table 2. Summary of nonsynonymous nucleotide substitutions relative to IFP2 common to all *Bactrocera* and lepidopteran sequences yielding indicated amino acid substitutions in the conceptual translated sequence

IFP2 nt	Nucleotide substitution	Amino acid substitution
338	T-C	S-P
347	G-A	D-N
354	A-G	H-R
533	G-A	V-I
857	T-C	Y-H
865	C-G	F-L
927	T-A	F-Y
1038	A-G	E-G
1099	A-T	Q-H
1634	G-A	V-I
1697	A-T	T-S
1831	A-T	K-N
1839	G-A	R-K
1847	T-A	Y-N
1963	T-A	D-E
1986	T-C	M-T
2095*	G-T	M-I

*this nucleotide is unchanged or deleted in known lepidopteran sequences.

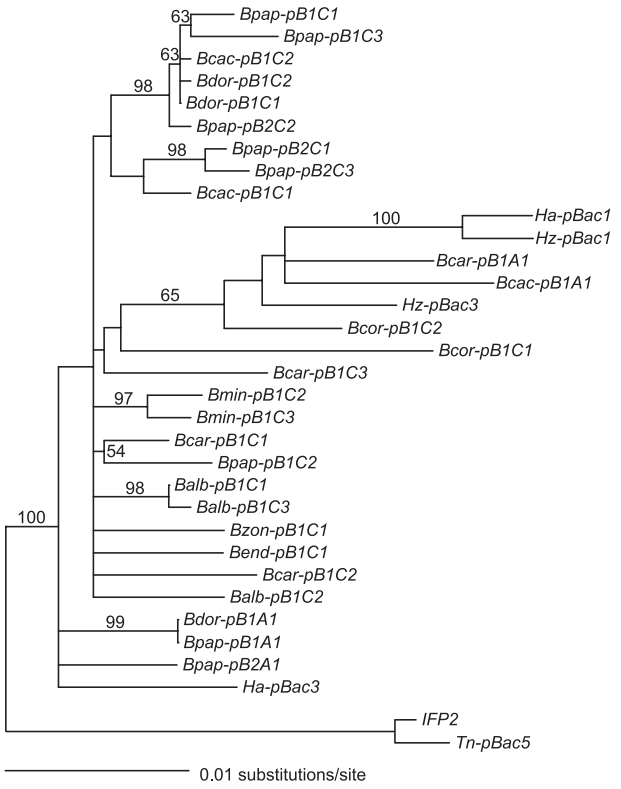

Figure 2. Phylogenetic relationships of nucleotide sequences corresponding to IFP2 nts 643–1293 amplified from the genomic DNA of indicated *Bactrocera* and noctuid species using the *piggyBac* primer sets 1F-1293R or 643F-1293R. GenBank accession numbers not shown in Table 1 include: IFP2, JO4364; Tn-pBac5, DQ236240; Ha-pBac1, DQ236220; Hz-pBac1, DQ236225; Ha-pBac3, DQ236222; Hz-pBac3, DQ236227. Phylogenetic relationships were constructed using the maximum likelihood method (–ln = 4169.80809) with the HKY + G model, placing IFP2 and Tn-pBac5 sequences as the outgroup. Bootstrap values (> 50%) are shown at the nodes based on 1000 replications of the data set.

deleted or nonrecovered sequences in the noctuids, this was the longest sequence available for comparison. Some of the *Bactrocera* relationships differ from the longer sequence analysis (Fig. 1) that is likely to be more accurate. While a differing clade structure appears, the noctuid sequences are clearly grouped within those from *Bactrocera*, which are distinct from the *T. ni* sequences. Similarly to the previous tree (Fig. 1), several of the sequence relationships are imprecise with little bootstrap support. A maximum likelihood tree for the conceptual translation of these sequences (~216 amino acids) yields a generally consistent branch pattern (data not shown).

Discussion

The discovery of *piggyBac* sequences in the tephritid fruit fly, *B. dorsalis s.s*., nearly identical to the original IFP2 *piggyBac* transposable element from the cabbage looper moth, *T. ni*, suggested that *piggyBac* might also exist in other *Bactrocera* species, and especially those included in the *B. dorsalis* complex (Handler & McCombs, 2000). This was first verified in preliminary studies (Handler, 2002), and by a recent report of highly similar *piggyBac* sequences in *Bactrocera* species from Asia and Australia (Bonizzoni *et al*., 2007). Here we expand upon the analysis of *piggyBac* elements in the *Bactrocera* subgenus by describing nearly identical sequences in six *B. dorsalis* complex species, two species in the closely related *B. zonata* complex, and one species from the *B. frauenfeldi* complex. Consistent with the initial analysis of *B. dorsalis s.s.* Hawaiian strains (Handler & McCombs, 2000), all the *piggyBac* sequences were highly similar to IFP2 and to one another. PCR products were not generated with three primer sets in the *Bactrocera* subgenus species *B. redunca* and *B. umbrosa*, or the more distantly related species *B. cucurbitae* and *Dacus porina* that reside in a separate subgenus and genus, respectively. While negative PCR results do not preclude the existence of slightly more degenerate sequences, Southern hybridization data more strongly argue against the existence of *piggyBac* in *B. cucurbitae* (Handler & McCombs, 2000). These data provide additional, and more conclusive, evidence for the horizontal transmission of *piggyBac* elements between Lepidoptera and bactrocerids, with high identity levels suggesting a relatively recent transfer.

The establishment of these elements in the *B. dorsalis* complex and closely related species probably resulted from vertical inheritance and expansion from a common ancestor; however, the precise relationship among some of the *B. dorsalis* complex species is unclear and understanding *piggyBac* distribution based on the current phylogeny is not straightforward (Muraji & Nakahara, 2001; Smith *et al*., 2003). Several of the complex species are considered to be conspecific and recent intermating may be contributing to spread, however the presence of *piggyBac* in the *Bactrocera* subgenus appears to be discontinuous. The three primer sets did not detect *piggyBac* in the *Bactrocera* subgenus species *B. redunca* and *B. umbrosa*, yet both molecular and morphometric phylogenies place *B. umbrosa* in between the *B. papayae* and *B. albistrigata* taxa that harbour *piggyBac* (Smith *et al*., 2003). We also found *piggyBac* in the more distantly related species *B. correcta,* and it was found previously in *B. jarvisi,* which is thought to be in another subgenus, *Afrodacus* (Bonizzoni *et al*., 2007).

None of the sequenced elements are identical to the functional IFP2 element, but the most similar element from *B. minuta* (Bmin-pB1C3) has the potential to encode a functional transposase. A previously sequenced *piggyBac* element from the *B. dorsalis white eye* strain also has an intact reading frame, but a 6-bp deletion, removing two amino acids, makes its potential function more suspect (Handler & McCombs, 2000). While none of the other sequences yield uninterrupted reading frames, all are highly similar to IFP2 with less overall degeneracy than any of the elements discovered thus far in the other lepidopteran species, especially in terms of deleted sequences (Zimowska & Handler, 2006). It was noted in this earlier study that three common nucleotide substitutions exist in all the noctuid sequences, as well as those from *B. dorsalis s.s*. Two are in the putative transposase coding region, one of which results in an amino acid substitution that does not eliminate transposase function (G. Zimowska & A. Handler, unpublished data). The current analysis confirms these common variations for all the *Bactrocera* sequences, but notably, 41 additional substitutions common to the *Bactrocera* and moth sequences also exist. While 17 are nonsynonymous resulting in amino acid substitutions, none result in a termination codon. If the variant transposase remains functional, this would support the notion that these elements co-evolved with IFP2 as a separate lineage.

Similar to other transposon families, more diverged *piggyBac*-like elements, having approximately 60% or less similarity in the coding region, have recently been discovered in the tobacco budworm, *Heliothis virescens* (Wang *et al*., 2006), the cotton bollworm, *Helicoverpa armigera* (Sun *et al*., 2008) and the silkworm, *Bombyx mori* (Xu *et al*., 2006). In addition, more highly diverged *piggyBac*-related sequences also exist in a wide range of organisms (Sarkar *et al*., 2003). While horizontal transfer probably contributed to the distribution of these elements, their distant relationships to IFP2 and each other are consistent with more recent vertical inheritance of stable integrants. The nearly identical *piggyBac* elements found in *Bactrocera* and several noctuid lepidopterans can be most simply explained by a relatively recent horizontal transfer and expansion among related species within each order. Whether this element originated in lepidopterans or other organisms cannot be determined as yet, but the greater degeneracy of the known noctuid

elements would suggest that they existed previous to the transfer to *Bactrocera*, and that the known elements in *T. ni* had an independent and more recent origin.

The effect of these elements on the use of *piggyBac* transformation vectors in these economically important species is uncertain. Functional elements have the ability to re-mobilize non-autonomous vectors, but the previously created *piggyBac* transformants in *B. dorsalis s.s.* remained stable for 15 or more generations when they were last tested (A. Handler & D. Oishi, unpublished data; Handler, 2002). If initial integrations are possible, the ability to stabilize transposon insertions post-integration minimizes, if not eliminates, the concern of re-mobilization (Handler *et al*., 2004; Dafa'alla *et al*., 2006). Another concern for host strains having only defective elements is the genomic effects of their potential remobilization by introduction of functional helper transposase during transformation. Possible repression of *piggyBac*-mediated transformation in *Bactrocera* hosts is also possible because of mechanisms similar to P cytotype repression in *Drosophila* (Josse *et al*., 2007). This did not, apparently, hinder the ability to transform *B. dorsalis s.s.*, although the frequencies were relatively low (Handler & McCombs, 2000). Whether other *Bactrocera* species or strains have cellular environments that can promote or repress *piggyBac* mobility remains to be tested, but these are important concerns as *piggyBac*-mediated transformants are considered for applied use, and possibly field release.

Experimental procedures

Insects

The *Bactrocera* species and strains tested are listed in Table 1 and except for *B. dorsalis s.s.* and *B. papayae* lab colonies, all samples were field-collected. The countries or regions for sample collection are given in Table 1.

PCR analysis

Direct PCR amplification was performed on genomic DNA samples prepared with DNAzol (Molecular Research Center, Cincinnati, OH, USA) and nontemplate controls using the Expand High Fidelity PCR System (Roche, Indianapolis, IN, USA) with proofreading capability under the following cycling conditions: 94 °C for 2 min, 94 °C for 30 s, 60 °C for 30 s, 68 °C for 2 min for 35 cycles with a final extension at 72 °C for 10 min. Rigorous protocols were used to avoid or detect potential sample contamination. Amplified DNA was separated and visualized on 0.8% agarose gels. Isolated products were subcloned into TOPO TA cloning vectors (Invitrogen, Carlsbad, CA, USA) and sequenced using M13 forward and reverse primers, and internal *piggyBac* primers, with most PCR products sequenced on both strands. Forward (F) and reverse (R) primers used for amplification of *piggyBac* sequences are given below. Numbers refer to the 5′ nucleotide position corresponding to the IFP2 sequence (Cary *et al*., 1989). All species were tested with primer sets A (1F-1293R), B (643F-1293R) and C (643F-2443R). Sequences generated by these primer sets had the terminal primer sequences deleted prior to analysis.

1F: 5′-CCCTAGAAAGATAGTCTGCG-3′
643F: 5′-GCGTAGCCGAGTCTCTG-3′
1293R: 5′-ACTCCGTTGGTCTGTGTTCC-3′
2443R: 5′-CGTTAAAGATAATCATGCGTAAAATTGAC-3′

Sequence analysis and comparisons

Nucleotide and amino acid sequence analysis and comparisons were performed using MegAlign (DNAStar, Inc., Madison, WI, USA), GeneWorks 2.5 (Oxford Molecular Group, Oxford, UK) software and BLASTP (Altschul *et al*., 1997), with alignments and sequence pair distances determined from a CLUSTALW multiple sequence alignment (Higgins *et al*., 1994). Amino acid sequences were aligned by pairwise alignment to IFP2 with manual introduction of frameshifts and some gaps for comparison of the consensus transposase sequence. Phylogenetic analyses were performed with the PAUP (version 4.0b10) program package (Swofford, 2002) with HKY + G selected as the substitution model to generate the best maximum-likelihood tree for the datasets. Tree reliability was assessed by 1000 bootstrap replicates.

Acknowledgements

Grateful appreciation is extended to Dr Fahong Yu and the University of Florida Interdisciplinary Center for Biotechnology Research for assistance with phylogenetic analyses, Dr Keng-Hong Tan for providing some of the *Bactrocera* samples, and to anonymous reviewers who provided insightful comments. Funding from the USDA-CSREES-National Research Initiative is gratefully acknowledged.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Bonizzoni, M., Gomulski, L.M., Malacrida, A.R., Capy, P. and Gasperi, G. (2007) Highly similar *piggyBac* transposase-like sequences from various *Bactrocera* (Diptera, Tephritidae) species. *Insect Mol Biol* **16**: 645–650.

Cary, L.C., Goebel, M., Corsaro, H.H., Wang, H.H., 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* **161**, 8–17.

Clarke, A.R., Armstrong, K.F., Carmichael, A.E., Milne, J.R., Raghu, S., Roderick, G.K. *et al*. (2005) Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the *Bactrocera dorsalis* complex of fruit flies. *Annu Rev Entomol* **50**: 293–319.

Dafa'alla, T.H., Condon, G.C., Condon, K.C., Phillips, C.E., Morrison, N.I., Jin, L. *et al*. (2006) Transposon-free insertions for insect genetic engineering. *Nat Biotechnol* **24**: 820–821.

Elick, T.A., Bauser, C.A., Principe, N.M. and Fraser, M.J. (1995) PCR analysis of insertion site specificity, transcription, and structural uniformity of the lepidopteran transposable element IFP2 in the TN-368 cell genome. *Genetica* **97**: 127–139.

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 viruses. *J Virol* **47**: 287–300.

Handler, A.M. (2002) 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.

Handler, A.M. and O'Brochta, D.A. (2005) Transposable elements for insect transformation. In *Comprehensive Insect Physiology, Biochemistry, Pharmacology, and Molecular Biology* (Gilbert, L.I., Iatrou, K. and Gill, S., eds), pp. 437–474. Elsevier Limited, Oxford.

Handler, A.M., McCombs, S.D., Fraser, M.J. and Saul, S.H. (1998) The lepidopteran transposon vector, *piggyBac*, mediates germline transformation in the Mediterranean fruitfly. *Proc Natl Acad Sci USA* **95**: 7520–7525.

Handler, A.M., Zimowska, G.J. and Horn, C. (2004) Post-integration stabilization of a transposon vector by terminal sequence deletion in *Drosophila melanogaster*. *Nat Biotechnol* **22**: 1150–1154.

Higgins, D., Thompson, J., Gibson, T., 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.

Josse, T., Teysset, L., Todeschini, A-L., Sidor, C.M., Anxolabehere, D. and Ronserreray, S. (2007) Telomeric *trans*-silencing: an epigenetic repression combining RNA silencing and heterochromatin formation. *PloS Genetics* **3**: e158

Muraji, M. and Nakahara, S. (2001) Phylogenetic relationships among fruit flies, *Bactrocera* (Diptera, Tephritidae), based on the mitochondrial rDNA sequences. *Insect Mol Biol* **10**: 549–559.

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

Smith, P.T., Kambhampati, S. and Armstrong, K.A. (2003) Phylogenetic relationships among Bactrocera species (Diptera: Tephritidae) inferred from mitochondrial DNA sequences. *Mol Phylogenet Evol* **26**: 8–17.

Sun, Z.C., Wu, M, Miller, T.A. and Han, Z.J. (2008) *piggyBac*-like elements in cotton bollworm, *Helicoverpa armigera* (Hübner). *Insect Mol Biol* **17**: 9–18.

Swofford, D.L. (2002) *PAUP*. Phylogenetic analysis using parsimony (* other methods)*. Version 4. Sinauer Associates, Sunderland, Maryland.

Wang, J., Ren, X., Miller, T.A. and Park, Y. (2006) piggyBac-like elements in the tobacco budworm, *Heliothis virescens* (Fabricius). *Insect Mol Biol* **15**: 435–443.

White, I.M. and Elson-Harris, M.M. (1992) *Fruit Flies of Economic Significance: Their Identification and Bionomics*. CAB International, Oxfordshire.

Xu, H.F., Xia, Q.Y., Liu, C., Cheng, T.C., Zhao, P., Duan, J. *et al*. (2006) Identification and characterization of piggyBac-like elements in the genome of domesticated silkworm, *Bombyx mori*. *Mol Genet Gen* **276**: 31–40.

Zimowska, G.J. and Handler, A.M. (2006) Highly conserved *piggyBac* elements in noctuid species of Lepidoptera. *Insect Biochem Mol Biol* **36**: 421–428.

Supplementary material

The following supplementary material is available for this article:

Figure S1. Multiple sequence alignments using CLUSTALW of common *piggyBac* nucleotide sequences from *Bactrocera* and noctuid species (Zimowska & Handler, 2006) generated from PCR using primers 1F-1293R (A) or 643F-2443R (B). For the latter sequences (B), only nucleotides corresponding to IFP2 positions 1294–2443 were aligned (indicated as nts 1–1150). GENBANK accession numbers not given in Table 1 are: IFP2, JO4364; Tn-pBac5, DQ236240; Ha-pBac3, DQ236222; Hz-pBac3, DQ236227; and Sf-pBac1, DQ236230. Nucleotide substitutions, additions or deletions differing from IFP2 are shown in black boxes, with deletions indicated by hyphens. Nucleotide positions for IFP2 are indicated above the sequences, with numbering for each sequence given at the right end of each sequence.

Figure S2. Amino acid sequence alignment of the conceptual translation of selected *piggyBac* nucleotide sequences from *Trichoplusia ni* and *Bactrocera* species generated from primers 643F–2443R (corresponding to aa 106–594 in IFP2), with introduced frameshifts to maintain a consensus sequence. Amino acid substitutions, additions or deletions differing from IFP2 are shown in black boxes, with deletions indicated by hyphens. Amino acid positions for IFP2 are indicated above the sequences, with numbering for each sequence given at the right end of each sequence.

Table S1. Sequence pair distance table of *piggyBac* nucleotide sequences (top) and their conceptual amino acid translation (bottom) aligned by CLUSTALW

This material is available as part of the online article from:
http://www.blackwell-synergy.com
doi:10.1111/j.1365-2583.2008.00813.x
(This link will take you to the article abstract).

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.