Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology) 2022 23(6):515-527
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2101090

Functional characterization of *piggyBac*-like elements from *Nilaparvata lugens* (Stål) (Hemiptera: Delphacidae)

Jun LYU, Qin SU, Jinhui LIU, Lin CHEN, Jiawei SUN, Wenqing ZHANG✉

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Abstract: *PiggyBac* is a transposable DNA element originally discovered in the cabbage looper moth (*Trichoplusia ni*). The *T. ni piggyBac* transposon can introduce exogenous fragments into a genome, constructing a transgenic organism. Nevertheless, the comprehensive analysis of endogenous *piggyBac*-like elements (*PLE*s) is important before using *piggyBac*, because they may influence the genetic stability of transgenic lines. Herein, we conducted a genome-wide analysis of *PLE*s in the brown planthopper (BPH) *Nilaparvata lugens* (Stål) (Hemiptera: Delphacidae), and identified a total of 28 *PLE* sequences. All *N. lugens piggyBac*-like elements (*NlPLE*s) were present as multiple copies in the genome of BPH. Among the identified *NlPLE*s, *NlPLE25* had the highest copy number and it was distributed on five chromosomes. The full length of *NlPLE25* consisted of terminal inverted repeats and sub-terminal inverted repeats at both terminals, as well as a single open reading frame transposase encoding 546 amino acids. Furthermore, *NlPLE25* transposase caused precise excision and transposition in cultured insect cells and also restored the original TTAA target sequence after excision. A cross-recognition between the *NlPLE25* transposon and the *piggyBac* transposon was also revealed in this study. These findings provide useful information for the construction of transgenic insect lines.

Key words: *Nilaparvata lugens*; *piggyBac*-like elements; Transposon; *NlPLE25*

1 Introduction

The *piggyBac* transposon, originally discovered in the cabbage looper moth (*Trichoplusia ni*), is 2472 bp in length and comprises three parts: a 13-bp terminal inverted repeat (TIR), a 19-bp asymmetric sub-terminal inverted repeat (STIR), and a 1782-bp transposase open reading frame (ORF) (Cary et al., 1989; Fraser et al., 1996). Its general features shared by various members of the *piggyBac* superfamily include the TTAA tandem site duplication (TSD), a triple-aspartate (DDD) motif, and a C-terminal cysteine-rich domain (CRD) in the *piggyBac* transposase core region (Sarkar et al., 2003; Keith et al., 2008). Numerous transposons similar to *piggyBac* exist in a variety of organisms, known as *piggyBac*-like elements (*PLE*s).

Previous study revealed that the silkworm (*Bombyx mori*) has 98 *PLE*s, of which five *PLE*s contain a single ORF and TIRs (Xu et al., 2006). Subsequently, *yabusame*, another *PLE* identified in the *B. mori* genome, showed the almost complete loss of its transposition activity in vitro (Daimon et al., 2010). Moreover, nine *PLE*s were found in the cotton aphid (*Aphis gossypii*), of which *AgoPLE1.1* has an intact transposase ORF (Luo et al., 2011). The *PLE-wu* was identified in Sf9 cells, which exhibits transposition activity in insect and mammalian cells (Wu and Wang, 2014). Certain *PLE*s were also reported in tobacco budworm (*Heliothis virescens*) (Wang et al., 2006), cotton bollworm (*Helicoverpa armigera*) (Sun et al., 2008), pink bollworm (*Pectinophora gossypiella*) (Wang et al., 2010), and red flour beetle (*Tribolium castaneum*) (Wang et al., 2008).

PiggyBac transposons with high transpositional efficiency are commonly used in the germline transformation of many insects, such as the fruit fly (*Drosophila melanogaster*), silkworm (*B. mori*), yellow fever mosquito (*Aedes aegypti*), honeybee (*Apis mellifera*), pink bollworm (*P. gossypiella*), and Asian corn borer (*Ostrinia furnacalis*), with success rates of 0.1%–10% (Gregory et al., 2016). The brown

✉ Wenqing ZHANG, lsszwq@mail.sysu.edu.cn
Wenqing ZHANG, https://orcid.org/0000-0002-5119-1658

Received Dec. 26, 2021; Revision accepted Mar. 3, 2022;
Crosschecked May 24, 2022

planthopper (BPH) *Nilaparvata lugens* is one of the most destructive insect pests of rice (*Oryza sativa*) in Asian countries. A major hurdle to BPH research is the lack of molecular tools required for its genetic analysis. Although the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system (Li et al., 2021) has been used for gene knock-out in BPH (Xue et al., 2018; Zhao et al., 2019; Chen et al., 2021), knock-in has not yet been reported. *PiggyBac* is a potentially valuable alternative tool to solve this challenge; however, endogenous *PLE*s might affect the stability of the *piggyBac* transposition system integrated into the genome (Jia et al., 2021). Therefore, a comprehensive analysis of endogenous *PLE*s is essential prior to the application of *piggyBac*.

Previously, two versions (GCA_000757685.1 and GCA_014356525.1) of the whole BPH genome have been released. Here, we report a genome-wide analysis of *PLE*s in BPH. Special attention was paid to *NlPLE25*, which has the highest number of copies in the genome. Our findings should contribute to the understanding of the distribution and characteristics of members of the *piggyBac* superfamily and their application in insects.

2 Materials and methods

2.1 Insect and cell cultures

A BPH population was reared on seedlings of the susceptible rice variety Huanghuazhan. The population was then maintained in a greenhouse at Sun Yat-sen University, Guangzhou, China, at (26.0±0.5) ℃ with (80±10)% relative humidity under a 16-h light/8-h dark photoperiod. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* medium (Thermo Fisher Scientific, Waltham, MA, USA) at 28 ℃. Sf9 cells were cultured in Sf-900 II serum-free medium (Thermo Fisher Scientific) at 28 ℃.

2.2 Identification of *PLE*s

The *N. lugens* assembly GCA_014356525.1 was adopted in this study. The RepeatModeler 5.8.8 (http://www.repeatmasker.org) and RepeatMasker 4.1.1 (http://www.repeatmasker.org) packages were employed to identify repeat families in BPH de novo. The interspersed repeat databases screened by RepeatMasker were based on the Repbase Update (Bao et al., 2015) and Dfam (Hubley et al., 2016). To gain more access to sequences of *PLE*s from the BPH genome, a search was carried out of the BPH nucleotide database of the National Center for Biotechnology Information (NCBI) using the key phrase "*piggyBac*-like elements." The 3-kb sections upstream and downstream of each sequence were extracted to look for TIRs (Bouallègue et al., 2017). The sequence was considered a complete *PLE* when TIRs were found on both sides. The TIR sequence logo was created using WebLogo (http://weblogo.berkeley.edu/logo.cgi). Multiple sequence alignment and phylogenetic tree construction were carried out using MEGA 6.0 (https://www.megasoftware.net). TBtools (Chen CJ et al., 2020) was used to map the location of *NlPLE*s on chromosomes. PSORT II (https://www.genscript.com/psort.html?src=leftbar) was utilized to predict the nuclear localization signal (NLS).

2.3 *NlPLE25* amplification and analysis

The genomic DNA was isolated from BPH using an Omega insect DNA extraction kit (Omega Bio-Tek, Norcross, GA, USA) and quantified using a NanoDrop ND-2000 instrument (Thermo Fisher Scientific). The primers were designed using NCBI primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast) based on the *NlPLE25* reference sequences in the NCBI database (Table S1). Phanta Max super-fidelity DNA polymerase (Vazyme Biotech, Nanjing, China) was employed to amplify the DNA fragments used for cloning according to the manufacturer's instructions. The amplified fragments were cloned using pEASY-Blunt Zero (TransGen Biotech, Beijing, China). Sequencing was performed by Tianyi Huiyuan Biotech (Wuhan, China), and the data were analyzed using DNAMAN 8.0 (LynnonBiosoft, Quebec, Canada). The obtained pEASY-*NlPLE25* vector was used in this study.

2.4 Plasmid construction

Three plasmids (pEGFP, p*NlPLE25*-EGFP, and pReplace-EGFP) were constructed for the excision assays. The pIEX/Bac1 vector was purchased from Youbio Biotech (Changsha, China). The enhanced green fluorescent protein (EGFP) was placed under the regulation of homologous region 5 (hr5) enhancer, IE1 promoter, and P10 promoter in pIEX/Bac1 by homologous recombination cloning to construct the pEGFP

vector. The homologous recombination-based cloning method was performed using the In-Fusion HD Cloning Kit (TaKaRa Bio, Kusatsu, Japan). Similarly, full-length *NlPLE25* was inserted between the promoter and the EGFP to construct the p*NlPLE25*-EGFP, where the *N. lugens* carboxylesterase precursor gene (MF278673.1, 1.64 kb) was cloned into p*NlPLE25*-EGFP, replacing the *NlPLE25* transposase ORF, in order to construct pReplace-EGFP.

For the transposition assays, four reporter plasmids (pPB-BleoR, p*NlPLE25*-BleoR, p*NlPLE25*, and pPB) were constructed. For pPB-BleoR and p*NlPLE25*-BleoR, the OpIE2-driven bleomycin resistance gene from PIZ/V5-His was ligated into the flanking sequences of the TIRs and STIRs of *piggyBac* and *NlPLE25* transposons, respectively, to build transposon reporters. The flanking sequences were obtained from pXL-BacII-3XP3-RFP and pEASY-*NlPLE25*. To construct the transposase expression vectors p*NlPLE25* and pPB, the *NlPLE25* transposase and *T. ni piggyBac* transposase were cloned into pIEX/BacI vectors. All cloning experiments were performed using the In-Fusion HD Cloning Kit (TaKaRa Bio). The primers for vector construction are listed in Table S1.

2.5 Excision assays

The excision activity was tested using S2 cells. EGFP cannot be expressed in p*NlPLE25*-EGFP unless the *NlPLE25* transposase protein removes the full-length *NlPLE25* from the sequence end TTAA. pReplace-EGFP was used as negative control and pEGFP as positive control. Prior to transfection, the S2 cells were seeded into 6-well plates and cultured at 28 ℃ overnight. At 80%–90% confluence, cells were transfected using FuGENE HD (Promega, Madison, WI, USA) with 2 μg of plasmid. The green fluorescence was visualized under a fluorescent microscope (Leica Biosystems, Nussloch, Germany) at 48 h post-transfection, and green fluorescence cells and total cells were counted in 10 fields at 200× magnification. The ratio of green fluorescence cell number to total cell number reflects the excision efficiency of *NlPLE25*. The DNA was extracted using the Omega Tissue DNA Extraction Kit (Omega Bio-Tek) at 48 h post-transfection for sequence analysis. Primer pairs were designed based on adjacent areas of *NlPLE25* on p*NlPLE25*-EGFP (Table S1), and polymerase chain reaction (PCR) amplification was carried out as described in Section 2.3.

2.6 Resistance colony formation assay

Sf9 cells were divided into six groups for transfection. The two control groups were pPB-BleoR and p*NlPLE25*-BleoR, and the four experimental groups were: (1) p*NlPLE25* and p*NlPLE25*-BleoR; (2) pPB and pPB-BleoR; (3) p*NlPLE25* and pPB-BleoR; and (4) pPB and p*NlPLE25*-BleoR. The transfection and excision assays were carried out as described in Section 2.5. Upon transfection, cells were seeded at a density of 1×10^6 cells per 10-cm dish and selected with 0.2% (2 g/L) Zeocin (Jingxin Biotech, Guangzhou, China) for two weeks. The resultant colonies were fixed with 10% (volume fraction) methyl alcohol and stained with 0.1% (1 g/L) crystal violet (Macklin Biochemical, Shanghai, China). To detect the presence of plasmids in cells after drug screening, 500 ng of purified DNA was used to transform *Escherichia coli* DH5α competent cells. The Luria-Bertani plates were then incubated at 37 ℃ overnight.

2.7 RT-PCR

After two weeks of selection, the messenger RNA (mRNA) levels of the bleomycin resistance gene were measured using reverse transcription (RT)-PCR. The RNA was isolated using TRIzol Reagent (Thermo Fisher Scientific), according to the manufacturer's instructions, and used for RT by the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech) kits. PCR amplification was carried out as described in Section 2.3. The primers for RT-PCR are listed in Table S1.

2.8 Inverse PCR

The genomic DNA from Sf9 cells was digested with *Hae*III. The digestion products were ligated, generating circular molecules containing either the right end or the left end of the transposon. PCR amplification was carried out as described in Section 2.3. The primers for inverse PCR are listed in Table S1.

2.9 Spatial and temporal expression patterns of *NlPLE25*

The total RNAs from the seven developmental stages (egg, 1st instar nymph, 2nd instar nymph, 3rd instar nymph, 4th instar nymph, 5th instar nymph,

and adult female) and six adult tissues (head, fat body, cuticle, leg, midgut, and ovary) of insects were extracted and reversely transcribed as described in Section 2.7. Quantitative real-time PCR (qPCR) was carried out using the Roche 480 Real-Time PCR System (Roche, Basile, Switzerland) with SYBR Green qPCR Master Mix (EZBioscience, Roseville, MN, USA). The primers for qPCR are listed in Table S1.

2.10 Data analysis

Statistical analyses were performed using GraphPad Prism 8 (GraphPad Software; https://www.graphpad.com). For the excision activity of *NlPLE25*, statistical significance was assessed with paired *t*-test. A value was considered statistically significant at $P<0.05$, and highly significant at $P<0.01$. For the frequency of *NlPLE25* transposase integration, the average values from all the independent replicates are shown in the figures. Data are represented as mean±standard error, with five replicates.

3 Results

3.1 *PiggyBac* superfamily in BPH

The results revealed that the BPH genome assembly was composed of 59.06% of repeats, including 10.26% retroelements, 2.90% DNA transposons, 1.90% simple repeats, 0.35% low-complexity regions, and 43.65% unclassified regions (Table 1). The 988 *piggyBac* sequences predicted from DNA transposons had a total length of 316 836 bp, which accounts for 0.03% of the BPH genome (Table 1). The TBLASTn analysis revealed that 55 homologs of *T. ni piggyBac* were present in the BPH genome. Subsequently, we performed TTAA and TIR searches on the 3-kb upstream and downstream positions of these

Table 1 Repetitive element contents in brown planthopper (BPH)

Class	Number	Length (bp)	Ratio of genome (%)
DNA transposons	64 016	31 563 456	2.90
Hobo-Activator	4818	1 544 728	0.14
Tc1-IS630-Pogo	28 786	10 196 413	0.94
En-Spm	0	0	0
MuDR-IS905	0	0	0
PiggyBac	988	316 836	0.03
Tourist/*Harbinger*	5626	1 516 002	0.14
Other (*Mirage*, *P-element*, *Transib*)	988	568 632	0.05
Retroelements	219 959	111 649 813	10.26
*SINE*s	0	0	0.00
Penelope	18 531	5 300 981	0.49
*LINE*s	191 954	80 571 250	7.41
CRE/*SLACS*	0	0	0
L2/*CR1*/*Rex*	124 335	53 275 704	4.90
R1/*LOA*/*Jockey*	403	211 231	0.02
R2/*R4*/*NeSL*	1457	702 984	0.06
RTE/*Bov-B*	32 372	13 096 311	1.20
L1/*CIN4*	0	0	0
LTR	28 005	31 078 563	2.86
BEL/*Pao*	8446	11 603 870	1.07
Ty1/*Copia*	1351	648 582	0.06
Gypsy/*DIRS1*	18 208	18 826 111	1.73
Retroviral	0	0	0
Rolling circles	0	0	0
Unclassified	1 941 732	474 795 500	43.65
Total interspersed repeats		618 008 769	56.81

*SINE*s: short interspersed elements; *LINE*s: long interspersed elements; *SLACS*: spliced leader associated conserved sequence; *L1*: *LINE* 1; *L2*: *LINE* 2; *CR1*: chicken repeat 1; Rex: retroelement of *Xiphophurus*; *LTR*: long terminal repeat.

sequences, and filtered out sequences with ORFs coding for amino acid residues of <250. When TIRs were found on both ends of an ORF, the sequence was considered to be a *PLE*, while the remaining sequence was considered to be a *piggyBac*-derived (*PGBD*) element or *PGBD*-like element. Finally, 28 sequences with perfect flanks were determined and named as *NlPLE1–NlPLE28* (Table 2).

Each of the 28 *NlPLE*s had a total length of 2088–5215 bp with an ORF of 759–2022 bp (Table 2). The length of TIRs ranged between 15 and 23 bp, which is slightly longer than that of 12–19 bp previously reported by Fraser et al. (1983). The first six nucleotides were relatively well conserved following the motif C[**C**/A]C[**T**/A/G][**T**/C/A][**T**/A/C] (Fig. 1a). Each *NlPLE* had one or more copies in the genome (Table 2 and Fig. 1b) distributed across the BPH chromosomes, except for *NlPLE14*, which exhibited the highest enrichment in chromosome 1 (Fig. 1b). The amino acid sequence alignment showed that transposases of the 28 *NlPLE*s had a similarity of 6.23%–32.28% with those of the *T. ni piggyBac* (Table 2 and Fig. 1c). *NlPLE24* and *NlPLE1* showed the highest (32.28%) and lowest (6.23%) sequence similarity, respectively. The similarity of *NlPLE* amino acid sequences ranged between 1.98% and 62.12% (Table S2).

To confirm the accurate phylogenetic relationships of *NlPLE* transposases, 157 previously described PLE transposases were selected from different genomes for the construction of the phylogenetic tree. Despite the divergence, the sequence analysis showed that most PLE transposases were found to contain conserved domains for a total of about 300 residues that were roughly delimited by the putative DDD motif and conserved cysteine residues (Figs. 1c and S1). Thus, a phylogenetic tree was constructed based on these conserved regions using MEGA. The simplified phylogeny is shown in Fig. 1d (see Fig. S2 for the complete tree). The *NlPLE* transposases were classified into six major groups. The largest clade included 14 *NlPLE*s and was clustered together with the *T. ni piggyBac*.

Table 2 Terminal inverted repeats (TIRs) of complete *Nilaparvata lugens piggyBac*-like elements (*NlPLE*s)

Name	Full length (bp)	ORF length (bp)	Amino acid length	5' TTAA TIR	Percentage identity (%)*	Copies
NlPLE1	3383	1101	367	TTAACCCTTAATTACTAGGGATG	6.23	1
NlPLE2	2942	1953	651	TTAACCCTCTAGTGCAT	15.92	6
NlPLE3	3259	1173	391	TTAACCCTTTATTGCATG	15.33	2
NlPLE4	2527	921	307	TTAACACTCCGCTGCTCGC	26.69	3
NlPLE5	2848	2022	674	TTAACCCTTAATTACTACAA	15.17	3
NlPLE6	2921	1887	629	TTAACCCTACAATGCAT	9.63	2
NlPLE7	2808	1983	661	TTAACCCTACAATGCATAAT	8.86	2
NlPLE8	2990	1413	471	TTAACCCTTTAATGCATG	15.07	4
NlPLE9	3328	1209	403	TTAACCCTTCAAGTGATGGGTG	14.33	2
NlPLE10	2284	759	253	TTAACCCTTAGCGCTCG	11.00	1
NlPLE11	3173	1257	419	TTAACCCTACACCGCATG	7.30	2
NlPLE12	2967	1950	650	TTAACCCTCTAGTGCAT	12.71	6
NlPLE13	3213	1809	603	TTAACCCTTTATTGCACAAT	14.05	4
NlPLE14	2953	1887	629	TTAACCCTCGGGCACAGGCAC	8.58	2
NlPLE15	3241	771	257	TTAACCCTTAGTTACTATC	18.59	1
NlPLE16	3001	804	268	TTAACCCTTAAATGCATG	18.44	4
NlPLE17	2951	1014	338	TTAACACTATTACCGAAACCC	20.07	5
NlPLE18	2401	1146	382	TTAACCCTGCATTACTCGCGCGGG	23.60	4
NlPLE19	2762	1875	625	TTAACCCTTTGCGATC	18.22	2
NlPLE20	3337	1008	335	TTAACCCTCAATTACTACGGA	19.48	1
NlPLE21	3092	993	330	TTAACACGTGAGTACACGCCCCC	17.31	3
NlPLE22	2412	1887	628	TTAACCCTTTAATGCATG	12.20	4
NlPLE23	3247	1899	632	TTAACCCTTAGTTACTATC	27.46	1
NlPLE24	5215	1704	567	TTAACCCTGCGGTACTCGCGCGGGCTA	32.28	5
NlPLE25	2590	1761	586	TTAACCCTCCGGTAGGCGCGCGG	30.25	8
NlPLE26	2656	930	309	TTAACCCTTTATAAGGTAATGGCAGA	10.45	1
NlPLE27	2088	1563	520	TTAACCCAGTAAGAACCACGT	28.67	4
NlPLE28	2424	1899	632	TTAACCCTTAATTACTACAAGC	27.50	3

* Percent (%) indicates sequence similarity to *Trichoplusia ni piggyBac* transposase. ORF: open reading frame.

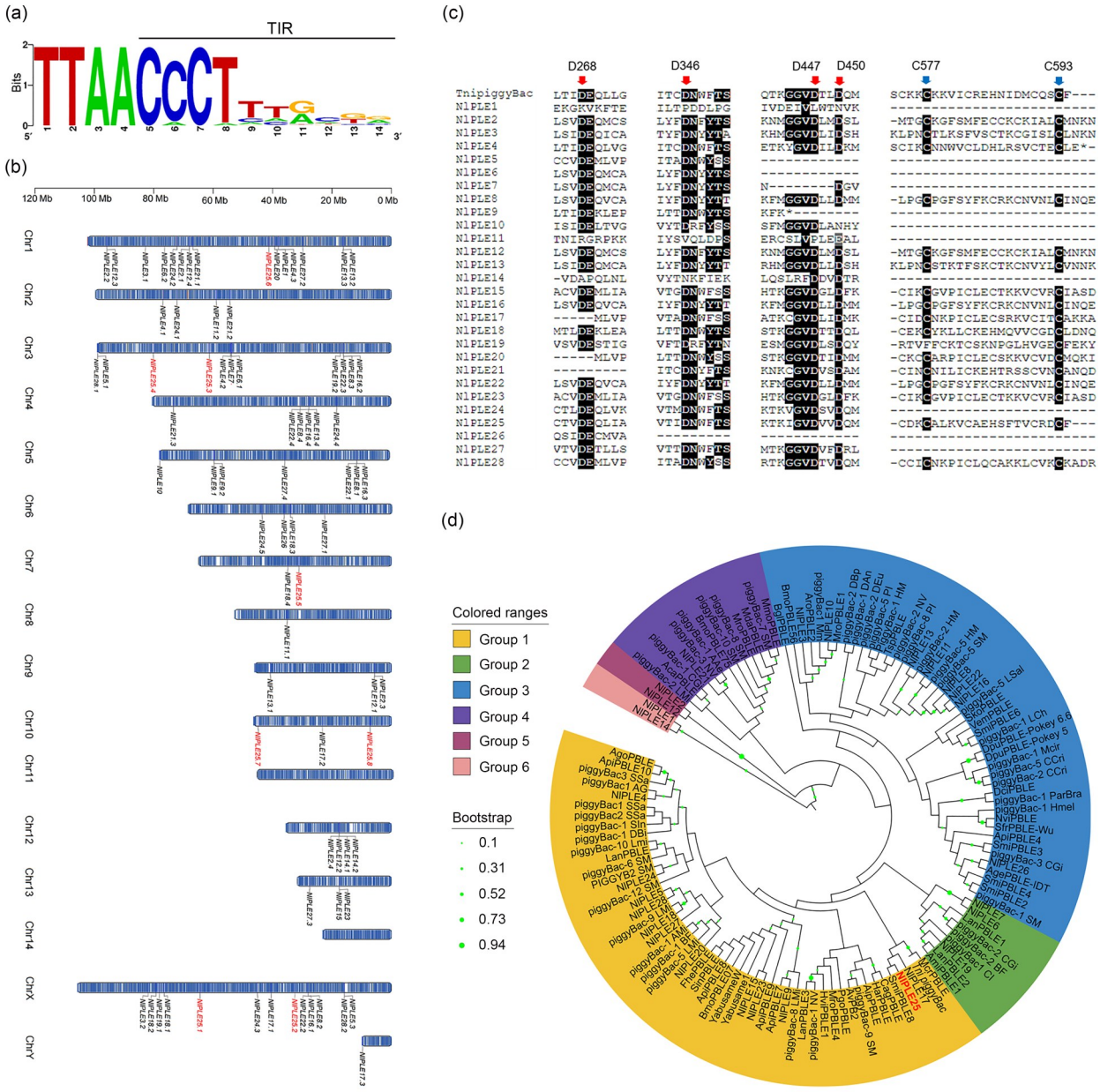


Fig. 1 ***PiggyBac*** **superfamily in the brown planthopper (BPH). (a) Sequence logo of terminal inverted repeats (TIRs) from** ***Nilaparvata lugens piggyBac*****-like elements (*NlPLE*s), generated using WebLogo (http://weblogo.berkeley.edu/logo.cgi). The horizontal axis represents sequential positions relative to TTAA integration target sites. The vertical axis represents the frequency of bases at each position. Height increases with the frequency of bases at a given position. (b) Chromosomal locations of *NlPLE*s. The scale refers to chromosomal size. The chromosome number is indicated at the bottom of each chromosome (Chr). The color on each chromosome represents gene density. The distribution of *NlPLE25* on chromosomes is indicated in red lettering. (c) Amino acid alignments of 28 *NlPLE*s and *Trichoplusia ni piggyBac.* The protein sequences were aligned using CLUSTAL. The triple-aspartate (DDD) domain was observed and indicated by a red arrowhead above the sequences. The C-terminal region was observed and indicated by blue arrowhead. D268, D346, D447, D450, C577, and C593 are the positions of amino acids in *T. ni piggyBac* transposase. (d) Phylogenetic relationships among *piggyBac*-like elements based on the transposase sequences. The phylogeny is based on amino-acid sequences covering approximately 300 residues (all information of the amino acid sequence is listed in Table S3; the alignment is given in Fig. S1; the complete tree is given in Fig. S2). The tree was constructed according to a maximum-likelihood method using MEGA 6.0. The bootstrapping values from 100 replications are shown in the green circle. The colored frames represent six well supported clades.**

Furthermore, we found that *NlPLE17* and *NlPLE25* showed a close evolutionary relationship with *T. ni piggyBac*.

The *NlPLE*s identified via whole-genome BLAST analysis were widely distributed throughout the BPH genome. *NlPLE25* was one of the most prominent *NlPLE*s, sharing a 30.25% similarity with *T. ni piggyBac* transposase. Therefore, it was selected for further study.

3.2 Structure and distribution of *NlPLE25*

The *NlPLE25* was found to be 2582-bp long and flanked by typical tetranucleotide target site TTAA duplications (Fig. 2a). Both ends of *NlPLE25* contain a 16-bp TIR and a 20-bp STIR, which are asymmetric with a 19-bp spacer at the 3' end and a sequence duplication between segments relative to the 5' end (Fig. 2a). The TIRs and STIRs of *NlPLE25* had a 50% similarity with *T. ni piggyBac* (Fig. 2b). The *NlPLE25* had an ORF encoding a 586-amino acid transposase (Fig. 2c), and a bipartite NLS (RRCQIDGLPVQLKKRGK) with a score of 0.89 detected using PSORT II. The *NlPLE25* transposase contains one N-terminal domain (NTD), one CRD, two dimerization and DNA-binding domains (DDBDs), and one catalytic domain with an insertion domain (Figs. 2b and 2c). The catalytic domain contains a conserved DDD motif (D263, D343, and D446), consistent with the amino acids of D268, D346, and D447 in *T. ni piggyBac* transposase (Fig. 2c).

The chromosomal location and copy number of *NlPLE25* were determined using the NCBI databank (Fig. 1b and Table 2). The BLAST results showed that *NlPLE25* exists with at least eight copies in the BPH genome, distributed on five chromosomes (1, 3, 7, 10, and X). These copies were designated as *NlPLE25.1*–*NlPLE25.8* (Fig. 2e). Five out of eight copies had the same sequence. Two copies (*NlPLE25.6* and *NlPLE25.8*) exhibited insertions and/or deletions (indels); the indels of *NlPLE25.6* suggested a premature translation termination, and the coding region of *NlPLE25.8* was disrupted because of the indels. The flanking sequences of *NlPLE25*, except for *NlPLE25.1* and *NlPLE25.8*, were localized at sites with a consensus TTAA, which is characteristic of *piggyBac* family members. The expression of *NlPLE25* transposase was significantly higher in the 5th instar nymph than in other stages, whereas the lowest expression was detected in the 3rd and 4th instars (Fig. S1a). *NlPLE25* was mainly expressed in the fat body, with significantly higher levels than those in other tissues (Fig. S1b).

3.3 Excision activity of *NlPLE25* in S2 cells

Since *NlPLE25* and *T. ni piggyBac* were found structurally similar (Fig. 2c), it was hypothesized that *NlPLE25* might have functions analogous to *T. ni piggyBac*. In p*NlPLE25*-EGFP, the expression of EGFP was blocked by the full-length *NlPLE25* until its removal by the expressed transposase (Fig. 3a). Thus, no green fluorescence was observed when pReplace-EGFP was transfected (Fig. 3b). However, 1.26% and 40.61% of the cells exhibited green fluorescence when p*NlPLE25*-EGFP and pEGFP were transfected, respectively (Figs. 3b and 3c). In addition, DNA amplification produced an approximately 500-bp band, and sequencing revealed the traceless excision of *NlPLE25* (Figs. 3d and 3e).

3.4 DNA transposition activity of *NlPLE25*

It was assumed that *NlPLE25* would retain its transposase activity because of its traceless excision activity. This was examined by analyzing *NlPLE25* transposition in cultured Sf9 cells using a synthetic transposon reporter in a transposase expression plasmid co-transfection assay. The synthetic transposon reporter was comprised of a bleomycin resistance gene flanked by TIRs and STIRs of the transposon (Fig. 4a). Cells were subjected to the excision assay two days after transfection. Six groups were included in the assay: two groups for single transfection (L1 and L2; Fig. 4b), and four groups for co-transfection (L3–L6; Fig. 4b). All of the co-transfected groups (L3–L6) were successfully excised, whereas such changes were not observed in the single transfection groups (L1 and L2; Fig. 4b). In L4 and L6, *NlPLE25* and *T. ni piggyBac* transposases recognized the TIRs and STIRs of each other (Fig. 4b). The clonogenic assay revealed significant rates of bleomycin resistance of cells conferred by the transposon reporter with transposase expression plasmids (Fig. S4); however, bleomycin resistance was not observed in cells expressing a control lacking the transposase (Figs. 4c and 4d). The relative integration frequency in the pPB+pPB-BleoR group was significantly higher than that in the other groups. In contrast, no significant differences were observed between any two of the three groups (pPB+

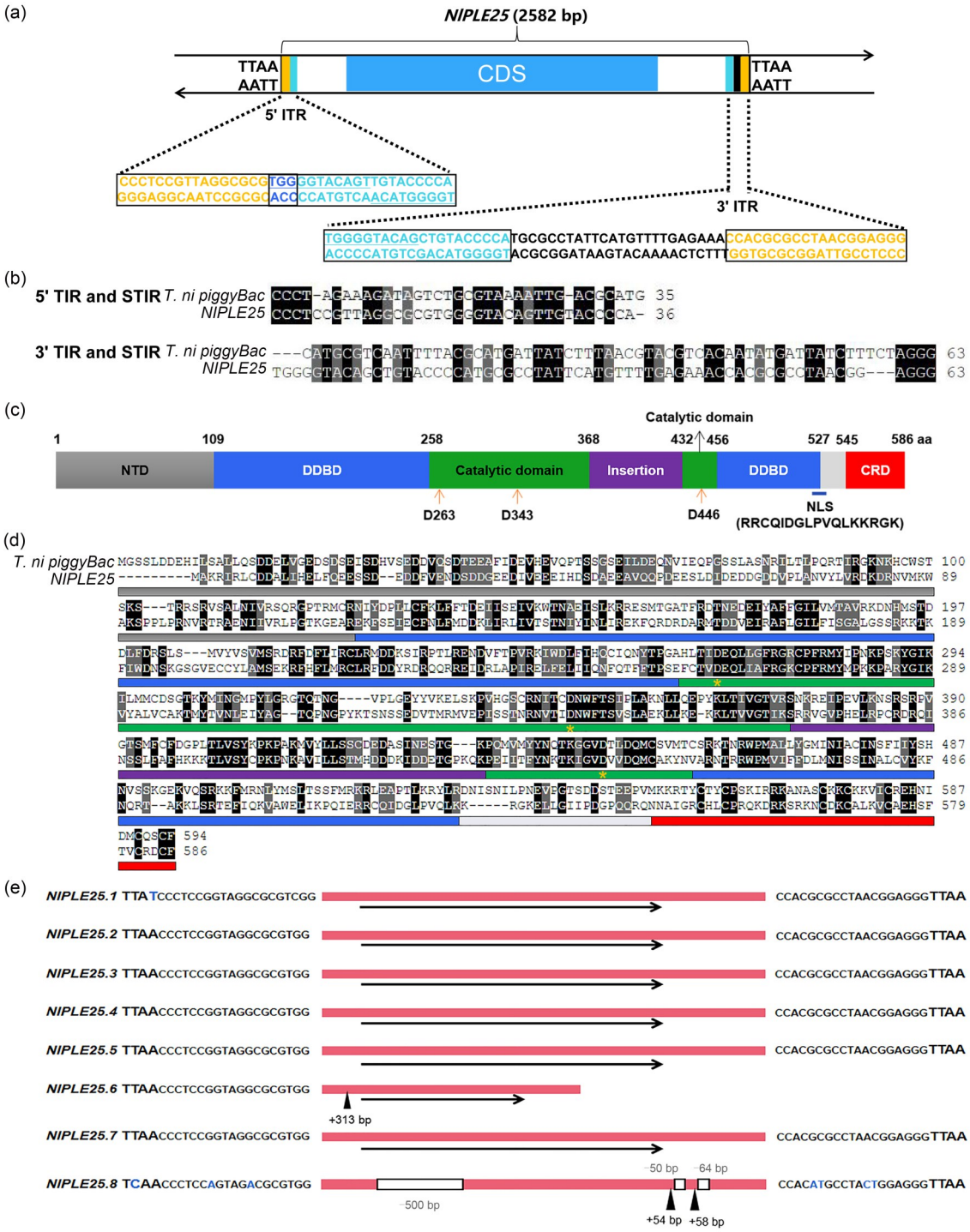


Fig. 2 Structure and distribution of *NlPLE25*. (a) Schematic representation of *NlPLE25* transposon flanked by TTAA, and the organization of 5'- and 3'-terminal inverted repeats (TIRs) and sub-terminal asymmetric inverted repeats (STIRs). (b) Sequence alignment. (c) Predicted domain organization of *NlPLE25* transposase according to *Trichoplusia ni piggyBac*. The catalytic domain contains the conserved DDD motif (orange arrow). Gray indicates disordered regions in the structure. (d) Sequence alignment of putative transposases. The different colors correspond to different domains from Fig. 2c. The orange asterisks indicate the DDD motif. (e) Schematic representation of different *NlPLE25* copies. The black arrow line indicates the open reading frame (ORF). The black triangle represents an insertion. The white square represents a deletion. CDS: coding domain sequence; NTD: N-terminal domain; DDBD: dimerization and DNA-binding domain; CRD: C-terminal cysteine-rich domain.

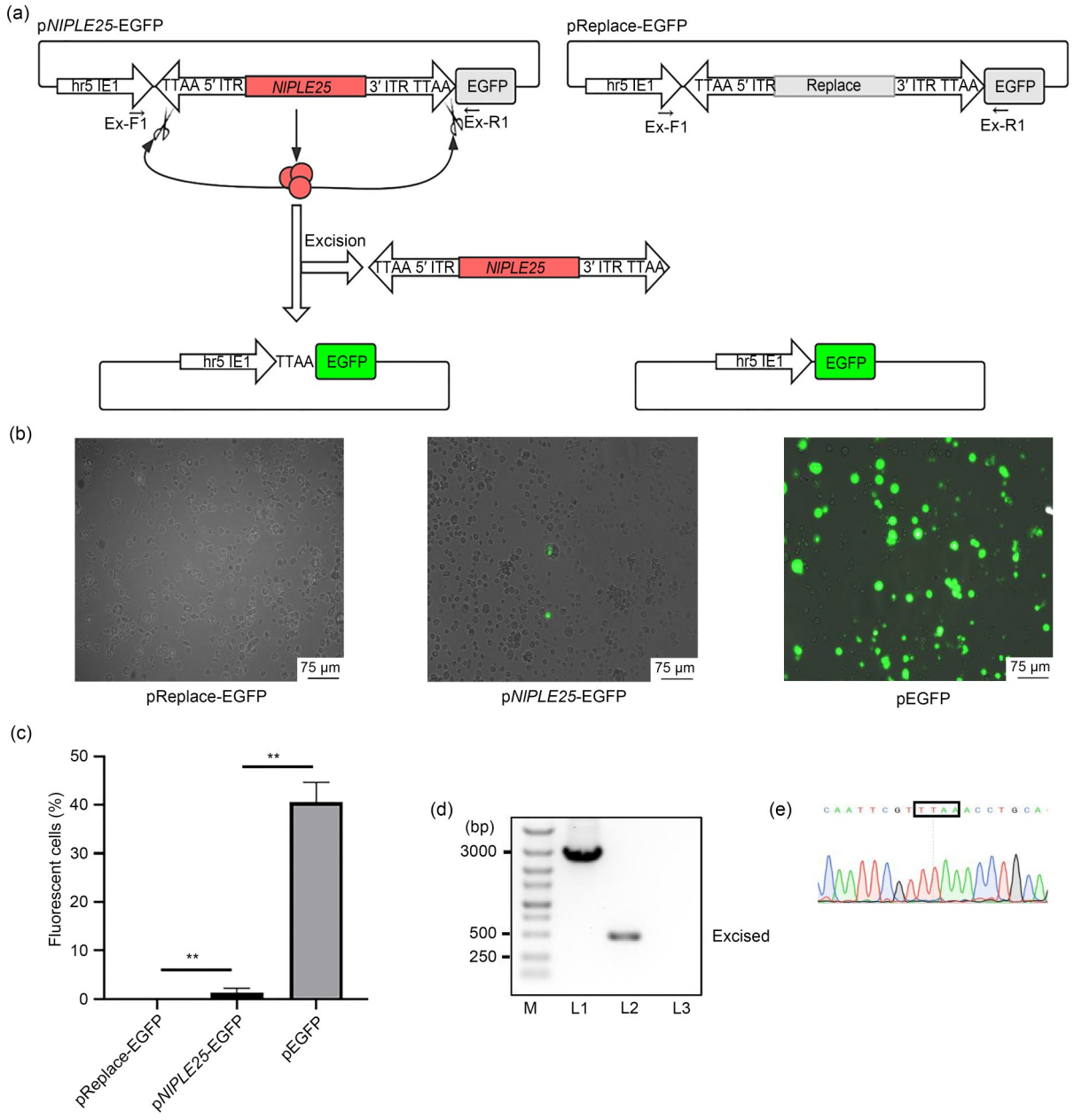
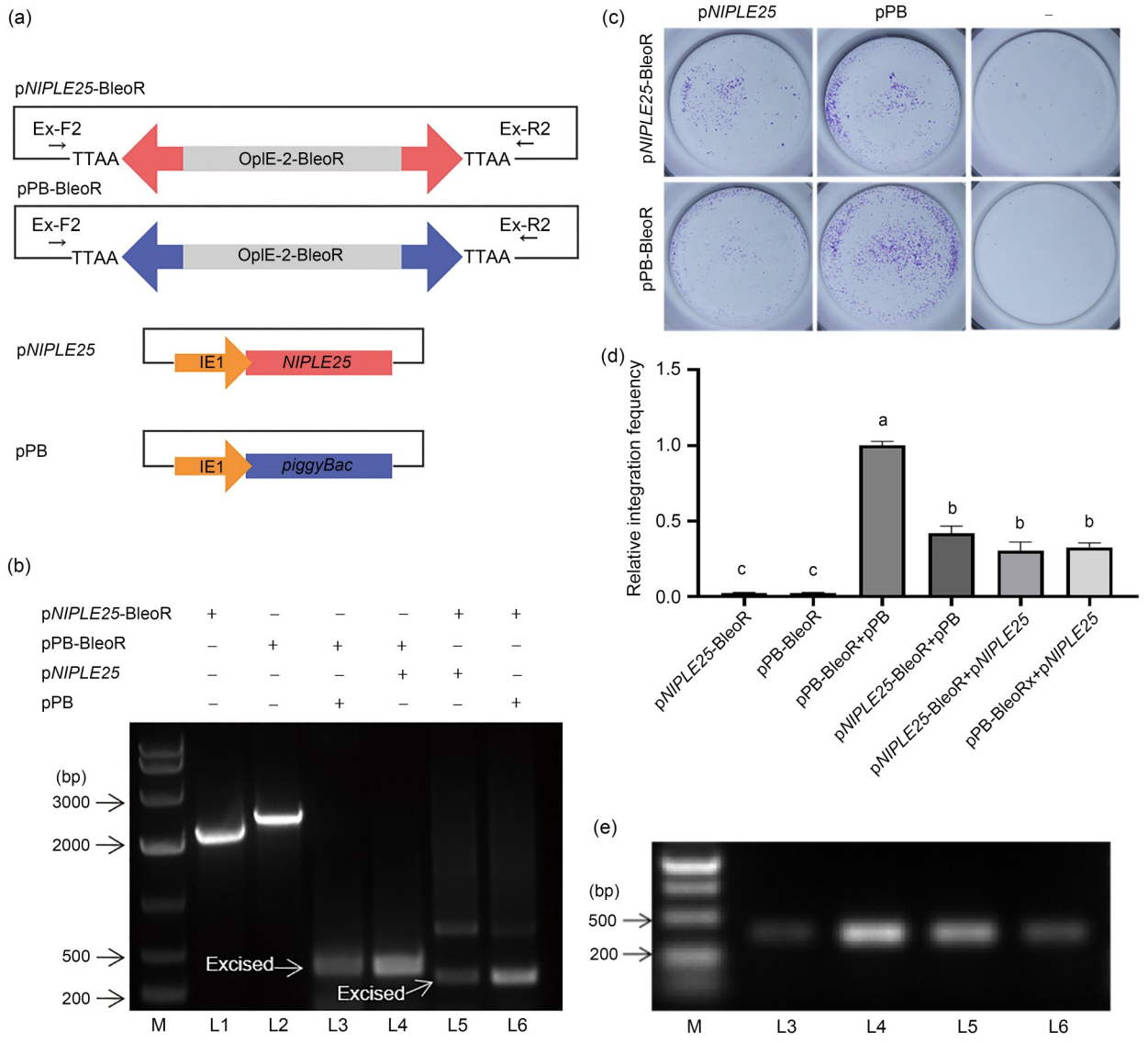


Fig. 3 Excision activity of *NlPLE25* in S2 cells. (a) Diagram showing the excision of *NlPLE25* transposon from the EGFP expression cassette (p*NlPLE25*-EGFP). PCR primers Ex-F1 and Ex-R1 are shown as black arrowheads. (b) Fluorescent image of S2 cells. (c) Excision efficiency of *NlPLE25* measured by counting GFP$^+$ cells in ten random fields. The data are represented as mean±standard error (n=5). ** P<0.01. (d) PCR products from transfected cells using primers Ex-F1 and Ex-R1. M, molecular weight marker; L1, cells transfected with pReplace-EGFP; L2, cells transfected with p*NlPLE25*-EGFP; L3, cells transfected with pEGFP. (e) Sequencing of excised PCR products. EGFP: enhanced green fluorescent protein; TIR: terminal inverted repeat; PCR: polymerase chain reaction; hr5: homologous region 5 enhancer; IE1: IE1 promoter.

p*NlPLE25*-BleoR, p*NlPLE25*+p*NlPLE25*-BleoR, and p*NlPLE25*+pPB-BleoR) (Fig. 4d). The frequency of *NlPLE25* transposase integration was 30.5% of that of *T. ni piggyBac* (Fig. 4d). The RT-PCR results confirmed the stable expression of the bleomycin resistance gene in the four co-transfection groups (Fig. 4e). Inverse PCR performed on the four lines generated using the four different co-transfection groups revealed that the 5' insertion sequence matched the 3' sequence (Fig. 4f), which argued for single insertions. Thus, the



(f)

		Spodoptera frugiperda genome	Transposon
pPB+pPB-BleoR	Chr16	10773676–10773696	ttttaaatgataaactcaagctcagaa**TTAA**CCCTAGAAAGATAATCATATTGTGACG
p*NIPLE25*+pPB-BleoR	Chr20	11486311–11486341	tagtaaattggcaatattcggatggcg**TTAA**CCCTAGAAAGATAATCATATTGTGACG
p*NIPLE25*+p*NIPLE25*-BleoR	Chr21	486228–486258	ggaattttgaaacatgccgcagctaga**TTAA**CCCTCCGTTAGGCGCGTGGGGTACAGT
pPB+p*NIPLE25*-BleoR	Chr18	8579447–8579477	gtacccagtatcgactcggtggcccat**TTAA**CCCTCCGTTAGGCGCGTGGGGTACAGT

Fig. 4 *NlPLE25* transposition in Sf9 cells. (a) Schematic representation of synthetic transposon substrates and transposase expression plasmids used for DNA transposition assays. (b) PCR products from transfected cells using primers Ex-F2 and Ex-R2. M, molecular weight marker; L1, cells transfected with p*NlPLE25*-BleoR; L2, cells transfected with pPB-BleoR; L3, cells transfected with pPB-BleoR and pPB; L4, cells transfected with pPB-BleoR and p*NlPLE25*; L5, cells transfected with p*NlPLE25*-BleoR and p*NlPLE25*; L6, cells transfected with p*NlPLE25*-BleoR and pPB. (c) Representative photographs of crystal violet-stained colonies obtained after zeocin selection of Sf9 cells co-transfected with transposon reporter plasmid and transposase cDNA expression plasmids. (d) Quantitative graph showing live cells from Fig. 4c. The results are presented relative to "pPB-BleoR+pPB." The data are represented as mean±standard error (n=5); the different lower-case letters above the columns represent significant differences at P< 0.05. (e) Expression of bleomycin resistance gene mRNA determined by RT-PCR, following zeocin resistance screening. (f) Alignment of the representative DNA sequences of identified genomic integration sites. RT-PCR: reverse transcription-polymerase chain reaction; cDNA: complementary DNA; mRNA: messenger RNA; Chr: chromosome.

observed results indicated that *NlPLE25* retained transposon activity and could act on the TIRs and STIRs of *T. ni piggyBac*, probably due to the high sequence similarity in the TIRs and STIRs between *NlPLE25* and *T. ni piggyBac* (Fig. 2b).

4 Discussion

In the present study, 28 *NlPLE*s with perfect TSDs were identified in the BPH genome using different methods and bioinformatics tools. These *NlPLE*s retained all the canonical features of the *piggyBac* superfamily, including TIRs at the extreme ends and an ORF encoding a transposase.

An *NlPLE* was observed to always insert into TTAA target sites, similar to the results reported by Bouallègue et al. (2017). The sequence of the TIRs varies among different *NlPLE*s, but there was a notable conservation of the first six nucleotides C[**C**/A]C[**T**/A/G][**T**/C/A][**T**/A/C], which is also similar to the result of Bouallègue et al. (2017). Furthermore, the first three nucleotides in TIRs play an essential role in transposase recognition, since they are cleaved from the target site, and even a single nucleotide deletion is sufficient to abolish excision (Elick et al., 1997). The sequence of the first three nucleotides is commonly CCC; however, other sequence types can also be recognized and cleaved by PLE transposase (Elick et al., 1997).

A more in-depth analysis of *NlPLE25* allowed the identification of eight *NlPLE25* copies, some of which had defects (Figs. 1b and 2e). Incomplete fragments could be rapidly lost in the absence of *trans*-mobilization (le Rouzic et al., 2007; Hua-Van et al., 2011) or domesticated into a functional protein (Luo et al., 2014). Similar to *T. ni piggyBac*, *NlPLE25* was comprised of two sections (Fig. 2a): (1) asymmetric TIRs and STIRs at both ends, necessary for transposase to form the active synaptic complex in vivo (Chen QJ et al., 2020); and (2) a transposase ORF. The *NlPLE25* transposase has similar functional domains to the *T. ni piggyBac*, including NTD, CRD, DDBD, and the catalytic domain (Figs. 2c and 2d). The CRD is required for transposase activity (Chen QJ et al., 2020) and is frequently found in other *PLE*s. The catalytic domain of *NlPLE25* transposase contains DDD motif, corresponding to D268, D346, and D447 in *T. ni piggyBac* transposase (Figs. 2c and 2d). Furthermore, a putative NLS enables *NlPLE25* transposase to enter the nucleus for DNA translocation catalysis (Fig. 2c). Nevertheless, additional studies are required to validate the existence of this putative NLS.

Our experiments confirmed the speculation that *NlPLE25* might retain its transposase activity. The frequency of *NlPLE25* transposase integration was 30.5% compared to that of *T. ni piggyBac* (Fig. 4d). *NlPLE25* transposase recognized the TIRs and STIRs of *T. ni piggyBac*, probably because of the similar transposon and transposase structures between *NlPLE25* and *T. ni piggyBac*. A similar phenomenon was observed in human *PGBD5*, the most evolutionarily conserved transposable element-derived gene in vertebrates, which can recognize *T. ni piggyBac* TIRs and catalyze DNA transposition in human cells (Henssen et al., 2015). Nevertheless, the transposase activity of *NlPlE25* is a double-edged sword. On the one hand, it can be used as a potential insect transgenic tool. Hyperactive *piggyBac* transposase has seven amino acid substitutions related to *T. ni piggyBac* and is commonly used in transgenic applications, but its efficiency is not always high. On the other hand, *NlPLE25* may influence the genetic stability of *piggyBac* transgenic lines. Endogenous *PLE*s may repress the transposition of introduced *piggyBac* vectors and their stability (Lorenzen et al., 2003; Jia et al., 2021). In the present study, *NlPLE25* was found to share TIRs and STIRs with *T. ni piggyBac*; therefore, *NlPLE25* might transfer or remove foreign fragments in the *piggyBac* transgenic lines of BPH. The expression of *NlPLE25* was observed in nearly all BPH tissues and throughout the BPH development period (Figs. S3a and S3b). If *NlPLE25* removes foreign fragments of *piggyBac* from the genome of gonadal germline cells, the genetic stability of the transgenic lines will be destroyed. It is important to note that, beyond *NlPLE*s, other sequences flanked by a single TIR or no TIR exist, but these encode functional transposase in the BPH genome, and thus could still be capable of cross-reacting with integrated *piggyBac* transposons. Although host genomes have evolved multiple mechanisms to negatively regulate transposable element activity (Reik, 2007; Slotkin and Martienssen, 2007; Jacobs et al., 2014; Sanchez-Luque et al., 2019), this risk should not be ignored.

One of the potential limitations of this work is that it is necessary to assess the effect of endogenous

*PLE*s on the stability of *piggyBac* transgenic lines in an in vivo system, as previously discussed. Future work might need to assess this risk by using *piggyBac* transgenic BPH lines.

5 Conclusions

This study aimed to identify endogenous *PLE*s in the BPH. We identified 28 *PLE*s and found that *NlPLE25* had the highest copy number and was distributed across five chromosomes. *NlPLE25* transposase achieved precise excision and transposition in cultured insect cells. We also indicated cross-recognition between the *NlPLE25* transposon and the *piggyBac* transposon. The findings of this study will contribute to our understanding of the distribution and characteristics of members of the *piggyBac* superfamily and their application in insects.

Acknowledgments

This work was supported by the Foundation of Guangzhou Science and Technology Key Project (No. 201904020041), China.

Author contributions

Jun LYU designed and performed the experiment, wrote and edited the manuscript. Qin SU, Jinhui LIU, Lin CHEN, and Jiawei SUN participated in part of the experiment and edited the manuscript. Wenqing ZHANG designed and supervised the experiment, wrote and edited the manuscript. All authors have read and approved the final version of the manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jun LYU, Qin SU, Jinhui LIU, Lin CHEN, Jiawei SUN, and Wenqing ZHANG declare that they have no conflict of interest.

All institutional and national guidelines for the care and use of laboratory animals were followed.

References

Bao WD, Kojima KK, Kohany O, 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*, 6:11. https://doi.org/10.1186/s13100-015-0041-9

Bouallègue M, Rouault J, Hua-Van A, et al., 2017. Molecular evolution of *piggyBac* superfamily: from selfishness to domestication. *Genome Biol Evol*, 9(2):323-339. https://doi.org/10.1093/gbe/evw292

Cary LC, Goebel M, Corsaro BG, et al., 1989. Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, 172(1):156-169. https://doi.org/10.1016/0042-6822(89)90117-7

Chen CJ, Chen H, Zhang Y, et al., 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*, 13(8):1194-1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen JX, Li WX, Lyu J, et al., 2021. CRISPR/Cas9-mediated knockout of the *NlCSAD* gene results in darker cuticle pigmentation and a reduction in female fecundity in *Nilaparvata lugens* (Hemiptera: Delphacidae). *Comp Biochem Physiol Part A Mol Integr Physiol*, 256:110921. https://doi.org/10.1016/j.cbpa.2021.110921

Chen QJ, Luo WT, Veach RA, et al., 2020. Structural basis of seamless excision and specific targeting by *piggyBac* transposase. *Nat Commun*, 11:3446. https://doi.org/10.1038/s41467-020-17128-1

Daimon T, Mitsuhiro M, Katsuma S, et al., 2010. Recent transposition of *yabusame*, a novel *piggyBac*-like transposable element in the genome of the silkworm, *Bombyx mori*. *Genome*, 53(8):585-593. https://doi.org/10.1139/g10-035

Elick TA, Lobo N, Fraser MJ, 1997. Analysis of the cis-acting DNA elements required for *piggyBac* transposable element excision. *Mol Gen Genet*, 255(6):605-610. https://doi.org/10.1007/s004380050534

Fraser MJ, Smith GE, Summers MD, 1983. Acquisition of host cell DNA sequences by baculoviruses: relationship between host DNA insertions and FP mutants of *Autographa californica* and *Galleria mellonella* nuclear polyhedrosis viruses. *J Virol*, 47(2):287-300. https://doi.org/10.1128/jvi.47.2.287-300.1983

Fraser MJ, Clszczon T, Elick T, et al., 1996. Precise excision of TTAA-specific lepidopteran transposons *piggyBac* (IFP2) and *tagalong* (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect Mol Biol*, 5(2):141-151. https://doi.org/10.1111/j.1365-2583.1996.tb00048.x

Gregory M, Alphey L, Morrison NI, et al., 2016. Insect transformation with *piggyBac*: getting the number of injections just right. *Insect Mol Biol*, 25(3):259-271. https://doi.org/10.1111/imb.12220

Henssen AG, Henaff E, Jiang E, et al., 2015. Genomic DNA transposition induced by human PGBD5. *eLife*, 4:e10565. https://doi.org/10.7554/elife.10565

Hua-Van A, le Rouzic A, Boutin TS, et al., 2011. The struggle for life of the genome's selfish architects. *Biol Direct*, 6:19. https://doi.org/10.1186/1745-6150-6-19

Hubley R, Finn RD, Clements J, et al., 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res*, 44(D1):D81-D89. https://doi.org/10.1093/nar/gkv1272

Jacobs FMJ, Greenberg D, Nguyen N, et al., 2014. An evolutionary arms race between KRAB zinc-finger

genes *ZNF91/93* and SVA/L1 retrotransposons. *Nature*, 516(7530):242-245. https://doi.org/10.1038/nature13760

Jia XH, Pang XY, Yuan YJ, et al., 2021. Unpredictable recombination of PB transposon in Silkworm: a potential risk. *Mol Genet Genomics*, 296(2):271-277. https://doi.org/10.1007/s00438-020-01743-0

Keith JH, Schaeper CA, Fraser TS, et al., 2008. Mutational analysis of highly conserved aspartate residues essential to the catalytic core of the *piggyBac* transposase. *BMC Mol Biol*, 9:73. https://doi.org/10.1186/1471-2199-9-73

le Rouzic A, Boutin TS, Capy P, 2007. Long-term evolution of transposable elements. *Proc Natl Acad Sci USA*, 104(49):19375-19380. https://doi.org/10.1073/pnas.0705238104

Li C, Brant E, Budak H, et al., 2021. CRISPR/Cas: a Nobel Prize award-winning precise genome editing technology for gene therapy and crop improvement. *J Zhejiang Univ-Sci B (Biomed & Biotechnol)*, 22(4):253-284. https://doi.org/10.1631/jzus.B2100009

Lorenzen MD, Berghammer AJ, Brown SJ, et al., 2003. *piggyBac*-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol Biol*, 12(5):433-440. https://doi.org/10.1046/j.1365-2583.2003.00427.x

Luo GH, Wu M, Wang XF, et al., 2011. A new active *piggyBac*-like element in *Aphis gossypii*. *Insect Sci*, 18(6):652-662. https://doi.org/10.1111/j.1744-7917.2011.01406.x

Luo GH, Li XH, Han ZJ, et al., 2014. Molecular characterization of the *piggyBac*-like element, a candidate marker for phylogenetic research of *Chilo suppressalis* (Walker) in China. *BMC Mol Biol*, 15:28. https://doi.org/10.1186/s12867-014-0028-y

Reik W, 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447(7143): 425-432. https://doi.org/10.1038/nature05918

Sanchez-Luque FJ, Kempen MJHC, Gerdes P, et al., 2019. LINE-1 evasion of epigenetic repression in humans. *Mol Cell*, 75(3):590-604.E12. https://doi.org/10.1016/j.molcel.2019.05.024

Sarkar A, Sim C, Hong YS, et al., 2003. Molecular evolutionary analysis of the widespread *piggyBac* transposon family and related "domesticated" sequences. *Mol Genet Genomics*, 270(2):173-180. https://doi.org/10.1007/s00438-003-0909-0

Slotkin RK, Martienssen R, 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*, 8(4):272-285. https://doi.org/10.1038/nrg2072

Sun ZC, Wu M, Miller TA, et al., 2008. *piggyBac*-like elements in cotton bollworm, *Helicoverpa armigera* (Hübner). *Insect Mol Biol*, 17(1):9-18. https://doi.org/10.1111/j.1365-2583.2008.00780.x

Wang J, Ren X, Miller TA, et al., 2006. *piggyBac*-like elements in the tobacco budworm, *Heliothis virescens* (Fabricius). *Insect Mol Biol*, 15(4):435-443. https://doi.org/10.1111/j.1365-2583.2006.00653.x

Wang J, Miller ED, Simmons GS, et al., 2010. *piggyBac*-like elements in the pink bollworm, *Pectinophora gossypiella*. *Insect Mol Biol*, 19(2):177-184. https://doi.org/10.1111/j.1365-2583.2009.00964.x

Wang JJ, Du YZ, Wang SZ, et al., 2008. Large diversity of the *piggyBac*-like elements in the genome of *Tribolium castaneum*. *Insect Biochem Mol Biol*, 38(4):490-498. https://doi.org/10.1016/j.ibmb.2007.04.012

Wu CX, Wang S, 2014. PLE-wu, a new member of *piggyBac* transposon family from insect, is active in mammalian cells. *J Biosci Bioeng*, 118(4):359-366. https://doi.org/10.1016/j.jbiosc.2014.03.010

Xu HF, Xia QY, Liu C, et al., 2006. Identification and characterization of *piggyBac*-like elements in the genome of domesticated silkworm, *Bombyx mori*. *Mol Genet Genomics*, 276(1):31-40. https://doi.org/10.1007/s00438-006-0124-x

Xue WH, Xu N, Yuan XB, et al., 2018. CRISPR/Cas9-mediated knockout of two eye pigmentation genes in the brown planthopper, *Nilaparvata lugens* (Hemiptera: Delphacidae). *Insect Biochem Mol Biol*, 93:19-26. https://doi.org/10.1016/j.ibmb.2017.12.003

Zhao Y, Huang G, Zhang WQ, 2019. Mutations in *Nl*InR1 affect normal growth and lifespan in the brown planthopper *Nilaparvata lugens*. *Insect Biochem Mol Biol*, 115: 103246. https://doi.org/10.1016/j.ibmb.2019.103246

Supplementary information

Tables S1–S3; Figs. S1–S4