Article

The characterization of two non-LTR retrotransposons from *Sogatella furcifera* and *Nilaparvata lugens*

Jia Lin, **Mei Zhang**, **JingYi He**, **GuangHong Li**, **FangHai Wang**

State Key Laboratory for Biocontrol and Institute of Entomology, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

E-mail: lsswfh@mail.sysu.edu.cn

Received 14 November 2018

Abstract

We have cloned two retrotransposons, one named SF-RT from *Sogatella furcifera* and another named NL-Rt from *Nilaparvata lugens* genome. Both SF-RT and NL-RT are members of the *Daphne* clade, and encode two open reading frames (ORFs) required for retrotransposition.We have gotten a methylated DNA fragments screened from *S. furcifera* genomes by methylation sensitive amplified fragment length polymorphism, and displayed higher cytosine methylation level in macropterous female adults than in brachypterous female adults. The methylated DNA fragment locate in the first ORF from 21bp to 319bp in SF-RT. Semiquantitative PCR analysis indicated that the detected gene fragments of SF-RT had higher expression in brachypterous female adult than in macropterous female adult, it means that the DNA methylationcan decline the gene expression in SF-RT.

Keywords retrotransposons; *Sogatella furcifera; Daphne;* DNA methylation.

Arthropods
ISSN 2224-4255
URL: http://www.iaees.org/publications/journals/arthropods/online-version.asp
RSS: http://www.iaees.org/publications/journals/arthropods/rss.xml
E-mail: arthropods@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Long interspersed elements (LINEs) are repeat and mobile DNA sequences, which belong to non-LTR (long terminal repeats) and often encode an RT and an endonuclase. There is a large number of LINEs sequences in eukaryotic genomes (Gao et al., 2017; Ivancevic et al., 2018; Pucci et al., 2018). LINEs typically encode two ORFs required for own retrotransposition (Feng and Moran, 1996; Maringer et al., 2017).

Transposons might be through continuous replication and inserted of host genome. In manyinsects, such as fruit flies and silkworm, there are column of Transposons very large that take up in their genome (Zhang and Chen, 2009). The transposons like LINE with reverse transcriptase activitycan be inserted into the host gene and form false gene (Kajikawa, 2002). Theseprocesses is considered new seeds of genetic evolution (Brosius, 1991) and play a key role in phenotypic shape, as it might be take part in regulating downstream gene expression with cis-element (McGregor and Orgogozo, 2007; Guichard et al., 2018). Transposons usually

distribute in heterochromatin area, but also in euchromatin. In the process of long-term symbiosis with transposons, genomes has evolved a mechanism to resist the mutations caused by the mobile of transposons. Epigenetic mechanism has become the means of defense to control transposon excessive movement. Theexpressionof transposonmight be reduced through its own methylation (Henikoff and Matzke, 1997). Phylogenetic analyses of non-LTR elements are restricted to the RT domain, the only domain common to all elements (Vladimir and Kapitonov, 2009).

Here we reportedthe Characterizations of two retrotransposons from *Nilaparvata lugens* and Sogatella furcifera. We also found that retrotransposon own methylation in *S. furcifera* indeed reduce itself expression.

2 Materials and Methods

2.1 Computation of non-LTR retrotransposons

We aligned with *N. lugens* genomes used the fourteen methylated DNA fragments, which were screened from *S. furcifera* genomes by methylation sensitive amplified fragment length polymorphism and displayed variable cytosine methylation patterns between macropterous female adults and brachypterous female adults (Zhou et al., 2013). It was found that a sequence was homologous to the 299 bases methylated fragment (GenBank: KF179359). The sequence will be set as seed sequence and use its upstream sequences (about 10000 bases) and downstream sequences (about 30000 bases) to predict the probable gene structure by HMM-based gene structure prediction website (http://linux1.softberry.com/berry.phtml). A satisfactory gene sequence contained the seed sequence was identified as Non-LTR retrotransposon.

2.2 Insects

N. lugens and *S. furcifera* were collected from the rice field located in the campus of South China Agricultural University, Guangzhou, China, and store at -80 ℃.

2.3 RNA isolation and systhesis of first-strand cDNA

We extracted total RNA using Trizol Kit (Invitrogen), and degraded genomic DNA using recombinant DNase I Kit (Takara) to obtain the purified RNA. The first-strand cDNA was synthesized by oligo dT primer (Takara).

2.4 Amplification of the prediction gene

Using cDNA as a template for PCR. The primers as Table 1 were designed for amplified integrated ORF regions of the prediction gene. The reaction mixture for PCR processed in total 20 μl volume, which contains 10 μM each primer, 2.5 mM dNTP and 1 units Taq DNA Polymerase (Takara). The PCR condition were 94 ℃ for 3 min, followed by 35 cycles of 94 ℃ for 30sec, 48 ℃ for 30 sec and 72 ℃ for 45 sec, the last cycle is 72 ℃ for 10 min. The products were measured by agarose gel electrophoresis and purified using Agarose Gel DNA Extraction Kit Ver.3.0 (Takara), then ligated into PMD18-T Vector(Takara) and transformed into DH5αcell. Monoclonal cells were picked and sequenced.

2.5 Alignment and analysis of sequences

The amplified sequences were aligned by software named DNASTAR. Online software service ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) could forecast the region of expression, Using Blast tools of National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/) searched for homologous sequence and conserved domain and then a phylogentic tree was constructed by MEGA 5.5 software(Kumar & Nei, 2008).

2.6 Semiquantitative PCR

To investigatethe methylated fragment whether there are some effect on the expression of its downstream exons, we isolated total RNA from macropterous and brachypterous female adult of *S. furcifera* independently. The same concentration RNA in different samples was used to synthesize the first-strand cDNA by oligo dT primer (Takara). The primers for PCR were listed in Table1. PCR cycling condition were as follows: 94 ℃ for

3 min, and followed by 35 cycles of 94 ℃ for 30 sec, 48 ℃ for 30 sec and 72 ℃ for 45 sec, the last cycle is 72 ℃ for 10 min. Products of amplification were electrophoresed using 1% agarose gel.

Table 1 Primers used for PCR. Primers were designed according to the prediction gene from the genome of BPH. The amplified regions were calculated based on the gene structure from predictive soft. The prediction gene from *N. lugens* were amplified using the primers except 'e-1, e-2' and 'g-1, g-2', while in the process of amplifying the prediction gene from *S. furcifera* the same primers were used except the 'E-1, E-2' and 'G-1, G-2' replaced by 'e-1, e-2' and 'g-1, g-2'.

Primer	Orientation	Sequence(5'-3')	Amplified region (bp)	Tm(℃)	GC%
5.1	Sense	ATGTCGTGTGGTATGTGC	1-216	48.5	50
5.2	Antisense	TTGTCCCTCTTTGATTGG		50.6	44.4
A-1	Sense	TGGATACGCCAATCAAAGAG	191-680	56.1	45
A-2	Antisense	CGCTTTAACAGCTCCTCCTT		56.8	50
B-1	Sense	ATCGATAAGGAGGAGCTGTT	655-1332	53.2	45
B-2	Antisense	TCTATACACAGCGATAAGGC		50.7	45
C-1	Sense	AAGCCTTATCGCTGTGTA	1311-1789	48.5	44.4
C-2	Antisense	CGTCTGAAAGATCGAGGA		50.5	50
D-1	Sense	TAACTGGAAATGTGTCCTCG	1758-2174	53.2	45.0
D-2	Antisense	TCTCCATTTGTTGCTCTCAT		52.6	40.0
E-1	Sense	ATAACAGGAGAGAGCCAGTT	2129-2583	50.5	45.0
E-2	Antisense	AAGAGCGATAGGTCGGTAGT		53.4	50.0
e-1*	Sense	GGAAAGTAGTGAATACGCTA	1877−2370	47.6	40.0
e-2*	Antisense	CACTGAGAGAAGAGCGATAG		49.6	50.0
F-1	Sense	AACAACTACCGACCTATCGC	2560-3049	54.6	50.0
F-1	Antisense	CAAAGGCTGTGACAGATCCC		57.6	55.0
G-1	Sense	TAACGGCAACTTTGAGGGAT	3015-3498	57.4	45.0
G-2	Antisense	ATAAGCTCCTCCCCATACCG		58.4	55.0
g-1*	Sense	AACTTCGAGGGATCTGTCAC	3022-3485	53.6	50.0
g-2*	Antisense	CAAACCGCTATTCCATATTG		53.7	40.0
H-1	Sense	TCGTAAACAGTAGACTACAATATGG	3449-3900	53.9	36.0
H-2	Antisense	TTATCTGAGATCTCTAAACCACACT		54.5	36.0

3 Results

3.1 The structure of prediction gene contained the seed sequence

The seed sequence from *S. furcifera* genome has 95% identities with *N. Lugens* and sits at the first exon in the prediction gene which having two exons (3678bp), 128bp 5'-UTR and 577bp 3'-UTR. The structures of prediction gene is shown in Fig. 1.

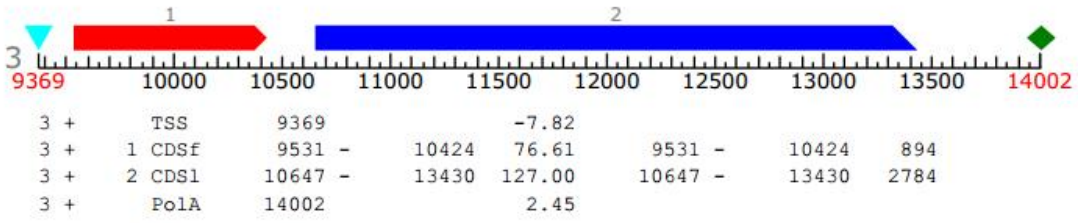


Fig. 1 The structure of prediction gene contained the seed sequence in *N. lugens* genome. The seed sequence sits in the red region. TSS, CDSf, CDSl and PolA represent the transcriptional initiation site, the first ORF, the second ORF and the transcription termination.

3.2 Sequence of non-LTR retrotransposonsin two species

The non-LTR retrotransposon ORF sequence obtained by cloning is 3900bp in *N. lugens* (NL-RT) (Fig. 2), 3904bp in *S. furcifera* (SF-RT) (Fig. 3).

ATGTCGTGTGGTATGTGCAACAAATCGGTGAGAGGTGCAAACAACTCTGAAGTGAAGTGCATAG
ACTGTAACAATCAGTTCCATGGAAACTGCGTGAGTATGAAAGTAGAGGAAATCAAATTTCTAATT
GAAAGTGGTAAATCGTGGAGATGTGATGGATGCACCAGAAACAAGAGACTCAGCATGTCTATGG
ATACGCCAATCAAAGAGGGACAAATCACCTTGGAGAAGTTAGCGCAAATGCTCACCAATGTGTC
AGAGAACATTAAAAGTGTGGAGAAGAGCTTAGGAGATTCTATACAATCGTGTCACGAATCTGTC
GGTGATGTACTCGAGAAAGTGAAAATACAAGAAAGTCAACTCAGGGTGTGCCTGGATAAAATTG
AGACACAATCTACCGAAATAATAGCACTAAAAAAGGAAAATGAAGAACTGCGTCGTGCCATCTC
TGATATTCAGCAATATACAAGATCGAACTGTCTGGAGCTCCATAATTTTCCTCAGGAGGAAAATG
AGGATCTTCTTGGTGTTGTGAAATCTGTTAGCAAAGCTTTGGGACATACAATAACCGACCTGCAA
ATAGACAATTGTCACCGTCTTCCAACTCGTGAAAAAGATAAAGTGCCACCAATTATAATAAAACT
GACTAGAAGGATCGATAAGGAGGAGCTGTTAAAGCGAAGGAGGGTGATGAGGAACTTTTCAACG
CGACACATGGACCTACCAACAGACATCCCGCTGTACCTCAACGAGAGTCTGTCGCCGGAGAACA
GGAAGGTGCTAGCGCTGGCGAGAGCAGCCAAGAAAGAAAAGGACTACAAGTATCTCTGGATTAG
GAACGGAAAAATTTTGATGCGAAAATCTGAGGGTCAACCGGTAATCTCATTGAGTTCAGTAAGT
GATATAAGCAAACTATAATGATTACAAATGTACTAGTTTATTTAATAAAAGTTTACTAATTTTTGA
AAAATCCACGAAAGTTAAAAATAAATGACCGACCTTAATGAACTATTCTGTTTAGTACAAAACAT
AAGAAGCTTAAGAGAAAACTTTGACCTTTTTTTAATTGAGCTTGAAAGTTATGCAAAAAAACCGC
AATGTATAATTTTAACAGAGATTTGGATATATAGTGATGAGTCTGAACTTTACCCAATAGATGGA
TACAATTGTTTTACAAATTGCAATGACAATTATAGAGCAGGTGGAGTAGCGGTTTATGTGAATAA
AGAAATAACAGCAATAGCCGATAAGATAGTTCTAATCTCTGTAGATGCCTTAAAAGTTGACTTTA
TTTTTGGTAATGTCGAAATAAGCCTTATCGCTGTGTATAGATTTCACAATATACCTGTAGATGAGT
TCTTCAAAGAGATAAAAACTGTATTAGATAAGTTAGATAATGATAATTCTATAATAACAGGTGAC
ACAAATATTGATCTGCTTGATAGCAATCGAGTATCAGATGAATACCAACTCATTATGTCAAGTTA
TGGGTTTACATCTTATGTTAACGAACCGACTAGAGTAACCAACATTTCAACTTCATGCATCGATC
ACTTATGGTTTAGATACGTAACAACAAGGAATTCTTTCAATCCCTCCACAATTATAAAGAATTTA
CTAATCACTGATCATTATGCTATAGAGTTTTATTTTGAGAAAATGATTAAGAAGCGAATAAATCA
TTCACCCAATACTAATCACAATAAAAAGCAGAAAAAAATAAACGTTGAAATTCTCGCCAATAAA
TTAAATAAAATTAACTGGAAATGTGTCCTCGATCTTTCAGACGTAGATTTAGCTCTGGATAAATTT
CTAGAAATTTTCAGTGAGTGTTTGGAAATAGCATGTACAGGAGAAAGTAATAAAAGGAAAATAT
TTACACAACCACCACATAAACCATGGCTCACTCCCTACTTGCTAAACCGAATAACAGTTAAAAAT
CGCTTATACAAAAAAACCACAAAACAACCATATAACGAGAACTTGCTAAATTATTATGAAAGGT
ACAGTGATAAACTCAAGAAAGAAATACAGGAATCCAAAAATAAATATTTTCGCGAGACTCTAAA
CAATTTGAGGGGAAACTCTAAGGAGACTTGGAAAGTAGTGAATACGCTATTAGGAGAAAATAAC
AGGAGAGAGCCAGTTAAAAAAATGAGAGCAACAAATGGAGAGACTCTCCTAAACGAGAATCTA
ATTGCAAATGAGCTTAATAATCACTTTATCAGTATATACAGCGCTAAGAATACAAAAAAGCTAAA
TATCAACGATTATAACAGTTATAAAATGCTTTTTGACAAACCAATTCACCATTACCATTCAATGTT
TTTCACACCAATAACTAATTTAGAGATAGAAGGAATTGTACTCTCAATGAAATCCAGAAAATCTC
CAGGCTATGATAATATAAGAATTGAGCTAATTAAAAATGTAATAAAATCGATCTCTAGCACCCTT
GCTCACATATATAACCTCAGTTTGGAGACTGGTGTGTATCCAAAAAAATTAAAAAATGCAATTGT
AGTACCAATTTTCAAATCTGGGGATAAAGAAAATCCAAACAACTACCGACCTATCGCTCTTCTCT

CAGTATTTGCAAAAATCCTAGAGAAAGTTGTCAGGATAAGATTAGTAAGCTTTCTAACCAAACAC
AGTTTTTTCAGCAAAAATCAATTCGGGTTTCAAAAAGGACTAAGTACAGAGGACGCCATGCTTAA
ATTTATCTCTGAAATATATAATGGAATGAACAATAATAAGAAATGTGCCGGCCTCTATTTGGATA
TCCGAAAAGCGTATGATACTGTGAATCACGATATATTGTTGGGAAAATTGCAAGACTCAGGAGTT
AGAGGTGTATGTAATAATTGGTTTAAAAGCTTTCTAAGTAATAGGTCACAACAGGTAAGGATTGG
GGATTCGCTAAGCGAGCTAAAACTTATAGATACAGGGGTAGGTCTGCCTCAGGGGTCCGGTCTAT
CAGCTGAGCTGTTCCTTGTGTATGTTAATGATCTTTGTAACGGCAACTTTGAGGGATCTGTCACAG
CCTTTGCCGACGACACAGCACTGAGTTACAGTGCAGATGATAGGGGTCAGTTGGCTCAGATGATT
AGTGAAGATTTAAAGAAATTGAATTTATGGCTGCAGGTGAACGCCCTAGAATTGAATGCCAGTA
AATCCCACATAATTGTCCACAAGCTGAGGCCTGAAGGAAATGATTTAATGAATATATCATTTCAT
TCAAATGAATGTGATAGTCCAATAAACTGCACCTGTGAAAAGATTTCAGAAGCACCCCAAGTTA
AATATCTAGGTATTATAATTGATTCCAAGCTTACTTGGAACCAACAAATAATTAAATTGAAGAGG
GAACTTACGTGTGTATGCAGAAAATTTTACTATATAAGAAATTTATGCCCCGAATATGTCATGGA
ATCGCTGTACTATGCACTCGTAAACAGTAGACTACAATATGGAATAGCGGTATGGGGAGGAGCT
TATTTTAATAACATTAAACCTCTTGTCACCGGGCAAAAATACATATTGAGGACCATAGACAAAAA
ACCAAGATTATTTAGCTCTTTGCCAATTTTCAGAAAATGGGGATTTCTACCACTAAGGTATCTGTA
TTTCTTCAAAGTATTGTCAATTTTCTTTTTTAGAAGTGGACAAGTGAACGTTGCTAACCGCGAATA
TTATCTCCGATCCGCAAGTAATTTAACTAGACCAAGACCACACAAAGAAATCTTTAAACGTTTTT
ATTTATTTATAGCCCCAAAAGTATACAATGAAATTCCAATTAGCATAAGGCAGCAAAAAAATCCG
AGAAGTTTCAAGTTTTTTCTGAAGAATTGGTTATTGGAAAAGCAGGATGTTGAAGTGTGGTTTAG
AGATCTCAGATAA

Fig. 2 The non-LTR retrotransposon ORF sequence in *N. lugens.*

ATGTCGTGTGGTATGTGCAACAAGTCGGTGAGAGGAGCAAACAACGCTGAAGTGAAGTGCATAG
ACTGTAACAACCAGTTCCATGGAAACTGCGTGAATATGAAACTCGAGGAAATCAAATTTCTAATT
GAAAGTGGTAAATCGTGGAGATGCGACGGATGCACCAGAGACAAGAGACTTAACTTGTCAATGG
ATACGCCAATCAAAGAGGGACAAATCACCTTGGAGCAGTTGGCTCAAATGCTCACGAATGTGTC
AGAGAACATTAAATAGGGTGGAGAAAAGCTTGGGAGATTCTATACAATCATGTCATGAATCCGT
CGAAGATGTACTCCAGAAAGTGAAAAAGCAAGAAAATCAACTCAAGGAGTGCCTTGACAAAATT
GAGTCTCAATCTACCGAAATAATAGCACTGAAGAAGGAAAACGAAGAACTACGTCGTGCCGTCT
CTGAAATTCAGCAATATACAAGATCGAACTGTCTAGAGCTCCATAATTTTCCTCAAGAGGAAAAT
GAAGATCTTCTCGGTGTTGTGAAGTCTGTTAGCAAGGCTTTGGGACATACAATAACCGACCTGCA
AATAGACAATTGTCACCGTCTTCCAACTCGCGAAAAAGATAAAGTGCCACCAATTATAATAAAA
CTGACTAGAAGGATCGACAAGGAGGAGCTGTTACATCGAAGGAGGGTGATGAGGAACTTCTCAA
CGCGGCACATGGACCTACCAACAGATATCCCGCTGTACCTCAACGAGAGTCTGTCGCCGGAGAA
CCGGAAGGTGCTAGCGCTGGCGAGAGCAGCCAAGAAGGAAAAGGACTACAAGTATCTCTGGATT
AGGAATGGAAAAAATTTAATGCGAAAATCTGAGGGTCAACCGGTAATCTCATTGAATGCAGTAA
GTGATATAAGCAAACTATAATGATAACAAACTTATTAGTTTATTTACTAAAAGGTTACTAATTTT
GAAAAAATCCACAAAATTTGAAACTAATTGACCGACCTTAATGAACTATTCTGCTTAGTACAAAA
TATAAGAAGCTTAAGAGAAAACTTTGACCTTTTTTTAATTGAGCTTGAAAGTTATGCAAAAAAAC
CGCTATGTATTATTTTAACAGAGATTTGGATCTATAGTGATGAGTCTGAACTTTACCCAATAGATG
GATACAATTGTTTTACAAATTGTAATGACAATTATAGAGCAGGTGGAGTAGCGGTTTATGTGAAT
AAAGAAATAACAGCAATAGCCGATAAGATAGTTCTAATCTCTGTAGATGCTTTAAAAGTTAACTT

TATTTTTGGTAATGTCGAAATAAGCCTTATCGCTGTGTATAGATTTCACAATATACCTGTAGATGA
GTTCTTCAAAGAGATAAAGACGATATTAGATAAGCTATATAATGATAATTCTATTATAACAGGTG
ACATAAATATTGATCTGCTTGATAATAATAGAGTAACAGATGAATACCAATTCATTATGTCGAGT
TATGGGTTTACATCATATGTTAACGAACCGACTAGGGTAACCAACATATCAACATCATGCATCGA
TCACCTATGGTTTAGATACGTAACAACAAGGAATTCTTTCAATCCCTCCACAATTATAAAGAATC
TACTAATCACTGATCATTATGCTATAGAGCTATATTTTGAAAAAAATATTAAGAAGCGTATAAAT
CATTCACCCAATTCTAATCCCAGCAAAAAGCAGAAAAAAATATACGTTGAAATTCTCGCCAATAA
ATTACAACATGGAGTTGACTGGAAATGTATCCTCGATATTCCAGACGTAGATTTGGCTCTGGATA
AATTTATAGAAATTTTTAGTGAGTGTTTGGAAGAAGCATGTACAGGAGAAAGTAATCCAAGGAA
AATATTTATACAACCACCACATAAACCATGGCTCACTCCATACCTACTAAACCGAATAACAGTTA
AAAATCGCTTATACAAAAAAACTACAAAACAACAGTACAATGAAAACTTACTAAATTATTATAA
GAGGTATCGGGATAAACTTAAGAAAGACATACAGGAATCCAAAAATAGATATTTTCGAGAAACT
CTAGATAATTTGAGGGGGGATTCTAAGGAGACTTGGAAAGTAGTGAATACGCTATTGGGAGAAA
ATAACAGGAGGGAAACAGTCAAAAAAATTAAAGCAACAAATGGRGAGACTCTCCTAAACGAGA
ATCTGATTGCAAATGAACTCAATAATCACTTTATTAGAATATATAGCGCTAAGAATACAAGAAAA
CTAAACAGTAACGATTATAATAGTTATAAAACACTTTTCCACAAACCAGTTCAGCATCACTATTC
AATGTTTTTTACACCAATAACTAATTCAGAAATAGAAGAAATTGTACACTCAATGAAATCCAGAA
AATCTCCAGGTTATGATAATATAAGAATTGAGCTAATTAAAAATGTAATAAAATCGTTCTCCAGC
ACCCTTGCTCACATATATAACCTCAGTTTGGAGACTGGTGTGTATCCAAAAAAATTAAAAAATGC
CATAGTAGTGCCAATTTTCAAATCTGGGGATAAAGAAAATCCAAACAACTACCGACCTATCGCTC
TTCTCTCAGTGTTTGCGAAAATCCTAGAAAAAATTGTCAGGATAAGATTAGTAAGTTTTCTAACC
AAACACAGTTTTTTCAGCAAAAATCAGTTCGGGTTTCAAAAAGGACTAAGTACAGAGGACGCCA
TGCTTAAATTTATCTCTGACATATATAATGGAATCAATAATAATAAGAAATGTGCTGGCCTTTATC
TTGATATCCGTAAAGCGTATGATACTGTAAATCATGATATATTATTGGGAAAATTACAAGATGCA
GGAGTAAGAGGTGTATGTAATAATTGGTTTCGAAGCTTTCTAAGTAATAGGTCACAACAGGTAAG
GGTTGGGGATTCGCTAAGCGAGCTAAAACATATAGATACAGGGATAGGTCTGCCTCAGGGGTCC
GGACTATCAGCGGAACTGTTTCTTATCTATGTTAATGATCTATGTAATGGTAACTTCGAGGGATCT
GTCACAGCCTTTGCCGACGACACAGCACTGAGTTACGGTGCAGAAGATAGGGGTCAGTTGGCTC
AGATGATTGGTGAAGATTTAAAAAAATTGAATTTATGGCTGCAGGTAAACGCTCTAGAATTAAAT
GCCAATAAATCTCACATAATTGTCCACAAGTTGAGGCCTGAGGGAAATGATTTAATGAATATAAC
ATTCCATTCAAATGAGTGTAATAGTCCTATCAACTGTAGTTGTGAAAAGATTTCAGAAGCACCCC
AAGTAAAATATCTAGGTATTATAATTGATTCCAAGCTTTCTTGGAATCAACAGATACTTAAATTG
AAGAGAGAACTTACTTACGTGTGCAGAAAATTATACTATATAAGAAGTTTATGCCCCGAATATGT
TATGAAATCGATGTACTATGCACTCGTACACAGTAGATTACAATATGGAATAGCGGTTTGGGGAG
GAGCTTATTTAAATAACATTAAACCTCTTGTCACAGGGCAAAAATACATATTGAGGACCTTAGAC
AAAAAACCAAGATTATTTAGCTCTTTGCCAATTTTCAGAAAATGGGGATTCCTACCACTAAGGTA
TCTGTATTTCTTCAAAGTATTGTCAATTTTCTTTTTTAGAAGTGAGCAAGTGAACGTAGTTAACCG
CGAATATTATCTCCGATCCGCAAGCAATGTAGCCAGACCAAGACCACACAAAGAAATTTTTAAA
CGTTTTTATCTATTTATAGCCCCAAAAGTATACAATGAAATTCCAAATAGTATAAGGCAGCAAAG
AAATCCGAGAAGTTTTAAGTTTCTTCTGAAGAATTGGCTATTAGAAAAGCAGGATGTTGAAGTGT
GGTTTAGAGATCTCAGATAA

Fig. 3 The non-LTR retrotransposon ORF sequence in *S. furcifera*. Horizontal bar markers in the sequence indicated the methylated DNA fragments screened from *S. furcifera* genomes by methylation sensitive amplified fragment length polymorphism.

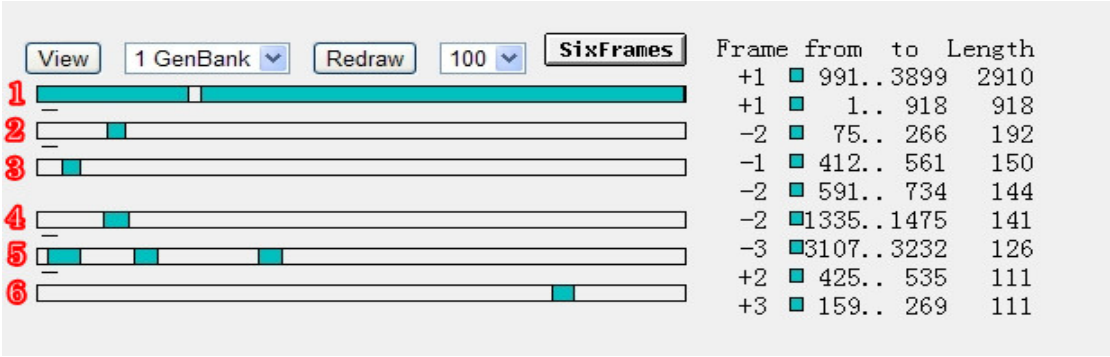
The structure of non-LTR retrotransposon in *N. Lugens* contains two non-overlapping ORFs and a region of non-coding DNA (Fig. 4 A). The non-coding DNA is 72bp long. The first ORF is 918bp long and encodes 305 amino acids. The deduced translated amino acid sequence may be a nucleic acid binding protein having an analogous PHD conserved motif.The second ORF is 2910bp long and encodes 969 amino acids.The deduced translated amino acid sequence is much similar with reverse transcriptase which shows specific motifs conserved with the endonuclease region and reverse transcriptase domain (Fig. 5 A).

The structure of non-LTR retrotransposon in *S. furcifera* also contains two non-overlapping ORFs and a region of non-coding DNA (Fig. 4 B). The non-coding DNA is 558bp long. The first ORF is 669bp long and encodes 222 amino acids, which may be a signal peptide. The second ORF is 2427bp long and encodes 808 amino acids. The deduced translated amino acid sequence is much similar with reverse transcriptase (Fig. 5 B).

3.3 Phylogenic construction

The initial BLAST searches indicated affiliation of NL-RT and SF-RT with the Daphne clades (http://www.girinst.org/RTphylogeny/RTclass1/). Phylogentic analysis of RT domain sequences was shown in Fig. 6 by MEGA 5.5 software.

(A)



(B)

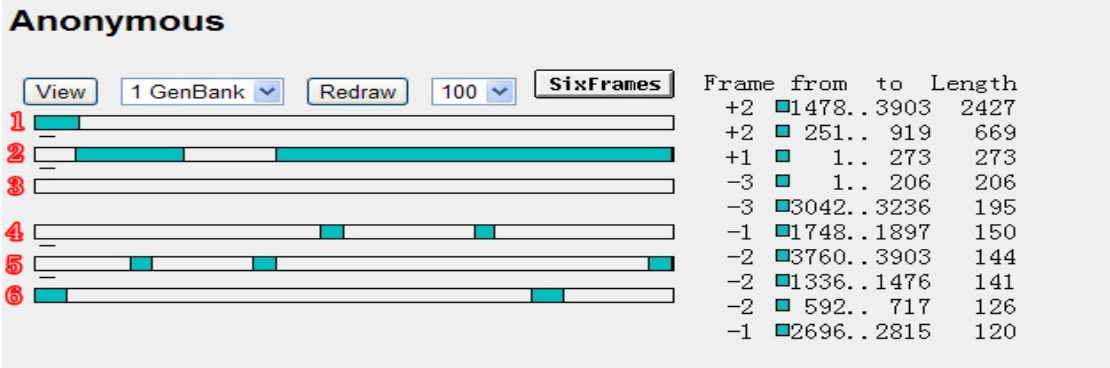
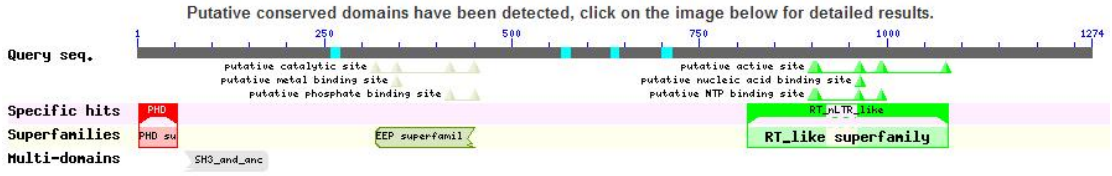


Fig. 4 (A) The ORF prediction sequence cloned from *N. lugens*. (B) The ORF predictionsequence cloned from *S. furcifera*. There were six possible open read frames and we chose the first predicted structure because it wassimilar with the beginning prediction in *N.* genome. The second predicted structure was chose in *S. furcifera* because it was similar with the predicted structure in *N. lugen*. The prediction from left to right is consisted with the sequence from 5' to 3'. The blue color marked that the sequence in the regions could encode amino acid. The blank regions marked those non-coding DNA.

(A)



(B)

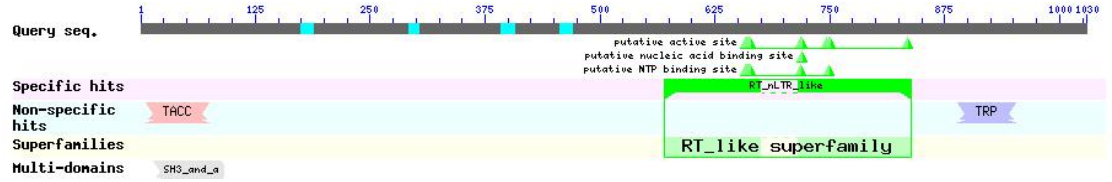


Fig. 5 (A) Putative conserved domains for the non-LTR retrotransposon in *N. lugens.* (B) Putative conserved domains for the non-LTR retrotransposon in *S. furcifera*. Different color represented those conserved domain respectively. The red marked represent PHD domain and the green marked indicated RT and EN domain.

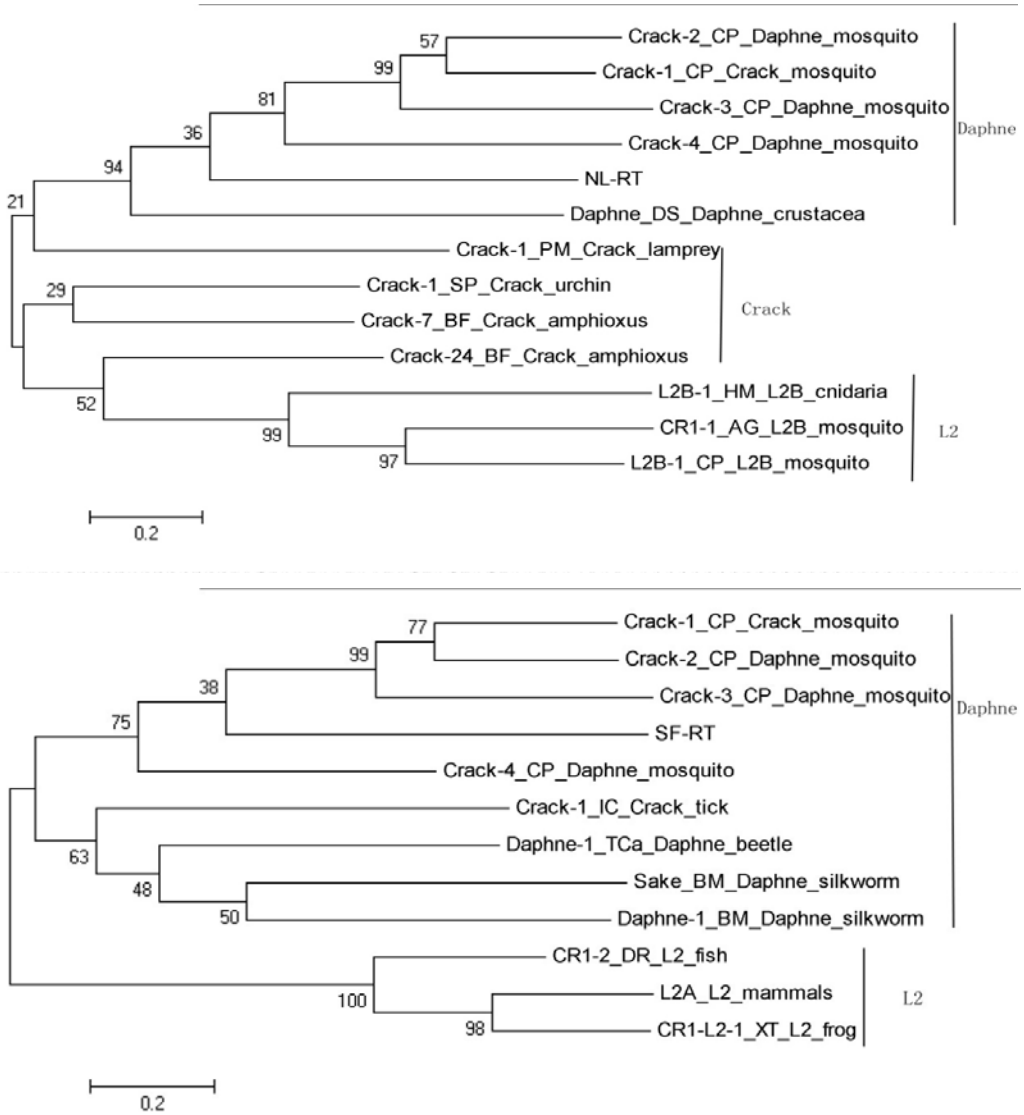


Fig. 6 Phylogentic analysis of RT domain sequencesfromNL-RT and SF-RT.

3.4 Semiquantitative PCR analysis

The seed sequence is a methylated DNA fragments screened from *S. furcifera* genomes by methylation sensitive amplified fragment length polymorphism, and displayed higher cytosine methylation level in macropterous female adults than in brachypterous female adults. The methylated DNA segment located in the first ORF from 21bp to 319bp in SF-RT. Semiquantitative PCR analysis indicated that the detected gene fragments had high expression in brachypterous female adult than inmacropterousfemale adult (Fig. 7), it means that the methylated DNA segment can decline the gene expression.

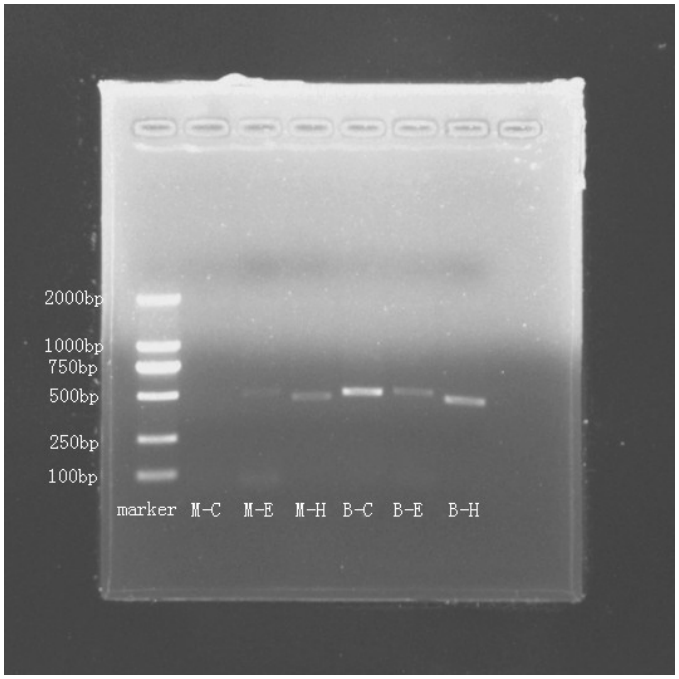


Fig. 7 The analysis of the gene expression pattern of SF-RT in *S. furcifera*. The levels of mRNA were normalized separately for each sample. 'M-C, M-E, M-H' represent that the primers (C-1/C-2, e-1/e-2, H-1/H-2) listed in Table 1 used as a probe to amplified gene fragments in macropterous female adult and 'B-C, B-E, B-H' represent that the detective samples come from brachypterous female adult using the same primers.

4 Discussion

This study provides the first insights into diversity, structural organization, and phylogeny of retrotransposable elements in the poorly explored family Delphacidae. Interestingly, we were able to identify a clade of non-LTR retrotransposons named Daphne, Jockey group, which currently includes representatives from crustaceans, insects, and echinoderms (Schonand Arkhipova, 2006).

TheSF-RT element is 3904bp long and encodes two ORFs (Fig. 4 B).The first ORF is homologous with SH3 domain protein.Therefore,it might be as a signal peptide located at 49-82 amino acid residue. The second ORF has RT domains located at 642-838 amino acid residue. A methylated DNA fragment screened from *S. furcifera* genomes by methylation sensitive amplified fragment length polymorphism, and displayed higher cytosine methylation level in macropterous female adults than in brachypterous female adults. The methylated DNA segment located in the first ORF from 21bp to 319bp in SF-RT. The detected gene fragments of SF-RT had higher expression in brachypterous female adult than in macropterous female adult by semiquantitative PCR analysis (Fig. 7), it indicated that DNA methylation can decline SF-RT elements expression and make some effect on wing polymorphism of planthopper.

The consensus sequence of NL-RTis 3900bp long and encodes two ORFs.The first ORF has two domain,

one named PHD-fingerhave been identified as a nucleic acid-binding protein that binds its own RNA, and another named SH3 domain protein behind PHD-fingercould be a signal peptide. The second ORFinclude anendonuclease domain (EN) and RT motifs. The motif conserved in EN domain, it is a central reverse transcriptase (Fig. 5 A), and belongs to the large EEP (exonuclease/endonuclease/phosphatase) superfamily that contains functionally diverse enzymes that share a common catalytic mechanism of cleaving phosphodiester bonds.

A phylogenetic tree constructed using RT amino acid sequence suggests that NL-RTisrelated to the *Daphne* clade. This result is consistent with the phylogenetic stance that *Daphne* clade existed in insects (Vladimir and Kapitonov, 2009), also provides insight into the evolutionary history of *N. lugens* retrotransposons.

Acknowledgments

This research was supported by national natural science foundation of China (31171844) and natural science foundation of Guangdong province (2015A030313114).

Reference

Brosius J. 1991. Retroposons-seeds of evolution. Science, 251: 4995:753

Feng Q, Moran JV, Kazazian Jr HH, Boeke JD. 1996. Human LI retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell, 87: 905-916

Gao B, Chen W, Shen D, et al. 2017. Characterization of autonomous families of Tc1/mariner transposons in neoteleost genomes. Marine Genomics, 34: 67-77

Guichard E, Peona V, Tagliazucchi GM, et al. 2018.Impact of non-LTR retrotransposons in the differentiation and evolution of anatomically modern humans. Mobile DNA, 9(28): 1-19

Henikoff S, Matzke M. 1997. Exploring and explaining epigenetic effects. Trends in Genetics, 13(8): 293-295

Ivancevic AM, Kortschak RD, Bertozzi T, et al. 2016. LINEs between species: Evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. Genome Biology and Evolution, 8(11): 3301-3322

Ivancevic AM, Kortschak RD, Bertozzi T, et al. 2018. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. Genome Biology, 19(85): 1-13

Kajikawa M. 2002. LINEs mobilize SINEs in the eel through a shared 3'sequence. Cell, 111(3): 433-444

Kapitonov VV, Tempel S, Jurka J. 2009. Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. Gene, 448(2): 207-213

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform, 9: 299-306

Maringer K, Yousuf A, Heesom KJ, et al. 2017. Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in Aedes aegypti. BMC Genomics, 18(1010): 1-18

McGregor AP, Orgogozo V, Dolon I, et al. 2007. Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature, 448(7153): 587-590

Pucci MB, Nogaroto V, Moreira-Filho O, et al. 2018. Dispersion of transposable elements and multigene families: Microstructural variation in Characidium(Characiformes: Crenuchidae) genomes. Genetics and Molecular Biology, 41(3): 585-592

Schon I, Arkhipova IR. 2006. Two families of non-LRT retrotransposons, Syrinx and Daphne, from the *Darwinulid ostracod*, *Darwinula stevensoni*. Gene, 371(2): 296-307

Zhang B, Chen B. 2009. Insect transposable elements and their functions. Chinese Bulletin of Entomology, 46(2), 187-195

Zhou XS, Chen JL, Zhang M, et al. 2013. Differential DNA methylation between two wing phenotypes adults of *Sogatella furcifera*. Genesis, 51(12): 819-826