BS a novel LINE-like element in Drosophila melanogaster

Citation for published version:

Digital Object Identifier (DOI):
10.1093/nar/23.8.1354

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Nucleic Acids Research

Publisher Rights Statement:
RoMEO green

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
ABSTRACT
Transposable elements with long terminal inverted repeats are rare and only one family of elements of this sort has been identified in the genome of Drosophila melanogaster. An insertion associated with the HsBS mutation of the achaete-scute complex has been reported to be a second element of this type. We have determined the complete sequence of this insertion and have shown that it is in fact two copies of a new LINE-like transposable element, that we have called BS, inserted in opposite orientation 337 bp apart. Like other elements of this type, BS has two open reading frames that appear to encode a gag-like polypeptide and a reverse transcriptase. There are few complete BS elements in the five strains of D. melanogaster that we have tested and they appear to transpose infrequently. The events that may have lead to the double BS insertion are discussed in terms of the supposed mechanism of transposition of LINE-like elements.

INTRODUCTION
Transposable elements are an important component of eukaryotic genomes, making up 15% or more of the total DNA. They occur as families of dispersed repeat sequences scattered throughout the genome, the number of copies varying from less than ten to several hundred thousand depending on the element and species concerned. They can be classified according to their structure and presumed mechanism of transposition, and fall into two main classes: elements that transpose by reverse transcription of an RNA intermediate, Class I elements, and elements that transpose directly from DNA to DNA, Class II elements.

There are two types of Class I elements. Class I.1 elements resemble retroviruses in having long terminal direct repeats (LTRs) and open reading frames with the potential to code for polypeptides similar to viral gag, pol and, in some cases, env proteins. Class I.2 elements also have open reading frames with similarities to retroviral gag and pol genes, but have no terminal repeats and end with A-rich sequences at the 3' end of their coding strands. Elements of this type are often referred to as LINE-like elements as the first examples to be detected were mammalian LINE or L1 elements (1).

The majority of Class I elements have short terminal inverted repeats, but some have been reported to have inverted repeats that are hundreds of base pairs long. Few elements with this structure have been identified, the best characterized being the FB elements of Drosophila melanogaster (2). These have inverted repeats 1–2 kb long that are themselves made up of short tandem repeats. Some FB elements contain nothing more than these tandem repeats, while others have a central region of unrelated and non-repeated sequence (3–5). In one case this central region appears to be due to insertion of a different Class II element (6), while the central region of another may code for a function that stimulates FB transposition (7,8).

Campuzano et al. (9) reported finding what appeared to be a member of a second family of elements with long inverted repeats that they named BS because it was found associated with the mutation Hw^RS. This is a derivative of Hw^I, a mutation of the achaete-scute complex (AS–C) that is associated with an insertion of a copy of the retrovirus-like element gypsy within the T5 AS-C transcription unit. Chromosomes carrying the Hw^RS mutation have a slightly reduced hairy wing phenotype and an additional 8 kb of DNA inserted within the Hw^I gypsy element. This extra DNA contains inverted repeats of 2.5 kb as judged by restriction site mapping and hybridization data. The 8 kb insertion does not contain FB sequences and was named BS.

In order to compare the structure of this putative new inverted repeat element with that of FB elements, we have determined the nucleotide sequence of the whole of the BS insertion. The results indicate that the Hw^RS mutation is not associated with insertion of a single inverted repeat element, but has two related sequences inserted 337 bp apart and in opposite orientation within the gypsy element of Hw^I. These are members of a new family of LINE-like elements as they show no strong sequence similarity to any of the five families of such elements, Doc, F, G, I and jockey, that have been identified in the genome of D. melanogaster (10,11).

MATERIALS AND METHODS
DNA sequencing
The sequence of the Hw^RS insertion was determined by subcloning the PstI fragments BS0.6, BS816, BS817, BS818 and BS819, and the PstI–BamHI fragment BS2.0 (Fig. 1A) of the plasmid p14RKHw^BS (9) in the vector pUBS (12). Double-stranded templates were prepared using QIAGEN™ columns and these were sequenced by dideoxynucleotide chain termination (13). The HindIII fragments BS1.1, BS5.8 and BS7.6 were used to orient and link the sequences from the PstI fragments. The entire

* To whom correspondence should be addressed
sequence has been determined on both strands of the element. Sequence data were assembled using programs written by Staden (14) and were analysed using the suite of programs from the Wisconsin Genetics Computer Group (15). The program PL-LEUP is from Version 7.0 of the GCG package (Genetics Computer Group, Inc.).

**Drosophila DNA preparation and Southern blotting**

Genomic DNA was prepared from strains of *Drosophila* by homogenizing 10 flies in homogenization buffer (8 M urea, 0.35 M NaCl, 10 mM Tris–HCl pH 8.3, 10 mM EDTA). Homogenates were extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and then with chloroform:isoamyl alcohol (24:1). Nucleic acids were ethanol precipitated and dissolved in 50 ml of 20 mg/ml RNase in 10 mM Tris–HCl pH 8.3, 1 mM EDTA.

Five micrograms of DNA from each strain were digested with the appropriate restriction enzymes and electrophoresed on 0.7% agarose gels together with amounts of control plasmids equivalent to two, five, eight and 10 copies per genome. The DNA was then transferred to GeneScreen Plus nylon membrane (DuPont) by capillary action in 1.5 M NaCl, 0.15 M Na citrate. Hybridization was carried out in 0.5 M Na phosphate pH 7.2, 7% sodium dodecyl sulphate and 1 mM EDTA. After hybridization filters were washed in 40 mM Na phosphate, 1% sodium dodecyl sulphate, 1 mM EDTA. Fragments used as probe were purified by agarose gel electrophoresis and labelled with [\(^{32}\)P]dCTP by random priming (16). The amount of hybridization in each track was quantified with the aid of a PhosphorImager (Molecular Dynamics).

**RESULTS**

We have determined the sequence of the BS insertion starting from a 14.8 kb EcoRI fragment of *Hw*\(^{BS}\) genomic DNA kindly supplied by Dr S. Campuzano. This contains the entire gypsy and BS insertions associated with the *Hw*\(^{I}\) mutation (Fig. 1A). The sequence of the BS insertion is diagrammed in Figure 1B and is in the EMBL data base as accession X77571. It is not an insertion of a single inverted repeat element, as had been thought, but is made up of two copies of an element lying close together and in opposite orientation within gypsy. We have called the left-hand element BS1 and the right-hand element BS2.

The BS1 and BS2 insertions are 5126 and 2571 bp long, respectively, and each is flanked by a duplication of a 12 bp sequence present only once in the corresponding part of the gypsy sequence. These sequences are different for BS1 and BS2, and were presumably generated when these elements inserted at these sites. These insertions are separated by 337 bp of gypsy DNA. Geyer et al. (17) have determined what they believed to be the ends of the whole BS insertion when analysing insertions that modify the phenotype of mutations due to gypsy insertions. The sequences that they reported were those of the ends of BS2.

The structure of the BS1 element indicates that it is a LINE-like (Class I.2) transposable element. It has no terminal repeats, has an A-rich sequence at the 3' end of the coding strand and has two long open reading frames (Fig. 1B). The first has three copies of
a CCHC motif like those found in the gag genes of retroviruses and in the first open reading frames of LINE-like elements (Fig. 2A). The second has motifs characteristic of a reverse transcriptase (Fig. 2B). Five other transposable elements of this type, Doc, F, G, I and jockey, have been found in the genome of D. melanogaster (10,11), but the sequence of BS1 has not been reported previously. Elements of this type are frequently truncated by varying amounts at their 5' ends (1,18). This is thought to happen at some point during transposition (19). The sequence of BS2 is identical to BS1, but is missing the first 2555 bp, suggesting that it has suffered such a deletion.

We have determined the relationships between the amino acid sequences of the putative reverse transcriptases of BS1 and the five Drosophila LINE-like elements mentioned above as well as those of two other potential transposable elements of this type, R1Dm and R2Dm (20), that are found inserted at specific sites within a proportion of the 28S rRNA coding sequences of D. melanogaster. These are believed to differ from the other LINE-like elements in being able to insert at these sites in a sequence-specific manner. This was done by comparing the complete amino acid sequence encoded by the second open reading frame of each element using the GCG program PILEUP. The result (Fig. 3) suggests that BS is most closely related to jockey.

We have used Southern transfer experiments to measure the number of BS elements in the genomes of five different stocks of D. melanogaster and to estimate their degree of conservation (Fig. 4). Three restriction fragments of BS1 were used to probe digests of genomic DNA using enzymes that should excise the corresponding fragments from complete genomic copies. Digests of a plasmid containing the fragment used as probe were used as size markers and to give an indication of the number of copies per genome. The intensity of hybridization of fragments corresponding in size to the probe decreases from the 3' to 5' end of the element (Fig. 4A–C) confirming that BS elements are a family of sequences containing variable 5' deletions.

There are about five copies of the 3' most fragment tested (fragment A in Fig. 1B) in the strains tested. The higher molecular weight bands presumably correspond to BS elements truncated somewhere within fragment A. This gives an upper limit of about 10 to the number of BS elements in the genome. The results from using fragment C as probe suggest that the strains tested contain few if any complete BS elements.

Campuzano et al. (9) estimated the copy number of BS elements to be 15 using the 1.7 kb PstI fragment (fragment D in Fig. 1B) as probe. This fragment includes both internal gypsy sequences and sequences from the 5' end of BS so this estimate reflects the number of copies of both elements. We were unable to detect fragments corresponding to this 1.7 kb HindIII–PstI fragment in DNA of any of the five strains tested (Fig. 4D), indicating that the double insertion of BS elements within gypsy does not occur in any of these strains.

We have compared the distribution of BS elements in the genomes of five strains of D. melanogaster by probing PstI-digested DNA with a BglII fragment from the 3' end of BS1. Each BS element should give a band of hybridization that depends on its position in the genome. The patterns of hybridization are similar from strain to strain (Fig. 5) suggesting that BS elements transpose infrequently in them. Unfortunately we have not been able to obtain the Hw-BS strain to check that it does not have an unusually high frequency of BS transposition.
but is inverted repeat elements within Figure 4. Track M probe was wild. Genomic 3, Charolles; per haploid allow Lb. In Figure the Figure 5. Southern in the of genome. shows the conservation of BS elements within the genomes of five strains of D. melanogaster isolated from the wild. Genomic DNAs were digested with PstI in (A) and (B) and with PstI and HindIII in (C) and (D). The fragments used to probe these DNAs are shown in Figure 1B. In each case a digest of a plasmid including the fragment used as probe was loaded on the gel in varying amounts to serve as a size marker and to allow the amount of hybridization of the genomic DNAs to be quantified. Track M shows the plasmid DNA loaded in an amount equivalent to 15 copies per haploid genome. The strains of D. melanogaster are as follows; 1, w; 2, Samarkand; 3, Charolles; 4, Oregon R; 5, Canton S.

**Figure 4.** Southern transfer experiment showing the conservation of BS elements within the genomes of five strains of D. melanogaster isolated from the wild. Genomic DNAs were digested with PstI in (A) and (B) and with PstI and HindIII in (C) and (D). The fragments used to probe these DNAs are shown in Figure 1B.

**Figure 5.** Southern transfer experiment showing the degree of heterogeneity in the distribution of BS elements in the genomes of the strains of D. melanogaster analysed in Figure 4. Genomic DNAs were digested with PstI and probed with the BglII fragment of BS, Fragment E in Figure 1B.

**DISCUSSION**

The insertion associated with the HwBS mutation is not an inverted repeat element like FB, as had been supposed previously, but is formed by two copies of a new element, BS, inserted in opposite orientation. These elements have all the characteristics of LINE-like elements and BS1 may be full-length.

Although the proportion of the D. melanogaster genome that comprises transposable elements is not unusually large, the number of different transposable elements of all kinds that can be found in this species is greater than that known for any other eukaryote (21). Eight LINE-like elements have now been identified in D. melanogaster as well as a short sequence that may represent a ninth family (22). This is the greatest number of elements of this type to have been found in any species. Five of these, Doc F, I, jockey and, presumably, BS, can insert at many sites in the genome while two of them, R1Dm and R2Dm, are site-specific elements (23,24). G elements are concentrated in the non-transcribed spacer DNA of ribosomal gene clusters and may not be active in most strains (25).

The BS2 element has inserted within one copy of a tandemly repeated 12 bp sequence that is the binding site for the product of the suppressor of Hairy wing gene (26). Binding of su(Hw) protein to these repeats in a gypsy element inserted upstream of the yellow gene alters yellow expression by interfering with the interaction of the yellow promoter and some of its upstream regulatory sequences (27,28). Insertion of the transposable elements hobo or jockey within these repeats suppresses the mutant phenotype of y2 presumably because this reduces the binding of su(Hw) protein (17).

The gypsy element associated with HwI is inserted within the transcribed region of the T5 (achaete) gene of the AS–C and the mutant allele is transcribed to give a hybrid T5–gypsy transcript that is about 10-fold more abundant than T5 transcripts from the wild-type allele (9). The level of this hybrid transcript is reduced in flies carrying the su(Hw) mutation suggesting that in this case binding of su(Hw) protein may prevent interaction of the T5 promoter and a cis-acting negative regulator. The binding of su(Hw) protein to its target may be reduced somewhat by insertion of the BS2 element within one of the 12 bp repeats as the HwI phenotype is slightly attenuated in HwBS flies, although no effect on the level of the T5–gypsy transcript has been detected (9).

The HwBS insertion must have been produced by either two independent transposition events or a single transposition event that generated two copies of BS. Insertions of P elements have been found that are close together and in opposite orientation (29). These are probably the results of independent insertions as P elements transpose at high frequency and preferentially to sites adjacent to a donor element (30). Elements that transpose by reverse transcription of an RNA intermediate, as is the case for LINE-like elements (19,31,32), are unlikely to transpose preferentially to adjacent sites and the frequency of BS transposition appears to be low. This suggests that the HwBS double insertion is the result of a single transposition event.

LINE-like elements are thought to integrate by a mechanism in which the 3'-OH at a nick in a strand of chromosomal DNA is used to prime DNA synthesis by an element encoded reverse transcriptase that uses the RNA transposition intermediate as template (33,34). The DNA synthesized in this way is in turn used as the template for synthesis of the second strand to complete integration. If the RNA intermediate were to be released after synthesis of the first DNA strand it might serve as primer for a second integration event close by resulting in a double insertion.
ACKNOWLEDGEMENTS

This work was supported by research grants from the Medical Research Council and the Human Frontier Science Program and a studentship to AU from the Government of Thailand. We are grateful to S. Campuzano for supplying the original BS clone.

REFERENCES