Insect Molecular Biology (1997) **6**(1), 83–88

Rapid cloning of insect transposon insertion junctions using 'universal' PCR

R. W. Beeman and D. M. Stauth

U.S. Grain Marketing Research Laboratory, USDA, ARS, Manhattan, Kansas, USA

Abstract

Very highly degenerate primers with short specific 3′ anchor sequences and 5′ adaptors were used in conjunction with nested specific primers to amplify large numbers of unknown insertion junctions of the insect retrotransposon *Woot,* using genomic DNA as template for the polymerase chain reaction (PCR). This technique, sometimes referred to as universal PCR, is a powerful method for molecular characterization of transposon insertions into genomes, and more generally for short-distance chromosome walking through unknown DNA. Twenty-four unique insertion junctions were cloned and sequenced from two strains of *Tribolium castaneum* and one strain of *T. freemani*. Inspection of these sequences revealed that integration of the *Woot* retrotransposon is cued by the insertion target motif, GTAC, in both species.

Keywords: one-sided PCR, *gypsy*-class retrotransposon, *Tribolium castaneum*.

Introduction

Gene transfer systems are being developed for a growing number of economically important invertebrate species (e.g. Tamura *et al.*, 1990; Handler & O'Brochta, 1991; Presnail & Hoy, 1992, Carlson *et al.*, 1995). Methods used to obtain evidence of chromosomal integration of foreign DNA include amplification of internal vector sequences from host DNA by the polymerase chain reaction (PCR), *in situ* hybridization of vector sequences to polytene chromosomes, Southern hybridization analysis of host/vector insertion junctions, and inverse PCR or genomic library screening to molecularly characterize such junctions. Here we describe the adaptation of a simple PCR-based method for rapid cloning of large numbers of unknown insertion junctions of an insect retrotransposon from host genomic DNA. Various forms of the general method, termed 'universal' or 'anchored' PCR, have been described previously (Sarkar *et al.*, 1993; Rosenthal *et al.*, 1994).

Conventional PCR requires knowledge of the sequence at both ends of a segment to be amplified. More recent one-sided PCR approaches enable amplification of DNA fragments bounded on only one side by known sequence, making it possible to perform short-distance chromosome walks by PCR. Most published techniques for one-sided PCR involve enzymatic cleavage of genomic DNA followed by the addition of artificial primer binding regions to the ends of the restriction fragments. These artificial regions are either oligonucleotide cassettes, enzymatically ligated onto the ends of the genomic restriction fragments, or oligonucleotide tails added by enzymatic synthesis (Jones & Winisforter, 1992; Lagerstrom *et al.*, 1991; Arnold & Hodgson, 1991; Ochman *et al.*, 1993). In the 'Inverse PCR' technique, the restriction fragments are self-ligated to produce DNA circles. Specific primers within the region of known sequence are oriented in directions opposite to those of traditional PCR, so that the unknown region of the DNA circle will be amplified (Triglia *et al.*, 1988; Ochman *et al.*, 1988).

All the above methods for one-sided PCR suffer from the same disadvantage: the need for enzymatic digestion and modification of template DNA. Methods that involve direct PCR using unmodified genomic DNA as template have also been described. Parker *et al.* (1991) used a single gene-specific primer in conjunction with each of a series of random 'walking' primers. Although the walking primers were nondegenerate, misannealing was frequent enough that they could function as nonspecific primers. Sarkar *et al.* (1993) used nested gene-specific primers in conjunction with highly degenerate walking primers with short specific anchor sequences at their 3′ ends. The 'universal' PCR procedure described below is based on the Sarkar method.

Woot is an active *gypsy*-class retrotransposon found in *Tribolium castaneum* and several related *Tribolium*

Received 8 April 1996; accepted 20 June 1996. Correspondence: Dr R. W. Beeman, U.S. Grain Marketing Research Laboratory, USDA, ARS, 1515 College Avenue, Manhattan, KS 66502, USA.

© 1997 Blackwell Science Ltd

species (Beeman *et al.*, 1996a). Most strains of *T. castaneum* appear to harbour about twenty to thirty copies per genome (Beeman *et al.*, 1996a). The aim of this research was to clone most or all *Woot* insertion junctions from two unrelated strains of *T. castaneum* in order to gain a better understanding of retrotransposon mobility in *Tribolium*.

Results and Discussion

To date, eighteen unique insertion junctions have been cloned and sequenced from two strains of *T. castaneum,* including one previously known and seventeen found in the current work (Fig. 1). Thirteen insertion junctions were found in both strains, whereas five have so far been found in the A^4/Ey strain but not in *Tiw-1*. We did conventional PCR using specific primers designed to amplify individual strain-specific insertion junctions. For each of the five strain-specific insertions, specific-junction PCR amplified the expected fragment from the strain known to carry the insertion, but not from the apparent empty strain.

Figure 1 shows a manual-alignment of sequences of the 3′ insertion junctions representing the eighteen unique *Woot* insertion junctions identified to date from *T. castaneum* strains A^4/Ey and *Tiw-1*, as well as six *Woot* junctions from the related species, *T. freemani*. Inspection of these sequences reveals several features: firstly, *Woot* shows a strong preference for insertion directly adjacent to the four-base target motif GTAC, and an absolute requirement for RTAY, where R = purine and Y = pyrimidine. This target sequence is known to have undergone direct duplication at either end of the *Woot* retrotransposon upon insertion in the homeotic *Abdominal* mutation (Beeman *et al.*, 1996a). Such direct duplication of the target sequence is a typical consequence of transposon insertion. G occurs directly adjacent to the downstream LTR in nineteen out of twenty-one insertions in the two species, whereas the next two bases are invariably TA. C occurs at position 4 in nineteen of the twenty-one junctions. Secondly, Fig. 1 provides a glimpse at species differences in *Woot* sequence between *T. castaneum* and *T. freemani*. *Woot* elements from *T. castaneum* terminate in ACA, whereas those from *T. freemani* usually terminate in ATA. Furthermore, all *T. freemani* elements have three small deletions near the 3′ end of the 3′ LTR. Other species differences in *Woot* LTR sequence are indicated in the Fig. 1, and consist of base substitutions. Thirdly, flanking sequences beyond the four-base target sequence show no consistent properties or motifs among the twenty-four insertion junctions shown. The insertion specificity is apparently conferred only by the four-base target, and does not depend on adjacent sequences.

Since the *Woot* LTR primers can anneal in either the upstream or downstream LTR, nontarget internal fragments can also be amplified. Figure 2 shows the sequences of four internal fragments derived by universal PCR, including an unusual arrangement of two *Woot* LTR fragments in tandem (head-to-tail) orientation. This clone, designated A15(9) in Fig. 2, could represent either a *Woot* insertion into another *Woot* LTR (although the RTAY target sequence is absent), or an internal *Woot* deletion that starts 7 nt downstream of the 5′ LTR and extends through nt 1764 in the 3′ LTR (Beeman *et al.*, 1996a).

PCR screening of several unrelated strains of *T. castaneum* using junction-specific primer pairs revealed a considerable degree of insertion site polymorphism (data not shown). We cloned and sequenced both empty and occupied variants of two of these insertion sites, A2 and A3 (refers to Fig. 1 nomenclature), each of which contained an inserted *Woot* element in the A^4/Ey strain, but not in the *Tiw-1* strain. For both occupied variants, we cloned and sequenced both the upstream and downstream junctions from the A^4/Ey strain. After sequence analysis of the empty insertion sites derived from the *Tiw-1* strain, specific PCR primers were designed that spanned the insertion sites. Using these primer pairs with A^4/Ey DNA as as template, we found that empty sites could be amplified, cloned and sequenced from the A^4/Ey strain. Thus, the A^4/Ey strain is dimorphic for the presence of *Woot* insertions at both the A2 and A3 sites. Since all ten individuals from the A^4/Ey strain tested by PCR carried both empty and occupied junctions at the A3 site (data not shown), it is possible either that most individuals in this strain are somatic mosaics for the presence of this

Figure 1. Nucleotide sequences of twenty-four unique 3′ insertion junctions of *Woot* in *Tribolium* genomic DNA. The left-hand column lists names of individual insertion junction clones. A and T = A^4/Ey and *Tiw-1* strains, respectively, of *T. castaneum*; F = *T. freemani*. The number following the strain abbreviation designates individual junctions within each strain. The number in parentheses indicates the universal primer used to amplify a particular junction (refer to Table 1). Aorig corresponds to the insertion junction of the original *Woot* element (Beeman *et al.*, 1996a). Insertion sites Aorig and A1–A4 (in boldface type) are apparently unoccupied in the *Tiw-1* strain. Numbering of nucleotides was arbitrarily begun 147 residues upstream of the LTR terminus. Junction fragments include the 3′-terminal 147 nt of *Woot*, followed by flanking host sequence. Sequences were aligned manually, since computer-assisted alignments were unsatisfactory. Dashes indicate gaps introduced to preserve the alignment. Nucleotides in boldface type indicate diagnostic *Woot* sequence differences between *T. castaneum* and *T. freemani*. The GTAC motif boxed at nt 148–151 represents the insertion target directly flanking the 3′ terminus of the 3′ LTR. GenBank accession numbers for the twenty-four unique junction fragments are U47893–U47916.

```
          1         *         *         *         *         *         *        70
Aorig     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCAAAAAAAGAAAACAAAAGTAAAATTCAAG-TT
A1(1)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCAAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A2(2)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A3(2)     TGTAGGGTATCAATTAACTAGACGAGCCAAACCATTT-AAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A4(3)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A5(5)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A6(1)            TCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAAACAAAAGTAAAATTCAAGATT
A7(8)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCAAAAAAGGAAAACAAAAGTAAAATTCAAGATT
A8(8)         GCGTATCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAAACAAAAGTAAAATTCAAGATT
T1(2)     TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTC-AAAAAAGAAAACAAAAGTAAAATTCAAGATT
A9(5)     TGTAGCGGATCAATTAACTAGACGAGCCAAACCATTTCGAAAAAAGAAA-GAAAAGTAAAATTCAAGATT
A10(2)    TGTAGCGGATCAATTAACTAGACGAGCCAAATCATTTAAAAAAAAACA--CATAAGTAAAATTCAAAATT
A11(3)    TGTAGCGGATCAATTAACTAGACGAGCCAAATCATTTCAAAAAAAAA---CAAAAGTAAAATTCAAGATC
A12(2)    TGTAGCGGATCAATTAACTAGACGAGCCAAATCATTTCGAATAAAA----CAAAAGTAAAATTCAAGCTT
A13(2)    TGTAGCGGATCAATTAACTAGACGAACCAAATCATTTCAAAAAAA-----CAAAAGTAAAATTCAAGATT
T2(1)                TTAACTAGACGAGCCAAATCATTTCAAAAAAA-----CAAAAGTAAAATTCAAGNTT
T3(1)          GGATCAATTAACTAGACGAGCCAAATCATTTCGAAAAAA-----CAAAAGTAAAATTTAAGATT
T4(2)     TGTAGCGGATCAATTAACTAGACGAACCAAATCATTTCAAAAAAA-----CAAAAGTAAAATTCAAGATT
F1(1)      TTAGAGACTCAATTAACTAGACGAGCCAAATCATTTCAAAAAT------CAAAAGCAAAATTCAAGATT
F2(1)         GAGACTCAATTAACTAGACGAGCAAAATCAATTCGAAAAT------CAAAAACAAAATTCAAGATT
F3(1)     TTTAGAGACTCAATTAACTAGACGAGCCAAATCATTTCGAAAAT------CAAAAACAAAATTCAAGATT
F4(7)     TTTAGAGACTCAATTAACTAGACGAGCCAAATCATTTCGAAAAT------CAAAAACAAAATTCAAGATT
F5(5)     TTTAGAGACTCAATTAACTAGACGAGCCAAATCATTTCGAAAAT------CAAAAACAAAATTCAAGATT
F6(3)     TTTAGAGACTCAATTAACTAGACGAGCCAAATTATTTCGAGAAT------CAAAAGCAAAATTCAAGATT

                  *         *         *         *         *         *       140
Aorig     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCGCAGTCT
A1(1)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACAGTCT
A2(2)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A3(2)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A4(3)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGNTCAACCATCATCAATCACGGTCT
A5(5)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A6(1)     AATTAAATAAATCAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A7(8)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCG-TCAACCATCATCAATCACAGTCT
A8(8)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCA-CCATCATCAATCACGGTCT
T1(2)     AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCG-TCAACCATCATCAATCACGGTCT
A9(5)     AATTAAATAAAACAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A10(2)    AATTAAATAAA----TTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A11(3)    AATTAAATAAATAAATTAAA-TAAGTTGTAGATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A12(2)    AATTAAATAAATAAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
A13(2)    AATTAAATAAATAAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
T2(1)     AATTAAATAAATAAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
T3(1)     AATTAAATAAATAAATTAAA-TAAGTTGTAACTCATTTAAACATCGATCAACCATCATCAATCACGGTCT
T4(2)     AATTAAATAAATAAATTAAA-TAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACGGTCT
F1(1)     AATTAAATAAATAAATT----------GTAAATAAGTTAAACATCAATCAACCTTT---AATCACGGTCT
F2(1)     AATTAAATAAATAAATT----------GTAAATAAGTTAAACATCAATCAACCTTT---AATCACGGCCT
F3(1)     AATTAAATAAGTAAATT----------GTAAATAAGTTAAACATCAATCAACCTTT---AATCACGGTCT
F4(7)     AATTAAATAAATAAATT----------GTAAATAAGTTAAACATCAACCAACCTTT---AATCACGGTCT
F5(5)     AATTAAATAAATAAATT----------GTAAATAAGTTAAACATCAATCAACCTTT---AATCACGGTCT
F6(3)     AATTAAATAAATAAATT----------GTAAATAAGTTAAA--TCAATCAACCTTT---AATCACGGTCT

                  *         *         *         *         *         *       210
Aorig     CACTACAGTACTTAAAGAAAAGTAGATGTGCCTGAA
A1(1)     CACTACAGTACTATTATTTATTTAAATTGTATTATAAGCATCACGCTACAGCCAAAAAATGCTTTTTAAA
A2(2)     CACTACAGTACTGATTCTTTTATTGATTTTGGAAACAGTGACCGCTGAGATGTCACTAATCGAACGATTG
A3(2)     CACTACAATACGAAGGATGGACAAAAGTTTTCATAAAAAATATAGAGCTGTGGATTAATAGGAAGGTAGT
A4(3)     CACTACAGTACACATTGGACAGCTTACGTAAAACGTGGGCGAGAAGCTATATATTTTGATAGTCATAGTA
A5(5)     CACTACAGTACATACATACATACATATTATAATACATAAACATTATAATACAATAACTACAATACAAATT
A6(1)     CACTACAGTACTAAAGACGGCGATTTATCAGACTCTTAACAATTTCACTTGTAACTCCATCAAGACCTGT
A7(8)     CACTACAGTACCTTGATAACTTAGTTGGGAAAATCGTCGGAAATGCTATCCGTTAGGGCCGGATAGCCTC
A8(8)     CACTACAGTACCTGTCGAACGCGTCGTCCGGGACGTGGACACCGTTTGCTTCTGCCTGAGTAAGGGGCTA
T1(2)     CACTACAGTACTTGTTCGCTATCGGTCTCGTGGTCATATTTAGCCTTAGATGGAGTTTACCACCCACTTA
A9(5)     CACTACAGTACTTCAAGCTT
A10(2)    CACTACAATACCAAGCAAATTCATATAATTTGGGCACTTCTTGATAGTGGAAGTCAGACATCGTTCATTA
A11(3)    CACTACAGTACCGAGGTTGAAATTAACGGGCAGGACGATGGTAATCGAGATCGAGCAGTACTTTTTAAGC
A12(2)    CACTACAGTACGTTCACGCTTAGAGAATGGGTGCTTAATTTGGTTTCCAATCTATCAAACTCATGTAACT
A13(2)    CACTACAGTACCATCTTATTTTGACTGATAGTGCAGGTGATAGTTTGTGCCTTTTGTTGCGTGATATGTT
T2(1)     CACTACAGTACTTATCTCCGCCGGTATAATTTCTTCCGCCCACGACTCACAAACCATAGGTATAGTTGTG
T3(1)     CACTGCAGTATATTTGCTTACCTGAAAATATTACCTGTCCAGAAAATACATAAAAAAAACAATCAACAAT
T4(2)     CACTACAGTATCAAACATGCACCTAATTTTTTTCCACATGATTTTTGCACT
F1(1)     CACTATAGTACTCTTTTTGTGGATGAGTGAAATTTGGAAGGGATCTTTCCAACCTGCCAATGGGCAGCTG
F2(1)     CACTATAGTACTATTTCAAATTCGTTATTTATCAGTAACCTTAATACATAGATTTTTTTACAATTTTTAC
F3(1)     CACTATAGTACCCAAGCACATGTTTCCATCCATGGGTTATCCCAGTTGCCATGCATACTAAAGCACTTTT
F4(7)     CACTATAGTACTTAAGTCTGGGTCGGGCTTCAGAAGGGACCGGAGTGTTACGAGAAAGCAACAACTTTGG
F5(5)     CACTATAGTACTCTGCACTTGTTATACACGCGTTAATACATTATATGTACTTCAAGAACGCCTATGGCGA
F6(3)     CACTACAGTACCTATTAGCTGTTTCAAAACTATAAAAACCGCCAACGTCGCATTTTCTCTTAAAATCACC
```

```
          1         *         *         *         *         *         *        70
Woot      TGTAGCGTATCAATTAACTAGACGAGCCAAACCATTTCAAAAAAGAAAACAAAAGTAAAATTCAAG-TT
T5(3)     TGTAGCGAATCAATTAACTAGACGAGCCAAATCATTTCGAAAAAAAAA--CAAAAGTAAAATTCAAGAGT
A14(4)    TGTAGCGTATCAATTAACTAAACGAGCCAAACCATTTCGAGAAAAGAAAACAAAAGTAAAATTCAAGATT
F7(2)                   AACTAGACGAGCCAAATCAATTCGAAAAT------CAAAAACAAAATTCAAGATT
A15(9)       GCGTATCAATTAACTAGACGAGCCAAACCATTTC-AAAAAAGAAAACAAAAGTAAAATTCAAGATT

                    *         *         *         *         *         *       140
Woot      AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCGCAGTCT
T5(3)     AATTAAATAAATAAATTAAA-TAAGTTGTAAATCATTTAAACATCGNTCAACCATCATCAATCACGGTCT
A14(4)    AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACAGTCT
F7(2)     AATTAAATAAATAAATT----------GTAAATAAGTTAAACATCAATCAACCTTT---AATCACGGTCT
A15(9)    AATTAAATAAATCAATTAAAATAAGTTGTAAATCATTTAAACATCGATCAACCATCATCAATCACAGTCT

                    *         *         *         *         *         *       210
Woot      CACTACA------------TTTTTATTGGTGCCGTGACCAGGATTCAAAGTCATAAAGTGTCATTTCCGA
T5(3)     CACTACAGATACTACA--TTTATTTTTGGTGCCGTGACCAGGATTCAAAGTCATAAAGTGTCATTTCCGA
A14(4)    CACTACA------------TTTTTATTGGTGCCGTGACCAGGATTCAAAGTCATAAAGTGTCATTTCCGA
F7(2)     CACTATAGACAGACAAAAAAAATTTTTGGTGCCGTGACCAGGATTCAAAGTCATAAAGTGTCATTTCCGA
A15(9)    CACTACA------------TTTTTAGAGTAAGTTAAAAACGAGACAAGTTTAGTGTTCGTTCGATTAGG
```

Figure 2. Nucleotide sequences of four internal (nontarget) *Woot* fragments amplified by universal PCR, compared to that of the original *Woot* copy (Beeman *et al.*, 1996a). See Fig. 1 legend for explanation of clone designations, numbering, dashes and boldface type. The ACA motif boxed at nt 145–147 represents the 3′ terminus of the LTR. The sequences immediately downstream correspond to the start of the *Woot* retrotransposon body, which contains the coding regions. The single exception is A15(9), which contains a *Woot* LTR inserted into the middle of (or in tandem with a fragment of) another *Woot* LTR in direct (head-to-tail) orientation. The 'host' LTR is italicized and angled to emphasize lack of homology to internal *Woot* body sequences.

insertion, or that the insertion is present as a balanced lethal. In contrast, the *Tiw-1* strain appears to be devoid of *Woot* insertions at these sites, since junction fragments could not be amplified by PCR using junction-specific primers and *Tiw-1* DNA as template. Comparison of sequences confirmed that the four-base target sequence had undergone direct duplication at either end of the *Woot* insertion at both A2 and A3 (Fig. 3). Such target duplications are typical of insertion footprints for both retrotransposons and DNA transposons.

We estimate that roughly half the total number of *Woot* junctions present in the A^4/Ey strain have been cloned. This estimate is based on direct counts of junction fragments resolved by field inversion gel electrophoresis (FIGE) and visualized by Southern hybridization analysis (Beeman *et al.*, 1996; M. S. Thomson, personal communication).

Universal PCR is a simple and effective general method for short-distance chromosome walking, applicable to either multiple- or single-copy sequences

A2 insertion site

```
   5' jnct AGTGCACCGCCTCGACCAAATCTATGCTCAAAGTACTGTGATGATCGCGGTATTTA
empty site AGTGCACCGCCTCGACCAAATCTATGCTCAAAGTACTGATTCTTTTATTGATTTTGGA
   3' jnct
                          TCAATCACGGTCTCACTACAGTACTGATTCTTTTATTGATTTTGGA
```

A3 insertion site

```
   5' jnct CTTTTACAAAAATGGCAGGAAAGTTGGAATACATACTGTGATGATCGCGGTATTTA
empty site CTTTTACAAAAATGGCAGGAAAGTTGGAATACATACGAAGGATGGACAAAAGTTTT
   3' jnct
                          TCAATCACGGTCTCACTACAATACGAAGGATGGACAAAAGTTTT
```

Figure 3. Nucleotide sequences of upstream and downstream insertion junctions, and the corresponding empty insertion site, for two sites exhibiting dimorphism for the presence of a *Woot* insertion. The A2 and A3 sites correspond to those shown in Fig. 2. Junction fragments were derived by universal PCR amplification from the A^4/Ey strain, whereas empty-site fragments were derived from both the A^4/Ey and *Tiw-1* strains (sequences were identical in the two strains). Host (non-*Woot*) sequences are indicated by normal font style, whereas *Woot* LTR sequences are italicized and angled to emphasize lack of homology to flanking host sequence at the corresponding empty site. The insertion target is boxed. Note that the target sequence is reiterated on both sides of the *Woot* insertion, indicative of duplication at the time of insertion.

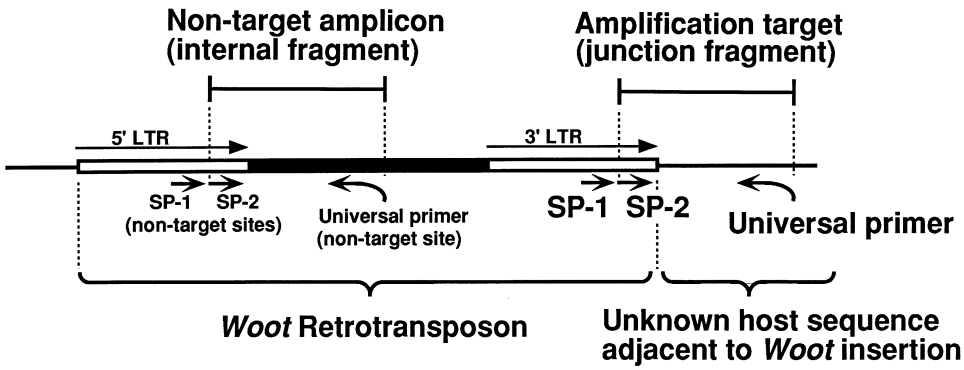

Figure 4. Amplification of unknown insertion junction fragments for a known retrotransposon by universal PCR. SP-1 and SP-2 indicate nested *Woot*-specific primers complementary to a region near the 3′ end of the retrotransposon, and with their 3′ ends facing downstream. The various DNA species in the highly degenerate universal primer mixture anneal at many places throughout the genome, but priming within 1 kb downstream of the *Woot* LTR can lead to amplification of the indicated target junction fragment. The internal nontarget amplicon is a result of the presence of the sequence identity of the LTRs. This amplicon derives from universal primer priming internally within the *Woot* element, and specific primer priming in the 5′ LTR.

in either small or large genomes (this work; Sarkar *et al.*, 1993). The method appears to offer an easy alternative to the more cumbersome inverse PCR approach. As shown in this work, it is the method of choice for molecular characterization of transposon insertions into genomes. It would appear to be ideally suited for molecular confirmation of putative transformant lines in gene transfer experiments, either in cultured cells or *in vivo*. We have used the method to clone pantropic retroviral insertion junctions in mosquito cultured cell chromosomes (Matsubara *et al.*, 1996).

Experimental procedures

The A^4/*Ey* and *Tiw-1* strains of *T. castaneum* are described in Beeman *et al.* (1996a, b). The former is derived from a strain collected in Georgia (USA) in 1980, and the latter was collected in the village of Puranpur in the Bareilly District of Uttar Pradesh, India, in 1988. The *T. freemani* strain was obtained from A. Sokoloff, and has been described by Nakakita *et al.* (1981). Genomic DNA was prepared as described (Brown *et al.*, 1990).

The universal PCR procedure was modified from that described by Sarkar *et al.* (1993) (Fig. 4). These authors used the method for short-distance chromosome walking in the coagulation factor IX gene in several mammalian species. We designed outwardly facing, nested primers based on sequence near the 3′ terminus of the long terminal repeat (LTR) of the *Woot* retrotransposon. The downstream insertion junctions were targeted in this work, since the 5′ end of the LTR is less suitable for primer design. The nested, *Woot* LTR (forward) primers were (1) 5′-ATAAGAGTGACTTCGTACATC-3′ and (2) 5′-AAACAACACAACGTAGGTAAG-3′, terminating 180 and 158 nucleotides (nt), respectively, upstream of the 3′ terminus (SP-1 and SP-2 in Fig. 4). The universal (reverse) primers each consisted of three regions: A 3′ anchor sequence (4–6 nt), a fully degenerate region (8–10 nt) and a 5′ linker (18–20 nt) (Fig. 4). The universal primers anneal at many sites throughout the genome, including sites internal to *Woot*. Since retroelements terminate on both ends with LTRs in direct orientation, annealing of the LTR primers to the upstream LTR can lead to amplification of nontarget internal fragments, as shown in Fig. 4. Table 1 lists the universal primers that amplified one or more *Woot* insertion junctions or internal fragments, as confirmed by cloning and DNA sequence analysis.

Two rounds of nested PCR were done. Round 1 used *Woot* LTR primer 1 (0.1 pmol/μl) and one of the universal primers (1.0 pmol/μl), 10–1000 ng of genomic DNA as template, 1.5 mM $MgCl_2$, 200 μM dNTPs, 1 × Promega Taq pol buffer and *c.* 0.5 units Taq pol in a total volume of 25 μl. The thermocycler program was 30 cycles of 1 min at 94°C, 2 min at 50°C, and 3 min at 72°C, followed by a 10 min final extension at 72°C. For round 2 we used 1 μl of round 1 PCR product as template, 1 pmol/μl each of *Woot* LTR primer 2 and a linker primer corresponding to the 5′ portion of the universal primer used in round 1. Linkers were the T7 phage promoter sequence for universal primers 1–

Table 1. Universal primers used to successfully amplify 3′ *Woot* insertion junctions.

Primer	Sequence (5′-3′)	Junctions*
Uni-1	TAATACGACTCACTATAGGGNNNNNNNNNNCTAT	A1, A6, T2–3, F1–3
Uni–2	TAATACGACTCACTATAGGGNNNNNNNNNAGTGC	A2–3, A10, A12–13, T1, T4, F7
Uni-3	ATTAACCCTCACTAAAGGGNNNNNNNNGAATTC	A4, A11, T5, F6
Uni-4	GGGTTTTCCCAGTCACGACNNNNNNNNAGTACT	A14
Uni-5	GGGTTTTCCCAGTCACGACNNNNNNNNAAGCTT	A5, A9, F5
Uni-7	GGGTTTTCCCAGTCACGACNNNNNNNNGGATCC	F4
Uni-8	GGGTTTTCCCAGTCACGACNNNNNNNNNNCTAG	A7–8
Uni-9	GGGTTTTCCCAGTCACGACNNNNNNNNNCCTAC	A15

Nomenclature of clones refers to Figs 1 and 2.
*Includes nontarget internal *Woot* fragments (see Fig. 2).

2, T3 for universal primer 3, and the M13 '−40' primer sequence for universal primers 4–5 and 7–9 (Table 1). Other conditions and temperature programme were the same as described for round 1.

Round 2 PCR products were resolved on low-melt 1.5% agarose gels. Amplified bands were individually excised, and the DNA eluted using Promega Wizard PCR columns. Eluted DNA was ligated into the InVitrogen TA cloning vector and cloned according to the manufacturer's specifications. Cloned inserts from single colonies were sized by direct PCR analysis of 'toothpicked' cells using *Woot* primer 2 and a vector primer. A colony was touched with the tip of a sterile toothpick, which was then briefly swirled in the PCR reaction cocktail.

As the study progressed we wanted to avoid wasting effort analysing new clones representing previously isolated junctions. Thus, we developed a 'multi-PCR' approach, in which toothpicked clones could be tested for the presence of internal fragments and any of up to thirteen unique junctions in a single PCR reaction. A single *Woot* LTR-specific forward primer was combined with an internal *Woot* reverse primer and thirteen junction-specific reverse primers representing thirteen unique junctions. These primers were selected to be mutually compatible using the Amplify software package developed by William Engels at the University of Wisconsin, Madison. This multi-PCR mixture was tested on individual standard clones representing an internal *Woot* fragment and each of the thirteen unique junctions. Clones scored as negative based on multi-PCR analysis were sequenced to identify new junctions. Plasmid DNA was isolated according to standard procedures (Sambrook *et al.*, 1989) and sequenced on the sense strand (refers to *Woot*), using the Promega Femtomole® cycle sequencing kit and LTR primer 2 as sequencing primer. Preliminary DNA sequence analysis was done using the AssemblyLign software package (IBI), while final alignments were done by hand.

Empty insertion sites were amplified from the *Tiw-1* strain using nested, specific reverse primers in conjunction with a universal forward primer. The specific primers were complementary to regions just downstream of 3′ *Woot* insertion junctions in the A^4*/Ey* strain. To amplify empty site A2 we used specific primers 5′-CGTCAATGGTGATTGCTGTTAG-3′ and 5′-TGATTTTGGCGAGGAGTC-3′ (nested) in conjunction with the universal primer designated 'Uni-2' (see below). To amplify empty site A3 we used specific primers 5′-ACGTGGTGGT-CAACGCCTAC-3′ and 5′-TCAACGCCTACTACCTTCC-3′ (nested), also in conjunction with Uni-2 (see below). Upstream junctions were cloned after specific PCR amplification using a (reverse) internal *Woot*-specific primer and a (forward) upstream junction-specific primer. The former was 5′-CACATGGCCTTTCGACCACC-3′, complementary to nt 229–248 of the *Woot* LTR (Beeman *et al.*, 1996a), and the latter was based on sequence upstream of the empty junction sites (5′-GCCTCGACCAAATCTATG-3′ for junction A2, and 5′-GGCTCTACTGGTTATCGCAG-3′ for junction A3).

References

Arnold, C. and Hodgson, I.J. (1991) Vectorette PCR: a novel approach to genome walking. *PCR Methods Applic* **1:** 39–42.

Beeman, R.W., Thomson, M.S., Clark, J.M., DeCamillis, M.A., Brown, S.J. and Denell, R.E. (1996a) *Woot*, an active, *gypsy*-class retrotransposon in the flour beetle, *Tribolium castaneum*. *Genetics* **143**: 417–426.

Beeman, R.W., Stuart, J.J., Haas, M.S. and Friesen, K.S. (1996b) Chromosome extraction and revision of linkage group 2 in *Tribolium castaneum*. *J Heredity* **87**: 224–232.

Brown, S.J., Henry, J.K., Black, W.C., IV and Denell, R.E. (1990) Molecular genetic manipulation of the red flour beetle: genome organization and cloning of a ribosomal protein gene. *Insect Biochem* **20**: 185–193.

Carlson, J., Olson, K., Higgs, S. and Beaty, B. (1995) Molecular genetic manipulation of mosquito vectors. *Annu Rev Entomol* **40**: 359–388.

Handler, A.M. and O'Brochta, D.A. (1991) Prospects for gene transformation insects. *Annu Rev Entomol* **36**: 159–183.

Jones, D.H. and Winisforter, S.C. (1992) Sequence specific generation of a DNA panhandle permits PCR amplification of unknown flanking DNA. *Nucleic Acids Res* **20**: 595–600.

Lagerstrom, M.J., Parik, J., Malmgren, H., Stewart, J., Peterson, U. and Landegren, U. (1991) Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA. *PCR Methods Applic* **1**: 111–119.

Matsubara, T., Beeman, R.W., Shike, H., Besansky, N.J., Mukabayire, O., Higgs, S., James, A.A. and Burns, J.C. (1996) Pantropic retroviral vectors integrate and express in cells of the malaria mosquito, *Anopheles gambiae*. *Proc Natl Acad Sci USA* **93**: 6181–6185.

Nakakita, H., Omura, O.I and Winks, R.G. (1981) Hybridization between *Tribolium freemani* (Hinton) and *Tribolium castaneum* (Herbst) and some preliminary studies on the biology of *Tribolium freemani* (Coleoptera: Tenebrionidae). *Appl Entomol Zool* **16**: 209–215.

Ochman, H., Ayala, F.J. and Hartl, D.L. (1993) Use of polymerase chain reaction to amplify segments outside boundaries of known sequences. *Methods Enzymol* **218**: 309–321.

Ochman, H., Gerber, A.S. and Hartl, D.L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**: 621–623.

Parker, J.D., Rabinovitch, P.S. and Burmer, G.C. (1991) Targeted gene walking polymerase chain reaction. *Nucleic Acids Res* **19**: 3055–3060.

Presnail, J.K. and Hoy, M.A. (1992) Stable genetic transformation of a beneficial arthropod, *Metaseiulus occidentalis* (Acari: Phytoseiidae), by a microinjection technique. *Proc Natl Acad Sci USA* **89**: 7732–7736.

Rosenthal, A., Platzer M. and Charnock-Jones, D.S. (1994) Capture PCR: an efficient method for walking along chromosomal DNA and cDNA. *The Polymerase Chain Reaction* (Mullis, K.B., Ferre, F. and Gibbs, R.A., eds), pp. 222–229. Birkhauser Press, Boston.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Press, New York.

Sarkar, G., Turner, R.T. and Bolander, M.E. (1993) Restriction-site PCR: A direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *PCR Methods Applic* **2**: 318–322.

Tamura T, Kanda, T., Takiya, S., Okano, K. and Maekawa, H. (1990) Transient expression of chimeric CAT genes injected into early embryos of the domesticated silkworm, *Bombyx mori*. *Idengaku Zasshi* **65**: 401–410.

Triglia, T., Peterson, M.G. and Kemp, D.J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequence. *Nucleic Acids Res* **16**: 8186.