GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Structural analysis of a 4414-bp element in *Drosophila melanogaster*

B. Yu[1], X.T. Wang[2], H.W. Li[1], C.J. Zhao[1], C.X. Wu[1] and X.M. Deng[1]

[1]National Engineering Laboratory for Animal Breeding & the Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, College of Animal Science and Technology, China Agricultural University, Haidian District, Beijing, P.R. China
[2]Institute of Oceanology, Chinese Academy of Sciences, Qingdao, P.R. China

Corresponding author: X.M. Deng
E-mail: deng@cau.edu.cn

Genet. Mol. Res. 10 (2): 717-730 (2011)
Received September 13, 2010
Accepted December 7, 2010
Published April 26, 2011
DOI 10.4238/vol10-2gmr987

ABSTRACT. We cloned a 4414-bp element from a mutant of *Drosophila melanogaster*. Its insertion site was 18,929,626 bp. Analysis of the nucleotide and amino acid sequences demonstrated that the element is homologous to Pifo_I, first obtained from *D. yabuka*, which belongs to the *gypsy*/*Ty3* subfamily. We also obtained a 3754-bp length element from a wild-type fly by PCR, with a pair of primers designed from the conserved region of the 4414-bp length element. The two elements included a pair of long terminal repeats and part of the GAG and ENV proteins, but the POL protein was completely lost. This element is found in the subgenus of *D. melanogaster*, but it is a degenerate type of Pifo_I and is not infective. Also, a 714-bp region structured in 5.0 tandem repeats of 143 bp each was found in the 5'UTR of the degenerate element; these could interact with transcription factor CF2. Phylogenetic analysis and alignment of amino acids indicated that the Pifo_I element was closer to the *ZAM* retrotransposon, which gave us some clues to their functional similarity. Based on these data, we propose that there is a relationship between the degenerate element

and the mutant phenotype, which would provide a foundation for further research.

Key words: *gypsy*/*Ty3*; 4414-bp element; 3754-bp element; Pifo_I; *ZAM*

INTRODUCTION

Retrotransposons are genetic elements that can amplify themselves in a genome and are ubiquitous components of the DNA of many eukaryotic organisms: they consist of two sub-types, long terminal repeat (LTR) and non-LTR retrotransposons. LTR retrotransposons are further sub-classified into the *copia*/*Ty*1 and *gypsy*/*Ty*3 subfamilies based both on their degree of sequence similarity and the order of encoded gene products. The LTR retrotransposons and, more specifically, those of the *gypsy*/*Ty3* class share some common structures with retroviruses, such as LTR and *gag* and *pol* genes and, in some cases, incomplete and nonfunctional *env* genes (Capy et al., 1997). Besides, all retroviruses and nearly all LTR retrotranposons share the terminal sequence TG…CA (Dej et al., 1998). On the basis of these similarities, many scientists propose that retroviruses may evolve from LTR retrotransposons by acquiring a functional *env* gene (Capy et al., 1996).

The members of *gypsy*/*Ty*3 subfamily have been described in yeast, filamentous fungi, plants, nematodes, insects, and some marine species. In *Drosophila* species, many retrotransposons belong to this subfamily, such as *gypsy* (Marlor et al., 1986), *412* (Yuki et al., 1986), *17.6* (Saigo et al., 1984), *297* (Inouye et al., 1986), *ZAM* (Leblanc et al., 1997), and so on. In these retrotransposons of *Drosophila*, *gypsy* is well studied in *Drosophila melanogaster* because it can be genetically tractable and is closely related to the non-infectious LTR retrotranposons but infectious like retroviruses (Song et al., 1994). Another main reason is that sequences homologous to *gypsyDm* (*gypsy* element exists in *D. melanogaster*) are widely distributed in *Drosophila* species, found in both the *Sophophora* subgenus and the *Drosophila* subgenus, and also detected in some species of the *Scaptomyza* genus (Herédia et al., 2004), all of which make *gypsy* a canonical retrotransposon to research in the evolutionary history of and the relationship between retrotransposons and retroviruses. *gypsy* is one of a very small number of elements that have a non-TG…CA terminal sequence and can insert into many different loci in genomic DNA but has weak consensus target sequences. Despite that it is reported that *gypsy* also has a "preferable site" for its insertion into genomic DNA, in nearly every example, the target sequence TATATA or TACATA is duplicated upon insertion (Marlor et al., 1986). However, Dej et al. (1998) have suggested that other target sequences could also be recognized and the consensus could be replaced by a non-strict type consisting of alternating pyrimidine (Y) and purine (R) residues, that is, a YRYRYR consensus target sequence, by sequencing three *gypsy* insertions into the *ovo* gene. Compared to *gypsy*, *ZAM*, another member of the *gypsy*/*Ty3* class, is highly sequence-specific in its integration, namely, GCGCGCg (the lowercase "g" indicates a <50% occurrence of that base) (Leblanc et al., 1997).

Although retroviruses and retrotransposons share the common intracellular mechanism of expansion on the basis of the homologous products encoded by open reading frame 1 (ORF1) and ORF2, the essential difference between them lies in the presence of the ENV protein encoded by ORF3, which is necessary for the vital extracellular stage (Alberola et al., 1997). In general, the envelope (*env*) gene is absent or nonfunctional in LTR retrotransposons.

Up to now, ORF3 has only been described in *gypsy*, *17.6*, *297*, *Tom*, and *ZAM*. However, it has been reported that the *gypsy* retrotransposon has the ability to be infectious, and its infectious properties are due to its *env*-like gene (Teysset et al., 1998). A functional ENV protein could be produced in 10 species of the *obscura* group of the *Drosophila* genus and one species from the *Scaptomyza* genus (Llorens et al., 2008). Besides, *ZAM* can give rise to an 8.6-kb full-length RNA and a 1.7-kb spliced message for *env* gene, which provides evidence that *ZAM* is mobilized through a reverse transcriptional process in the germ line of flies (Leblanc et al., 1997).

In this study, a 4414-bp element was cloned in a mutant of *D. melanogaster* by long and accurate polymerase chain reaction (LA-PCR) and depicted the characteristics of the element, which further confirmed that the element is a degenerate type of Pifo_I, which belongs to the *gypsy*/*Ty3* subfamily. Besides, a 714-bp region structured in 5.0 tandem repeats of 143 bp each was found in the 5'UTR of the degenerate element, which could bind tightly to transcription factor CF2-II. Based on phylogenetic analysis and the alignment of amino acids, we infer that Pifo_I, like *ZAM* retrotransposon, functions as an insulator. This hypothesis could explain the relationship between the degenerate element and the downstream gene in the mutant flies and demonstrate the reason of the mutant phenotype.

MATERIAL AND METHODS

Fly stocks

Both the wild-type flies (*Canton Special*) and the mutant flies belonged to *D. melanogaster* group. All flies were reared in cornmeal medium at constant temperature and humidity (25°C; 60% relative humidity).

DNA extraction and LA-PCR

Approximately 30 adult flies per sample were anesthetized and collected in a 1.5-mL microcentrifuge tube. Genomic DNA was extracted according to Jowett (1986).

We then carried out LA-PCR to amplify the extracted genomic DNA. The primers were designed using Primer 3 (Whitehead Institute/MIT Center for Genome Research; http://www.genome.wi.mit.edu). The total volume was 50 µL containing 10 pM primers, 25 mM $MgCl_2$, 2.5 mM of each dNTPs, 10X LA-PCR buffer (Mg^{2+} free) and LA Taq DNA polymerase (TaKaRa, Japan). Cycle conditions were 94°C for 1 min, followed by 30 cycles at 98°C for 10 s and 68°C for 15 min, with a final extension at 72°C for 10 min. The products of PCR amplification were purified with the Gel Band Purification kit (TIANGEN, China), cloned into the pMD-18T vector (TaKaRa), and then sent to SinoGenoMax Co., Ltd. (China) to carry out DNA sequencing after detection on a 1% agarose gel.

DNA-binding assays

For electrophoretic mobility shift assays with nuclear extracts, reaction mixtures contained 1.5 µg poly(dI-dC), 1.5 µL 10X binding buffer, 2.75 µL nuclear extracts, 0.6 µL biotin-probe DNA (500 fM), and competitor at 50 and 100X molar ratio. Water was added to a final volume of 15 µL. All components except for biotin probe were mixed and incubated for 15

min at room temperature; the probes were then added and reactions were carried out at room temperature for 20 min. Samples were loaded onto a 6.5% non-denaturing polyacrylamide gel that was run on ice at 180 V for 60 min, then transferred to nylon membrane in 0.5X TBE at 500 mA for 30 min. After cross-linking DNA in a UV linker, the results were detected with chemiluminescence image analyzer.

Nucleotide and amino acid sequence analyses

The sequence obtained from the mutant flies by LA-PCR was submitted at the GIRI website (http://www.girinst.org/repbase/index.html) to search for the "censor" homologous portions. Meanwhile, the characteristics of the sequence were also analyzed.

The nucleotide sequences of retrotransposons and retroviruses used in this study were obtained from GIRI, EMBL and GenBank. Their names and accession numbers are given in Table 1. These nucleotide sequences and their respective amino acid sequences were aligned using the ClustalW 2.0 program (developed by University College Dublin) and manual adjustments of the alignments were performed when necessary. The phylogenetic analyses were carried out using MEGA version 4.0 (Kumar et al., 2008). The methods utilized for tree inference were neighbor-joining (NJ) and maximum parsimony (MP). Bootstrap tests with 500 samplings were performed for the trees obtained.

Table 1. Accession numbers of the different elements used in the study.

Element	Host	Type	Used for protein alignment	Accession No.
Pifo_I	*Drosophila yabuka*	LTR retrotransposon	Y	-
ZAM	*Drosophila melanogaster*	LTR retrotransposon	Y	AJ000387
gypsyDm	*Drosophila melanogaster*	LTR retrotransposon	Y	AF033821
17.6	*Drosophila melanogaster*	LTR retrotransposon	Y	X01472
297	*Drosophila melanogaster*	LTR retrotransposon	Y	X03431
gypsyDvir	*Drosophila virilis*	LTR retrotransposon	Y	M38438
Ulysses	*Drosophila virilis*	LTR retrotransposon	N	X56645
Cer-1	*Caenorhabditis elegans*	LTR retrotransposon	N	U15406
Ty3	*Saccharomyces cerevisiae*	LTR retrotransposon	N	M23367
HIV-1	Human immunodeficiency virus type 1	Retrovirus	Y	AY805330
SIV	Simian immunodeficiency virus	Retrovirus	Y	AY159322

LTR = long terminal repeat.

RESULTS AND DISCUSSION

Sequence cloning and nucleotide analysis

At the beginning of the experiment, we mapped the mutant site in a 5-kb region on the right arm of the 2nd chromosome. We then designed a pair of primers to amplify the region from genomic DNA of wild-type and the mutant flies by a standard PCR method. We also selected the Hn^{bp} strain as a reference (Wang et al., 2008), because our mutant flies were first screened in this strain. A unique, specific band with the expected size of 1 kb, confirmed by DNA sequencing, could be amplified in wild-type flies and the Hn^{bp} strain, but no band was obtained in the mutant flies. We then adopted LA-PCR to amplify the three types of DNAs, then positive results were obtained (Figure 1A). As indicated in Figure 1A, the same expected-size bands were amplified by LA-PCR in genomic DNA of wild-type and the Hn^{bp} strain,

while much bands with larger size were amplified in the mutant. In order to confirm that the appearance of the band with the larger size was not an occasional event, single fly genomic DNA was extracted and used to amplify. The same results were obtained, as shown in Figure 1B. The band with larger size amplified in the mutant flies' DNA was purified and sub-cloned into pMD-18T vector to carry out DNA sequencing at SinoGenoMax Co., Ltd.

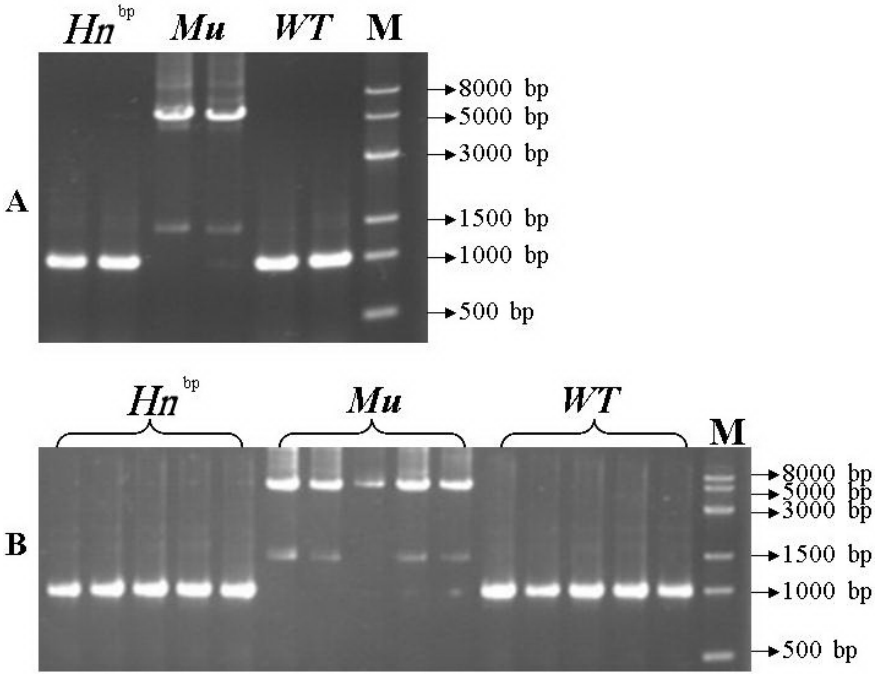

Figure 1. Results of long and accurate polymerase chain reaction. Bands in **A** were obtained by amplifying pooled genomic DNA of three types of strains, respectively. **B** showed the results of amplifying single fly DNA of three strains. *WT* = wild-type; *Mu* = mutant; Hn^{bp} = a fly strain with an allele of *Henna* gene. M = DNA marker.

After getting the sequence of the fragment, we analyzed the nucleotide structure and found that the fragment contained a 4414-bp retrotransposon-like element. We further confirmed that the element was inserted in the site of 18,929,626 bp upstream of the *twist* gene.

The element was submitted to the 'Repeat Masking' subtitle at the GIRI website to search for the "censor" homologous portions (Kohany et al., 2006). Table 2 shows that the 4414-bp element is highly similar to the Pifo_I element, which belongs to the *gypsy* class. Pifo_I is an internal portion of Pifo retrovirus-like element, which has a pair of 444-bp LTRs. The original sequence of Pifo_I element is cloned from *D. yakuba* and its length is 8112 bp with three ORFs that encode proteins homologous to GAG, POL and ENV in retroviruses, respectively (Bartolome et al., 2009). Table 2 shows that compared to the Pifo_I element, the 4414-bp element contains a pair of LTRs (similarity: 0.9775) and two internal portions of Pifo_I retrovirus-like elements, encoding a partial GAG protein (similarity: 0.9960) and ENV protein (similarity: 0.9756), but with the *pol* gene completely lost.

Table 2. Censor result of 4414-bp element in GIRI.

Query sequence			Original sequence			Class	Dir	Similarity	Score	Type
Name	From	To	Name	From	To					
4414-bp	2	447	Pifo_LTR	1	444	LTR/Gypsy	d	0.9775	3998	LTR
element	450	2710	Pifo_I	4	2285	LTR/Gypsy	d	0.9756	18405	Partial of GAG protein
	2713	3968	Pifo_I	6857	8112	LTR/Gypsy	d	0.9960	11288	ENV protein
	3969	4414	Pifo_LTR	1	444	LTR/Gypsy	d	0.9775	3998	LTR

LTR = long terminal repeat.

The 4414-bp element was homologous to Pifo_I element, allocated to the *gypsy* class. Does it then share some common features in nucleotide structure with the *gypsy* retrotransposon? *gypsy* with three ORFs corresponding to *gag*, *pol* and *env* gene in different species of *Drosophila* has unusual termini relative to other retrotransposons and retroviruses, which universally begin and end with the dinucleotide inverted repeat TG…CA. LTR terminal sequences of the 4414-bp element and Pifo_I conformed to the *gypsy* termini, and differed from other retrotransposons and retroviruses (Figure 2).

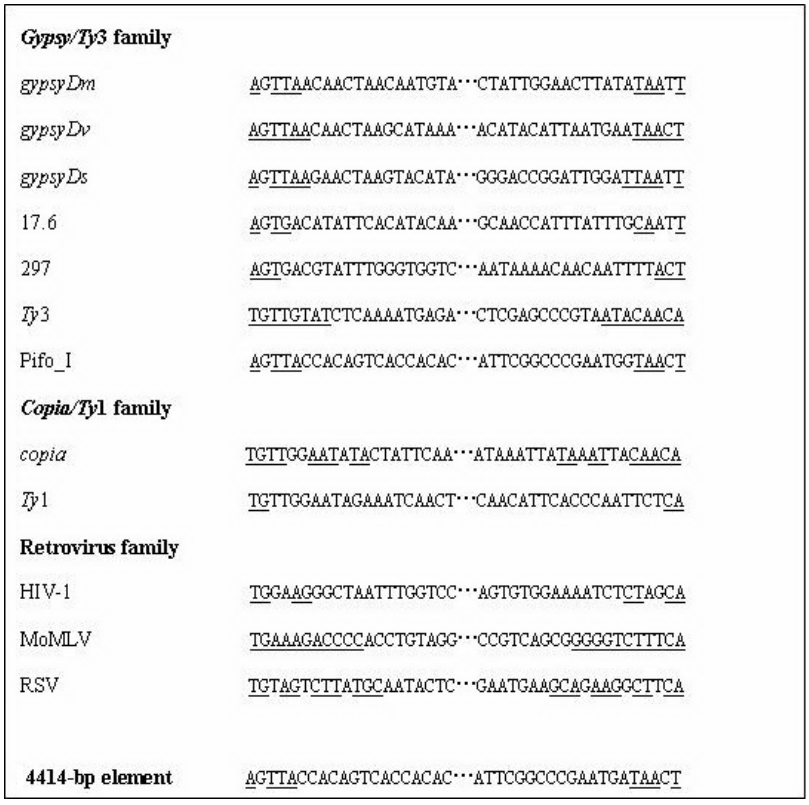

Figure 2. Long terminal repeat (LTR) termini of retroelements. Compared with *copia*/*Ty1* and retroviruses family, the members in *gypsy*/*Ty3* family had no TG…CA terminal sequence, while the 4414-bp element and Pifo_I shared the same LTR terminal sequence with *gypsy* retrotransposon in different *Drosophila* species.

We also analyzed the insertion site of the 4414-bp element in the genome. Examination of the target sites of *gypsy* retrotransposon by sequencing in nearly every example showed that the target sequence TATATA or TACATA is duplicated upon insertion (Marlor et al., 1986). By studying the "consensus sequences" of *gypsy* in the *ovo* gene, Dej et al. (1998) revealed that the consensus could be relaxed to a slightly more degenerate one consisting of alternating pyrimidine (Y) and purine (R) residues, that is, YRYRYR consensus target sequence. Because the alternating pyrimidine and purine residues are a typical feature of DNA sequence able to adopt the Z conformation, especially poly $(dGC)_2$ (Zhang et al., 2006), they further raised the possibility that *gypsy* recognized this DNA structure feature (Dej et al., 1998). The target site of the 4414-bp length element acquired was 5'-GCGCGCGCGCGC-3', which conformed to the alternating purine-pyrimidine sequence and the typical structure of Z-DNA. Besides, the target sequence was similar to the *ZAM* retrotransposon (5'- GCGCGCg-3'). Based on this result, we could infer that Pifo_I also has the $(GC)_6$ target site.

We confirmed that the 4414-bp element was inserted at the site of 18,929,626 bp, which is located upstream in the negative regulatory region of the *twist* gene. Meanwhile, this insertion destroyed the $(GC)_6$ structure, which could further destroy the conformation of Z-DNA. No definitive biological significance has been found for Z-DNA, but it is believed to provide torsional strain relief while DNA transcription occurs (Ha et al., 2005). The potential to form a Z-DNA structure also correlates with regions of active transcription. By analysis of the human chromosome 22 genomic sequence, Champ et al. (2004) showed that both Z-DNA-forming regions and promoter sites for nuclear factor-I correlate with the location of known and predicted genes across the chromosome and accumulate around the transcriptional start sites of the known gene. Therefore, we supposed that the insertion of the element influenced the *twist* gene downstream and further resulted in the mutant phenotype. However, Q-PCR data showed that the expression of *twist* gene in the mutant was not significantly different compared to the wild-type (data not shown).

The organization of 5'UTRs

In this study, the sequences of *gypsy*, *ZAM*, Pifo_I, and the 4414-bp element were analyzed for the presence of tandem repeats, using the "Dot Matrix of Comparison" function of DNAMAN, version 5.2.2 (Lynnon Corp., Quebec, Canada). This analysis revealed that all four sequences possessed tandem repeats in their 5'UTR (Figure 3). These repeats were further characterized for length, number and linguistic complexity using the "Tandem Repeat Finder" (Benson, 1999). This program can locate and display tandem repeats in a DNA sequence, and its output contains information about each repeat, including location, size, copy number, and nucleotide content.

The analysis of the four sequences performed by the Tandem Repeat Finder is shown in Table 3. Table 3 reveals that all four sequences contain the tandem repeats in their 5'UTRs, but differ in copy number of their tandem repeats and period size. The 4414-bp element showed a 714-bp region structured in 5.0 tandem repeats of 143 bp each, and the entropy was up to 1.98. Pifo_I also contained a sequence with 4.0 tandem repeats of 143 bp each. Therefore, the 4414-bp element and Pifo_I have a closer relationship.

We know that tandem repeats are very useful in linkage analysis and DNA fingerprinting because of its polymorphism in a specific population (Weber and May, 1989). As a matter

Table 3. Features of the four retrotransposons possessing repeats in their 5'UTR.

Retrotransposon	Position of repeats	Length of sequence	Period size	Copy number	Entropy (0-2)
gypsy	540-768	228	109	2.1	1.81
	541-603	62	27	2.3	1.77
	593-767	174	55	3.2	1.82
	650-712	62	27	2.3	1.77
ZAM	726-1425	699	307	2.3	1.89
Pifo_I	946-1518	572	143	4.0	1.98
	1982-2022	40	20	2.0	1.44
4414-bp element	797-1511	714	143	5.0	1.98

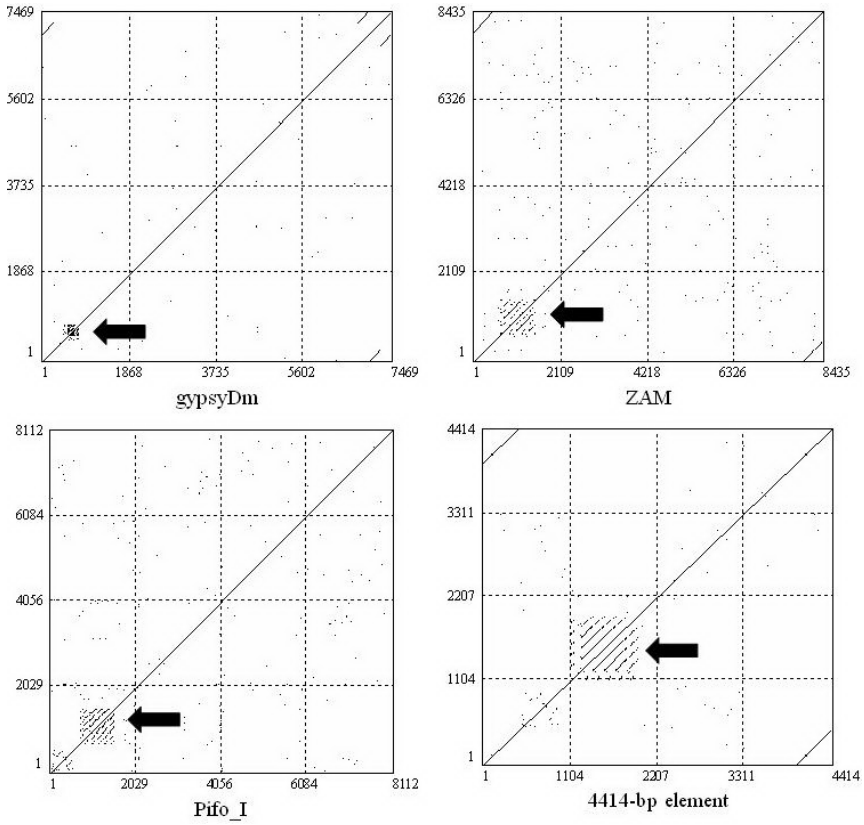

Figure 3. DNAMAN "Dot Matrix Comparison" analysis. The presence of repeated sequences within the four retrotransposons is shown by black arrows: the four sequences show different kinds of repeat in their 5'UTR.

of fact, tandem repeats may interact with transcription factors, alter the structure of chromatin or act as protein-binding sites, which provides evidence that the repeats are involved in gene regulation (Nisha et al., 2008). Minervini et al. (2007) reported that heterochromatin protein 1 could interact with 5'UTR of *ZAM* in a sequence-specific fashion. Therefore, the next step was to examine whether some protein could interact with the tandem repeats in the 5'UTR of the 4414-bp element.

The 715-bp tandem repeat was submitted to "ConSite" in JASPAR (Sandelin et al., 2004 - at http://www.phylofoot.org/) to detect a transcriptional factor binding site using phylogenetic footprinting. The result showed that CF2-II could bind to the repeat, which was confirmed by electrophoretic mobility shift assays (Figure 4). A 27-bp sequence in 143 bp, 5'-CAACGTATATACACACATATACACCAA-3', was used as a probe, which contained the core recognition sequence of the transcription factor CF2. Figure 4 shows that the transcription factor could bind to the probe, and with increasing concentration of cold competitor oligos, the intensity of the bound complex was weaker and weaker (lanes 3 and 4).

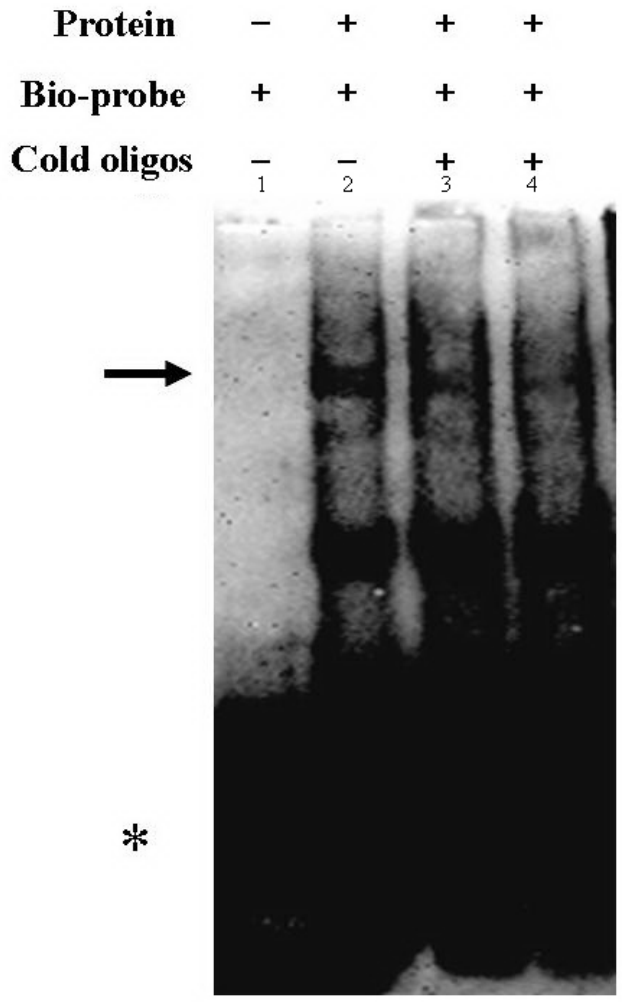

Figure 4. Identification of the site of CF2-II binding in 143-bp sequence. Combination of the probe with the nuclear extracts revealed the formation of a bound complex (arrow), which had significantly less mobility compared to the free probe (asterisk). There were only bio-probes in *lane 1*, no proteins and no cold oligos. The CF2-II could bind tightly with the probe (*lane 2*), and in the presence of competitor oligos, a strong diminution in intensity of the complex was observed with cold CF2 oligos at different concentrations (*lanes 3* and *4*, 50X and 100X, respectively).

CF2 is a Cys_2-His_2 zinc finger protein defined initially by virtue of its sequence-specific binding to the promoter of the chorion (eggshell) protein gene *s15* (Shea et al., 1990). It has been reported that CF2 can execute its function as a mediator of EGF-R-activated D/V patterning in *Drosophila* oogenesis (Hsu et al., 1996). Bagni et al. (2002) demonstrated that CF2 was also expressed in the developing muscles of the embryo and as a myogenic marker downstream of MEF2. Moreover, the interaction of MEF2 and CF2 can activate the enhancers of a number of muscle-specific genes and further control cell fate, which indicates the critical role of CF2 in the myogenic program (Tanaka et al., 2008). Therefore, the tandem repeats in the 5'UTR of the 4414-bp element in the mutant flies could interact with transcription factor CF2 and have an impact on the myogenesis pathway, which caused DVM I to be defective (data not shown).

Does infective Pifo_I exist among the *D. melanogaster* population? The 4414-bp element, degenerate type of Pifo_I element, was acquired in a mutant of *D. melanogaster* and may result in the mutant phenotype. Because Pifo_I is a retrotransposon, we began to think about the possibility that Pifo_I exists in other sites of the *D. melanogaster* genome.

To determine whether complete Pifo_I sequences are widespread present in *D. melanogaster*, PCR analysis was carried out on the genomic DNA of a wide-type strain of *D. melanogaster*. Specific oligonucleotides were designed following the conserved sequence of the 4414-bp element. The primers amplified the region including upstream sequence of *gag*, *gag*, *pol*, and part of the *env* sequence. PCR-positive bands were obtained, but shorter than expected size. Via sequencing and BLASTing, a 3754-bp element was acquired. Comparing Pifo_I, 4414-bp element (*DmMu*) and 3754-bp element (*DmWT*), we found that, like the 4414-bp element, the 3754-bp element was also similar to Pifo_I and had a pair of LTRs and portions of *gag* and *env*, and that the *pol* gene was also completely lost (Figure 5C).

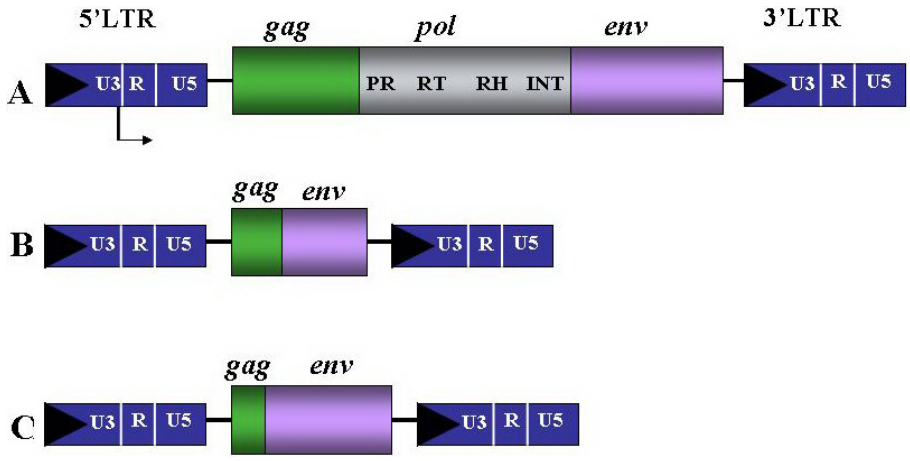

Figure 5. Schematic structures of Pifo_I, 3754-bp element (*DmWT*) and 4414-bp element (*DmMu*). **A.** The intact structure of Pifo_I, containing a pair of long terminal repeats (LTRs) and three open reading frames (ORFs) coding for proteins homologous to GAG, POL and ENV in retroviruses. ORF2 also included putative conserved domains homologous to PR, RT, RH, and INT in retroviruses and in the same order. **B.** and **C.** The 3754-bp element (*DmWT*) and 4414-bp element (*DmMu*), respectively. Both elements consisted of two LTRs and portions of *gag* and *env* genes, but the *pol* gene was entirely lost. They are degenerate types of Pifo_I.

The products of ORF1 and ORF2 are necessary for both intracellular and extracellular movements of retrotransposons. If the three elements are transpositionally active, the essential motifs of these proteins must be conserved, such as the putative domains PR, RT, RH, and INT in ORF2, which are homologous to the ones present in the *pol* gene of retroviruses and in the same order. In spite of the rapid evolution and high variability of ORF3, it has been established that the *gypsy* ORF3 product acts as a true envelope protein (Pelisson et al., 1994). Our results indicate that both the 4414-bp element (*DmMu*) and 3754-bp element (*DmWT*) consist of a pair of LTRs and part of the *gag* and *env* sequences, but that the *pol* gene is entirely lost (Figure 5B and C). The capacity of transposition in retrotransposons, in some respects, depends on whether the retrotransposons have a functional ORF3 product. Although the ORF3 sequence of the element cloned in *D. melanogaster* was highly similar to the corresponding portion of Pifo_I, the ORF3 product was not intact and lacked the signal peptide. Therefore, the two elements were deficient in critical sequences of transposition and had no infection capability

What is the relationship between the element in *D. melanogater* and Pifo_I in *D. yakuba*? As mentioned above, the 3754-bp element (or 4414-bp element) share high similarity in a pair of LTRs, nucleotide sequence (although the *pol* gene was completely lost) and structural characteristics, which indicates that the sequences are evolutionarily closer. It has been reported that *D. melanogaster* is a cosmopolitan species native to tropical Africa and that they colonized the rest of the world only relatively recently (Lachaise et al., 1988). Despite that *D. yakuba* is found across the tropical African mainland and nearby major islands, and is a close relative of *D. melanogaster*, the two species share a common ancestor of 12.8 million years ago (Stephan and Li, 2007). Therefore, these elements maybe derive from a common ancestral transposon or their similarity is due to a more recent horizontal transfer between the two species of an ancestral element that left a vestige in *D. melanogaster*. In the course of searching for the "censor" homologous portions in GIRI, we also matched the 4414-bp element to the genome of *Autographa californica*, which could provide an evidence for the horizontal transfer of an ancestral element. Horizontal transfer requires special conditions to be successful. A hypothesis was proposed that this process could be managed by exploiting vectors including those of animal parasites, such as wasps and nematode worms, or microorganisms like fungi or bacteria or viruses themselves (Heath et al., 1999). We know that the *gypsy*/*Ty3* class is the closest known group of retrotransposons to retroviruses (Springer and Britten, 1993), and several studies have provided evidence of the horizontal transfer of *gypsy* (Kotnova et al., 2007). Thus, Pifo_I should have the potential capacity of horizontal transfer in evolutionary history.

Phylogenetic analysis

We obtained phylogenetic trees based on the nucleotide sequences of the *gypsy*/*Ty3* group elements using the NJ and MP methods in order to clarify the relationship between the 4414-bp element (or 3754-bp element or Pifo_I) and the other members in this class. As outgroups, we selected HIV-1 and SIV (Table 1). Figure 6A shows the phylogenetic tree obtained with the NJ method. The topology of the trees obtained with the NJ and MP methods was similar, which indicated that the 4414- and the 3754-bp elements clustered with Pifo_I and they together clustered tightly with *ZAM*. Figure 6B shows the phylogenetic tree based on amino acid sequences of ORF2. The phylogenetic trees obtained for amino acid sequence of ORF1 and ORF2 showed the same results (data not shown). Besides, the similarity of the

ENV, GAG and POL proteins of the two elements were 46.56, 50.82 and 61.26%, respectively. These data indicate that Pifo_I could have a function similar to that of *ZAM*. It has been reported that HP1 protein is able to strongly bind to the 5'UTR of *ZAM in vitro* through the presence of a nucleation site, which could induce heterochromatization and thus element silencing (Minervini et al., 2007). As a matter of fact, the *gypsy* retrotransposon can also take advantage of endogenous protein Su(Hw), which binds to 5'UTR to further block enhancer-promoter communication (Wei and Brennan, 2001). These research results demonstrated that *ZAM* and *gypsy*, like insulators, performed their function by interaction with specific protein to silence elements. Savitskaya et al. (2006) concluded that these insulators could interact with each other over considerable distances, across inserted between enhancers or promoters and coding sequences, whereby enhancer blocking may be attenuated, cancelled, or restored. It has been demonstrated that the 4414-bp element was inserted upstream in the negative regulatory region of the *twist* gene in a mutant of *D. melanogaster*. Meanwhile, there were tandem repeats in the 5'UTR of the element that interacted with CF2-II protein. Therefore, a hypothesis was proposed that the 4414-bp element in the mutant functioned like an insulator that could silence or reduce the expressions of some genes. In the myogenesis pathway, we found that the expression of *Mhc* and *Tn* were reduced in the mutant flies and that the two genes were directly regulated by CF2 gene (Tanaka et al., 2008). We believe that further research on the 4414-bp element will disclose the mechanism of outheld wing in our mutant and the property of retrotransposon elements with a possible role in genome architecture.

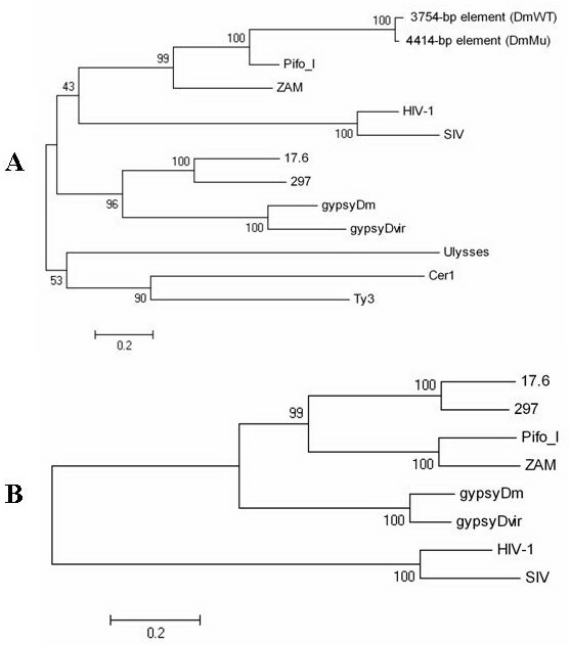

Figure 6. Phylogenetic trees obtained with the neighbor-joining method for nucleotide sequences (A) and amino acid sequences of POL protein (B) of different retrotransposons. HIV-1 and SIV sequences were used as outgroups. HIV-1 = human deficiency virus type 1; SIV = simian immunodeficiency virus.

CONCLUSION

In this paper, we depicted some characteristics of a 4414-bp element cloned from a mutant of *Drosophila melanogaster*. We confirmed that there is no intact sequence of Pifo_I in *D. melanogaster* and that the 4414-bp element is a degenerate type of Pifo_I, which belongs to the *gypsy*/*Ty3* subfamily. Besides, a 714-bp region structured in 5.0 tandem repeats of 143 bp each was found in the 5'UTR of the degenerate element. Based on the characteristics of nucleotide sequence and the phylogenetic analysis, we infer that Pifo_I, like *ZAM* retrotransposon acting as an insulator, could interact with some protein (transcription factor CF2) to execute its functions. This hypothesis could explain the relationship between the element and the downstream gene and explain the mutant phenotype.

ACKNOWLEDGMENTS

Research supported by the Major State Basic Research Development Program of China (#2006CB102100), the National Nature Science Foundation of China (#30771535) and the National Science & Technology Pillar Program (#2008BADB2B01).

REFERENCES

Alberola TM, Bori L and de Frutos R (1997). Structural analysis of *Drosophila subobscura gypsy* elements (*gypsyDs*). *Genetica* 100: 39-48.

Bagni C, Bray S, Gogos JA, Kafatos FC, et al. (2002). The *Drosophila* zinc finger transcription factor CF2 is a myogenic marker downstream of MEF2 during muscle development. *Mech. Dev.* 117: 265-268.

Bartolome C, Bello X and Maside X (2009). Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol.* 10: R22.

Benson G (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27: 573-580.

Capy P, Vitalis R, Langin T, Higuet D, et al. (1996). Relationships between transposable elements based upon the integrase-transposase domains: is there a common ancestor? *J. Mol. Evol.* 42: 359-368.

Capy P, Bazin C, Higuet D and Langin T (1997). Dynamic and Evolution of Transposable Elements. R.G. Landes Company, Austin.

Champ PC, Maurice S, Vargason JM, Camp T, et al. (2004). Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation. *Nucleic Acids Res.* 32: 6501-6510.

Dej KJ, Gerasimova T, Corces VG and Boeke JD (1998). A hotspot for the *Drosophila gypsy* retroelement in the *ovo* locus. *Nucleic Acids Res.* 26: 4019-4025.

Ha SC, Lowenhaupt K, Rich A, Kim YG, et al. (2005). Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases. *Nature* 437: 1183-1186.

Heath BD, Butcher RD, Whitfield WG and Hubbard SF (1999). Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism. *Curr. Biol.* 9: 313-316.

Herédia F, Loreto EL and Valente VL (2004). Complex evolution of gypsy in Drosophilid species. *Mol. Biol. Evol.* 21: 1831-1842.

Hsu T, Bagni C, Sutherland JD and Kafatos FC (1996). The transcriptional factor CF2 is a mediator of EGF-R-activated dorsoventral patterning in *Drosophila* oogenesis. *Genes Dev.* 10: 1411-1421.

Inouye S, Yuki S and Saigo K (1986). Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. *Eur. J. Biochem.* 154: 417-425.

Jowett T (1986). Preparation of Nucleic Acids. In: *Drosophila*: a Practical Approach (Roberts DB, ed.). IRL Press, Washington DC, 275-286.

Kohany O, Gentles AJ, Hankus L and Jurka J (2006). Annotation, submission and screening of repetitive elements in Repbase: repbasesubmitter and censor. *BMC Bioinformatics* 7: 474.

Kotnova AP, Glukhov IA, Karpova NN, Salenko VB, et al. (2007). Evidence for recent horizontal transfer of gypsy-homologous LTR-retrotransposon gtwin into *Drosophila erecta* followed by its amplification with multiple

aberrations. *Gene* 396: 39-45.
Kumar S, Nei M, Dudley J and Tamura K (2008). MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* 9: 299-306.
Lachaise D, Cariou ML, David JR, Lemeunier F, et al. (1988). Historical Biogeography of the *Drosophila melanogaster* Species Subgroup. In: Evolutionary Biology. Vol. 22 (Hecht MK, Wallace B and Prance GT, eds.). Plenum Press, New York, 159-225.
Leblanc P, Desset S, Dastugue B and Vaury C (1997). Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J.* 16: 7521-7531.
Llorens JV, Clark JB, Martinez-Garay I, Soriano S, et al. (2008). Gypsy endogenous retrovirus maintains potential infectivity in several species of Drosophilids. *BMC Evol. Biol.* 8: 302.
Marlor RL, Parkhurst SM and Corces VG (1986). The *Drosophila melanogaste*r *gypsy* transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell Biol.* 6: 1129-1134.
Minervini CF, Marsano RM, Casieri P, Fanti L, et al. (2007). Heterochromatin protein 1 interacts with 5'UTR of transposable element ZAM in a sequence-specific fashion. *Gene* 393: 1-10.
Nisha P, Plank JL and Csink AK (2008). Analysis of chromatin structure of genes silenced by heterochromatin in trans. *Genetics* 179: 359-373.
Pelisson A, Song SU, Prud'homme N, Smith PA, et al. (1994). Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J.* 13: 4401-4411.
Saigo K, Kugimiya W, Matsuo Y, Inouye S, et al. (1984). Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature* 312: 659-661.
Sandelin A, Wasserman WW and Lenhard B (2004). ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res.* 32: W249-W252.
Savitskaya E, Melnikova L, Kostuchenko M, Kravchenko E, et al. (2006). Study of long-distance functional interactions between Su(Hw) insulators that can regulate enhancer-promoter communication in *Drosophila melanogaster*. *Mol. Cell Biol.* 26: 754-761.
Shea MJ, King DL, Conboy MJ, Mariani BD, et al. (1990). Proteins that bind to *Drosophila* chorion cis-regulatory elements: a new C2H2 zinc finger protein and a C2C2 steroid receptor-like component. *Genes Dev.* 4: 1128-1140.
Song SU, Gerasimova T, Kurkulos M, Boeke JD, et al. (1994). An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.* 8: 2046-2057.
Springer MS and Britten RJ (1993). Phylogenetic relationships of reverse transcriptase and RNase H sequences and aspects of genome structure in the gypsy group of retrotransposons. *Mol. Biol. Evol.* 10: 1370-1379.
Stephan W and Li H (2007). The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity* 98: 65-68.
Tanaka KK, Bryantsev AL and Cripps RM (2008). Myocyte enhancer factor 2 and chorion factor 2 collaborate in activation of the myogenic program in *Drosophila*. *Mol. Cell Biol.* 28: 1616-1629.
Teysset L, Burns JC, Shike H, Sullivan BL, et al. (1998). A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral gypsy envelope can infect *Drosophila* cells. *J. Virol.* 72: 853-856.
Wang Q, Zhao C, Bai L, Deng X, et al. (2008). Reduction of drosopterin content caused by a 45-nt insertion in Henna pre-mRNA of *Drosophila melanogaster*. *Sci. China C. Life Sci.* 51: 702-710.
Weber JL and May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* 44: 388-396.
Wei W and Brennan MD (2001). The gypsy insulator can act as a promoter-specific transcriptional stimulator. *Mol. Cell Biol.* 21: 7714-7720.
Yuki S, Inouye S, Ishimaru S and Saigo K (1986). Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. *Eur. J. Biochem.* 158: 403-410.
Zhang H, Yu H, Ren J and Qu X (2006). Reversible B/Z-DNA transition under the low salt condition and non-B-form polydApolydT selectivity by a cubane-like europium-L-aspartic acid complex. *Biophys. J.* 90: 3203-3207.