Alteration of the *Ping* and *Pong* ORF1 Proteins Allows for Hyperactive Transposition of *mPing*

Lisette Payero[1], Giselle Outten[1,2], Courtney Burckhalter[1,3] and C. Nathan Hancock[1*]

[1] *Department of Biology and Geology, University of South Carolina Aiken*
[2] *Current Affiliation: Faculty of Medical Sciences, The University of the West Indies, Mona, Jamaica*
[3] *Current Affiliation: Savannah River National Laboratory, Savannah River Site, South Carolina*

The ORF1 and Transposase (TPase) proteins are essential for mobilizing the *PIF/Harbinger* transposable elements, including the *mPing* element from rice. Even though *mPing* is derived from the larger *Ping* element, it is mobilized more effectively by the ORF1 and TPase proteins encoded by the *Pong* element. This study aimed to identify the domains of ORF1 responsible for this difference in activity, and to create an optimized ORF1 construct. Three recombinant ORF1 constructs derived from *Ping* ORF1 and *Pong* ORF1 domains were made and tested in a yeast transposition assay. The results of these assays indicated that the Nuclear Localization Signal (NLS) containing region from *Pong* ORF1 contributed significantly to the increased activity. To further enhance this effect, a strong NLS was added, which resulted in even higher transposition. Testing of this high activity construct (Shuffle1 NLS) in Arabidopsis showed that this increase in activity can also be observed in plants. These results provide additional evidence that access of the ORF1 and TPase proteins to the nucleus is a limiting factor in *mPing* transposition. The domain swapping results also suggested that the *Ping* Myb-like binding domain was more effective than its *Pong* counterpart, and was consistent with the ORF1 C-terminal interacting with TPase. The role of a short repetitive sequence, present in the N-terminal *Ping* ORF1, but absent from *Pong* ORF1, was also tested. Removing one or both copies of this sequence, resulted in significantly higher activity than controls. Thus, by removing one or both copies of the *Ping* ORF1 repetitive sequence and using the *Ping* Myb domain, a strong NLS, and the *Pong* C-terminal together with *Pong* TPase resulted in about 50 times more *mPing* transposition events. This optimized construct will increase the effectiveness of *mPing*-based gene discovery tools. The finding that the ORF1 protein can be greatly optimized is consistent with the native proteins being adapted to only produce moderate activity, regulating transposition and preventing excessive host damage.

Introduction

Transposable elements are mobile segments of DNA that move from one genomic location to another[1]. Class II transposable elements (DNA transposons) transpose via a "cut and paste" mechanism, in which the transposable element is excised out of its genomic location and inserted into a new one by transposase proteins [2, 3]. These transposase proteins facilitate transposition by binding to the terminal inverted repeats (TIRs) of the transposable element using a DNA binding domain, and cleaving it from its genomic location using a catalytic domain [4]. Transposable elements constitute a large portion of most genomes, and play a large role in genome evolution due to their ability to induce mutations and alter genome size [5]. Determining the mechanisms that control transposable element activity provides clues about how these elements interact with their hosts. The activity and abundance of these elements in plants has also allowed for transposable element mutagenesis systems to be developed for a wide variety of crop species including maize, rice, soybean, and *Medicago* [6-10]. Understanding the factors that influence transposition activity is important to fully utilize transposable elements in gene discovery research.

Transposable elements are divided into transposon superfamilies based on their transposition and sequence characteristics [11]. One such superfamily is the *PIF/Harbinger* family, which is present in the genome of many plant and animal species [12]. Members of the *PIF/Harbinger* superfamily encode two proteins, ORF1 and TPase, that are thought to function together to mobilize the transposable element[12]. The ORF1 protein has a Myb-like domain believed to be essential in DNA binding, while the TPase protein contains a catalytic domain for DNA cleavage [12, 13]. Experiments in Arabidopsis, human cells, and *Saccharomyces cerevisiae* show that both of these proteins are required for mobilization of *PIF/Harbinger* elements [13-15]. The ORF1 protein is of particular interest to this study because it has been shown to exhibit more sequence divergence than the TPase protein [12].

Two naturally active members of the *PIF/Harbinger* superfamily are the closely related *Ping* and *Pong* elements from rice [16-18]. The miniature inverted repeat transposable element (MITE) *mPing* is a deletion derivative of the *Ping* element, and does not encode its own ORF1 and TPase. In order to transpose, *mPing* utilizes the ORF1 and TPase proteins from either *Ping* or *Pong* [13, 14, 16, 19, 20]. In some rice cultivars, *mPing* is highly active, reaching over 1000 copies and about 40 new insertions per generation in the Gimbozu line [17, 18, 21].

Previous experiments in yeast have shown that the ORF1 and TPase proteins encoded by the *Pong* element are more effective at mobilizing *mPing* than their *Ping* counterparts [13]. This is surprising given that *mPing* is derived from the *Ping* element and suggests that the *Ping* element proteins are not optimized for transposition. Our goal was to determine the nature of this difference in ability to mobilize *mPing*, with special focus on the ORF1 proteins, as they show more sequence divergence compared to the TPase proteins [*Ping* vs. *Pong* amino acid identity for ORF1 = 66.1%, TPase = 76.7%] [22]. Analysis of the *Ping* and *Pong* ORF1 proteins shows that while the Myb-like domain is highly conserved, only the *Pong*-encoded ORF1 protein contains predicted bipartite nuclear localization signals (NLSs) after the Myb domain (Figure 1). Despite this, experiments in yeast show that the *Ping* ORF1 is recruited to the nucleus [13]. For comparison, the ORF1 (Myb-like) protein from the *HARBI1* element from zebrafish contains a predicted NLS and it was shown to be localized to the nucleus, bringing its associated TPase (TNp) protein along with it [15]. Thus, we hypothesized that the *Pong* ORF1 NLS is stronger than any cryptic NLS present in *Ping* ORF1, and thus, more efficiently functions to recruit ORF1 and TPase to the nucleus. This hypothesis is consistent with previous studies that showed that mutation of the Nuclear Export Signals encoded by the *Ping* and

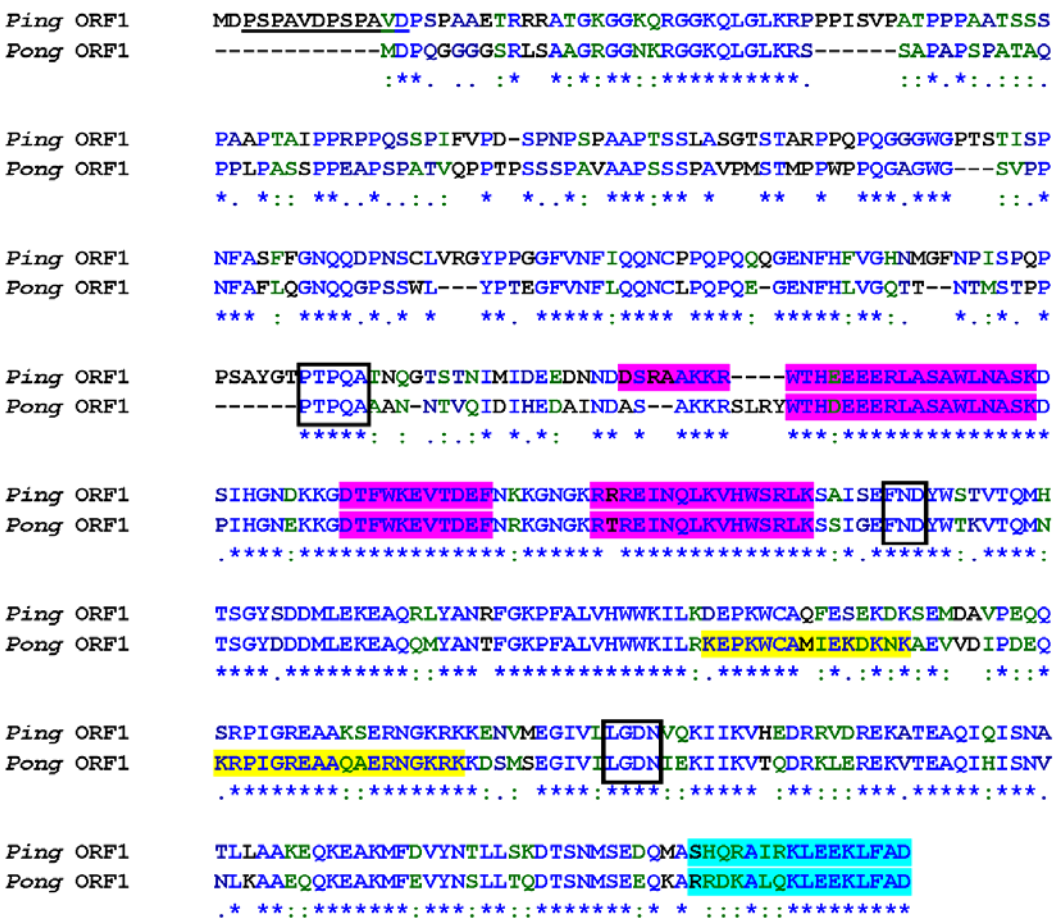


Figure 1. Amino acid sequence alignment of *Ping* ORF1 and *Pong* ORF1. CLUSTAL W [32] multiple sequence alignment with fully conserved residues shown in blue (*), conserved strong groups in green (:), and conserved weak groups in purple (.). Selected domains found with MyHits [33] are highlighted: Myb-like domain = pink, NLS = yellow, PFTA-like domain = cyan. Boxes indicate regions where the sequence was divided into sections to create recombinant versions. The repetitive region from *Ping* ORF1 is underlined.

Pong TPases increases transposition activity [13]. We predicted that domain swapped ORF1 constructs with the *Pong* ORF1 NLS would show increased ability to mobilize *mPing* compared to those with the corresponding region from *Ping* ORF1. Sequence analysis also showed that the *Ping* ORF1 protein has a short repetitive sequence not present in *Pong* ORF1 (Figure 1). We hypothesized that the extra copy of the repetitive sequence of interest found in *Ping* ORF1 was due to a replication error and may actually disrupt the structure of the protein. Thus, we predicted that removing the extra repeat would increase ORF1's ability to mobilize *mPing*.

To test these hypotheses and determine the factors that make *Pong* ORF1 more effective than *Ping* ORF1, we created domain swapped constructs from *Ping* and *Pong* ORF1 sequences. ORF1 constructs with strong canonical NLSs were also developed and tested in both yeast and Arabidopsis. The role of the *Ping* ORF1 repetitive sequence on *mPing* transposition was determined by comparing the transposition rates produced by ORF1 proteins with one, two and zero copies of the sequence. Together these experiments provided information about the transposition mechanism of *mPing* and allowed for the optimization the ORF1 protein to more effectively mobilize *mPing*.

Materials and Methods

ORF1 domain swapping and addition of a strong NLS

A *Ping* ORF1 fragment 3-4 NLS fragment (302bp) that contained 15 base changes, resulting in a stronger NLS was synthesized as a gBlocks® Gene Fragment (Integrated DNA Technologies, Coralville, Iowa) [Supplemental Figure 1A]. Domain swapped mutants were created utilizing USER friendly DNA recombination as described [23]. DNA fragments with complementary overlapping ends were created by performing PfuTurbo Cx Hotstart PCR (Agilent Technologies, Santa Clara, CA) with *Ping* ORF1, *Pong* ORF1, or the *Ping* ORF1 fragment 3-4 NLS template and the appropriate uracil containing primers (Table 1). The USER enzyme (New England Biolabs, Ipswich, MA) was used to excise the uracil residue from the PCR product, creating complementary overhangs that were joined by ligation. The resulting product was amplified using high fidelity PCR and the full length ORF1 sequence was selected using gel purification. A Gateway® pDONR ZEO entry clone was created by performing a BP Clonase® II reaction. An LR Clonase® II reaction (Invitrogen, Waltham, MA) was then performed using pAG413GAL-ccdB (Susan Lindquist, Addgene #14141) to create an expression clone [24].

Table 1. Primers used for USER friendly DNA recombination

Fragment Amplified	Forward Primer	Sequence	Reverse Primer	Sequence
Ping ORF1 Frag 1	*Ping* ORF1 ATTB F	GGGGACAAGTTTGTACAAAAA AGCAGGCTTCATGGATCCGTCG CCGGCCGTGGATCCGTCGCCGG CCGTGGATCCGTCGCCGGCT	*Ping* O1F1 Rev	AGCTTGGGGUGTTGGTGTT
Ping ORF1 Frag 2	*Ping* O1F2 For	ACCCCAAGCUACGAACCAA	*Pong* O1F2 Rev	ATTTGAGUTACCTTAGTCCA GTAATCATTG
Ping ORF1 Frag 3-4	*Ping* O1F3 For	ACTCAAAUGCATACAAGCGGA	*Ping* O1F4 Rev	ATAATTTUCTGGACATTGTC CCC
Ping ORF1 Frag 5	*Ping* O1F5 For	AAAATTAUAAAGGTCCACGAA GACC	*Ping* ORF1 ATTB R	GGGGACCACTTTGTACAAGA AAGCTGGGTCCTTTAATCCGC AAATAGCTTCTCCTCT
Pong ORF1 Frag 1	*Pong* ORF1 ATTB F	GGGGACAAGTTTGTACAAAAAA GCAGGCTTCATGGATCCACAAG GAGGTGG	*Pong* O1F1 Rev	AGCTTGGGGUGTTGGTGG
Pong ORF1 Frag 2	*Pong* O1F2 For	ACCCCAAGCUGCAGCTAAC	*Ping* O1F2 Rev	ATTTGAGUAACCGTACTCCA ATAGTCA
Pong ORF1 Frag 3-4	*Pong* O1F3 For	ACTCAAAUGAATACAAGCGGAT ATG	*Pong* O1F4 Rev	ATAATTTUTTCAATATTGTCCC CTAGG
Pong ORF1 Frag 5	*Pong* O1F5 For	AAAATTAUCAAAGTGACGCAAGA TCG	*Pong* ORF1 ATTB R	GGGGACCACTTTGTACAAGAAA GCTGGGTCTTAGTCAGCAAATA ACTTTTCCTCCA

Arabidopsis Transposition Assay

The Ribosomal Protein S5a (Rps5a) promoter [25] was cleaved out of dpGreen RPS5a-tdTomato (gift from Dr. Wolfgang Lukowitz, University of Georgia) by digestion with the *Cla*I and *Eco*RV and cloned into *Xho*I and *Stu*I digested pEarleyGate 100 [26] after T4 polymerase blunting to create pEG100R. The *Pong* ORF1 or Shuffle1 NLS ORF1 construct was inserted into pEG 100R with an LR Clonase®II (Invitrogen, Waltham, MA) reaction. The GmUbi3 promoter: *Pong* TPase L418A, L420A: OCS Terminator expression cassette was digested from pUQ213 GmUbi *Pong* (gift from Dr. Robert Stupar, University of Minnesota) with *Hind*III and *Sca*I, T4 polymerase blunted, and inserted into the *Pme*I site of pEG100R to produce the pWMD constructs. These plasmid were transformed into Arabidopsis containing the previously described *mPing*:GFP reporter construct [13, 14] by the floral dip method [27]. Plants were sprayed with 1:1000 dilution of Finale® herbicide (Bayer, Leverkusen, Germany) in order to select for the BAR gene. The T1 plants that survived this selection were examined using an Olympus SZX12 dissecting microscope to detect GFP expression.

Removal of the *Ping* ORF1 repeat sequence

Thermo Scientific ® Phusion (Thermo Scientific, Waltham, MA) high fidelity PCR of Shuffle1 NLS was performed using a forward primer (ORF1SC1 ATTB For short – GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGATC CGTCGCCGGCCG) that only included the first copy of the repeat (allowing for priming at multiple positions). The PCR products were then cloned into pDONR Zeo using a BP Clonase®II (Invitrogen, Waltham, MA) reaction. Colonies were sequenced to identify a clone with the desired one repeat variation and no other mutations and denoted as ORF1SC1 ONE. An LR Clonase®II (Invitrogen, Waltham, MA) reaction was performed to move ORF1SC1 ONE into pAG413GAL-ccdB. The ORF1SC1 no repeat (ORF1SC1 NO) construct was created by digesting Shuffle1 NLS with *Bam*HI to remove the repetitive sequences, then ligating the DNA back together.

Yeast Transposition Assay

Novel and previously described [pAG415 *Ping* TPase L384A, L386A, pAG415 *Pong* TPase L418A, L420A, pAG413 Ping ORF1, pAG413 Pong ORF1[13]] constructs were transformed into CB101 yeast [28] and a 100µl was plated onto both YPD (10^{-4} dilution) and galactose CSM-His-Leu-Ura-Ade. The YPD plates were used to determine the total number of yeast cells plated, while the Galactose plates allowed for selection of cells in which *mPing* had transposed out of the *ade2* gene. The YPD plates were incubated for two days, while the galactose plates were incubated for ten. The number of colonies were recorded and used to calculate transposition frequency.

Results and Discussion

ORF1 domain swapping

To determine which regions of the *Ping* and *Pong* ORF1 proteins are contributing to the difference in transposition, three domain swapped constructs (Figure 2) were created using the USER friendly recombination method [23]. The first construct, Shuffle1, is the *Ping* ORF1 sequence except the C-terminal domain has been replaced with the *Pong* sequence. The second, Shuffle2, roughly contained the N-terminal half of *Ping* ORF1 and the C-terminal half of *Pong* ORF1, including the *Pong* NLS sequence. The third, Shuffle3, consisted of the *Pong* N-terminal domain, *Ping* Myb-like domain, *Pong* NLS region, and the *Ping* C-terminal. These constructs were tested for their ability to induce transposition with the *Ping* and *Pong* TPases [lacking the nuclear export signal, L384A, L386A and L418A, L420A respectively [13]] (Figure 2). All three recombinant constructs exhibited significantly higher levels of *mPing* transposition than the *Ping* ORF1 protein. Shuffle2 with *Pong* TPase mobilized *mPing* at significantly higher rates than all other combinations.

The increased activity shown for Shuffle1 compared to *Ping* ORF1 suggests that the *Pong* ORF1 C-terminal facilitates transposition better than the *Ping ORF1* C-terminal. We also noted that this construct had higher transposition with the *Pong* TPase than the *Ping* TPase, consistent with the previous results that indicate that the C-terminal is responsible for dimerization between ORF1 and TPase [15].

The shuffle constructs containing the *Pong* NLS (Shuffle2 and Shuffle3) show even higher transposition than the Shuffle1 construct, supporting the hypothesis that the stronger *Pong* ORF1

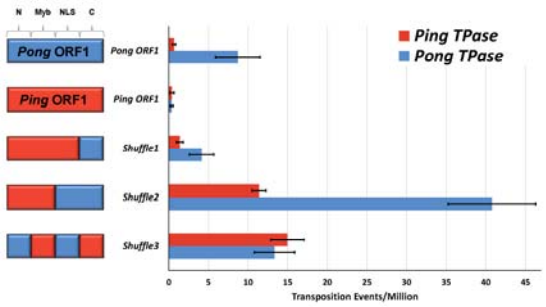


Figure 2. *mPing* transposition activity utilizing wild type and recombinant ORF1 proteins. Diagrams representing the ORF1 constructs and the locations of domains are shown on the left. *Ping* derived sequences are shown in red and *Pong* sequences are shown in blue (N = N-terminal, Myb = Myb-like domain, NLS = NLS Region, C = C-terminal). Results of the yeast transposition assay using *Ping* TPase (red) and *Pong* TPase (blue) are shown on the right. The chart shows the average transposition events per million cells for six replicates and the error bars indicate standard error.

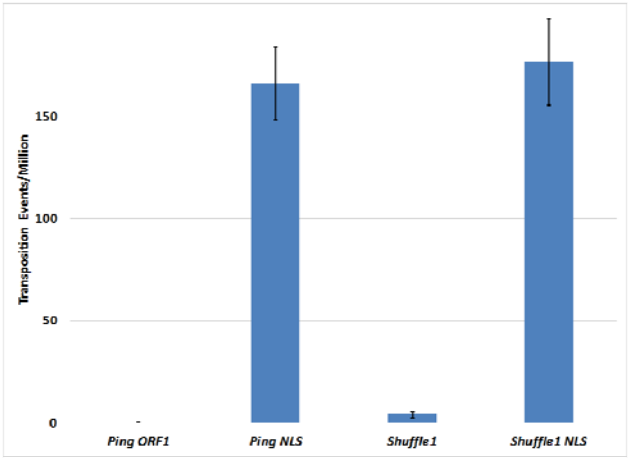


Figure 3. *mPing* transposition activity of ORF1 with or without a strong NLS. Transposition rates of *Ping* ORF1 and Shuffle1 proteins with either the native *Ping* encoded NLS region or a mutated version containing a strong NLS (*Ping* NLS and Shuffle1 NLS). *Pong* TPase was used for all treatments. The chart shows the average transposition events per million cells for six replicates and the error bars indicate standard error.

NLSs contribute to the difference in performance between the *Ping* and *Pong* ORF1 protein. The fact that the Shuffle2 construct was the most effective, especially with *Pong* TPase, is consistent with it containing both the *Pong* NLSs and the *Pong* C-terminal domain. The fact that its activity was even higher than the *Pong* ORF1 suggests that the Myb-like domain of *Ping* ORF1 may have a stronger *mPing* binding capacity than that of the *Pong* Myb-like domain.

The decreased activity of the Shuffle3 construct compared to Shuffle2 when utilizing *Pong* TPase corresponds with our previous hypothesis that the *Ping* C-terminal domain is not as optimized for transposition. Surprisingly, the Shuffle3 construct exhibited equal performance with either the *Ping* or *Pong* TPase, unlike the other constructs. In fact, with *Ping* TPase the Shuffle3 protein mobilized *mPing* at slightly higher rates than Shuffle2 with *Ping* TPase. This higher than expected activity with the *Ping* TPase is possibly the result of better cooperation between the *Ping* C-terminal and the *Ping* TPase protein, but could also be a result of including the *Pong* N-terminal region. To differentiate between these two possibilities explanations, additional experimentation will be needed.

Addition of Strong NLS

The results of the domain swapping experiment confirmed that the quality of the NLS is largely responsible for ORF1 performance. The overall consensus bipartite sequence that has been identified is made up of two adjacent basic amino acids (R or K) followed by a 10 amino acid spacer and then 3 to 5 basic residues in the next 5 amino acids [29]. The *Pong* ORF1 NLS sequence only has three basic residues, so only barely fall within these qualifications. Therefore, it was hypothesized that converting the *Ping* ORF1 encoded sequence into a strong NLS sequence would result in higher transposition activity. To achieve this, seven amino acids were mutated to form two complete bipartite NLS sequences, each with a full complement of basic residues (Supplemental Figure 1B). Addition of this strong NLS sequence resulted in dramatically higher performance of both the *Ping* ORF1 and Shuffle1 ORF1 proteins (Figure 3). These results support the hypothesis that protein localization is directly correlated to transposition activity. However, addition of a strong NLS signal to the *Ping* TPase lacking the nuclear export signal (L384A, L386A) did not result in any additional increase in transposition activity (data not shown). Together this suggests that TPase protein is effectively recruited to the nucleus together with ORF1.

To confirm that the addition of the NLS represents a biologically relevant change that would also occur in plants, the Shuffle1 NLS construct was tested in *Arabidopsis*. Matching constructs with *Pong* TPase L418A, L420A expressed with the *Glycine max* ubiquitin 3 promoter [30] and either *Pong* ORF1 or Shuffle1 NLS expressed from the Ribosomal Protein S5a promoter [25] were made (Figure 4). These constructs were transformed into plants containing the *mPing*:GFP reporter [14]. In this assay, excision of *mPing* allows for GFP expression. After selection of transgenic events, T1 plants transformed with the Shuffle1 NLS construct exhibited a significantly (χ^2 Test Stat = 7.273 [1 d.f.], P value = 0.007) higher rate of GFP expression (15/32 = 46.9%) than those transformed with the *Pong* ORF1 construct (5/32 = 15.6%). This result confirms that despite differences in nuclear envelope behavior during cell division, addition of a strong NLS increases the efficiency of entering the nucleus to gain access to the DNA substrate.

Removal of the *Ping* ORF1 repeat sequence

After deciding to use the Shuffle1 NLS construct because of its high transposition rate, we tested the role of the *Ping* ORF1 repetitive sequence found at the N-terminal of this construct (Figure 1). Two additional constructs, the first with one copy of the repeat removed (ORF1SC1 ONE), and the second with both iterations of the repeat sequence removed (ORF1SC1 NO), were made. The results of the yeast transposition assay utilizing these constructs are shown in Figure 5. Both ORF1SC1 ONE and ORF1SC1 NO mobilized *mPing* at significantly higher rates than the Shuffle1 NLS control, though they did not differ significantly from each other. This result indicates that the repetitive sequence found on the *Ping* ORF1 protein acts to hinder the protein's ability to mobilize *mPing*.

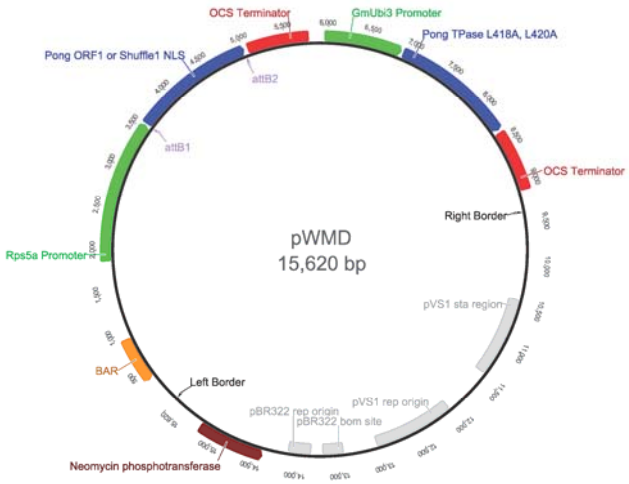


Figure 4. Plasmid for mobilization of *mPing* in plants. Plasmid map of pWMD generated using Geneious software[34]. The tDNA that inserts into the genome is composed of the region between the left and right borders. Promoters are shown in green, genes of interest are shown as blue arrows, terminators are shown in red, and the selectable marker gene (BAR) is shown in yellow.

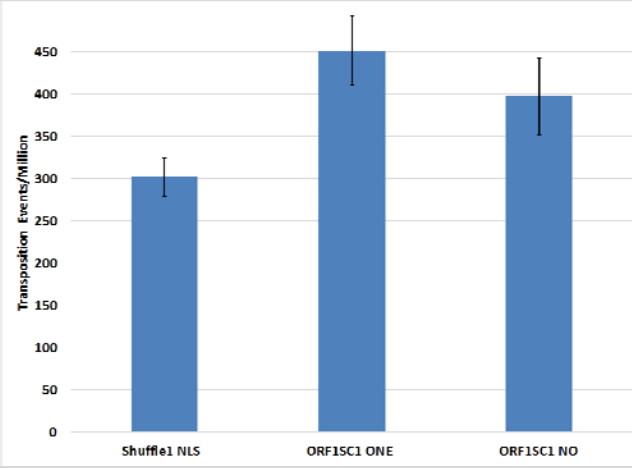


Figure 5. *mPing* transposition activity with zero, one, or two copies of the *Ping* ORF1 repetitive sequence. Chart comparing the Shuffle1 NLS (control) transposition rate, to version with one (ORF1SC1 ONE) or no (ORF1SC1 NO) copies of the *Ping* ORF1 repetitive sequence. Each bar represents the average transposition events per million cells for six replicates and the error bars indicate standard error.

Conclusion

This study provides insight into the factors that determine the effectiveness of the ORF1 protein. The results of this study show that the ORF1 protein can be optimized by creating a construct composed of the N-terminal of *Ping* ORF1 and the C-terminal of *Pong* ORF1, adding a strong nuclear localization signal, and removing one or both copies of the short repetitive sequence present within *Ping* ORF1. The resulting constructs, ORF1SC1 ONE and ORF1SC1 NO, mobilize *mPing* at significantly higher rates than either the wild type *Ping* or *Pong* ORF1 proteins. These constructs can be used to increase the transposition efficiency when using *mPing* as a research tool.

Together these results also provide evidence that the *Ping* and *Pong* ORF1 proteins are not optimized for transposition. *Pong* and *Ping* ORF1 do not contain the strongest possible NLSs and *Ping* ORF1's repetitive sequence inhibits transposition. This suggests that the presence of these underperforming domains may serve as self-regulatory mechanisms, acting to prevent high amounts of transposition that would damage the host genome. The fact that different elements contain different self-regulatory mechanisms may actually provide a partial explanation for the "burst" of transposition observed for some transposons including *mPing* [31]. Specifically, we hypothesize that the *mPing* element showed a drastic increase in transposition by overcoming the inherent regulatory mechanisms of the *Ping* element by reducing its size and using the *Pong* ORF1 and *Pong* TPase proteins for mobilization.

Acknowledgements

We would like to thank Ashley Strother for technical assistance with the experiments. We also would like to thank David Gilbert for carful editing of the manuscript. Some of the funding for this work was from was provided by NSF Award # 1127083 - A Genetic Resource for Gene Discovery in Soybean and NSF Award # 1444581 – A Resource for Functional Genomics to Support Soybean Genetics and Breeding.

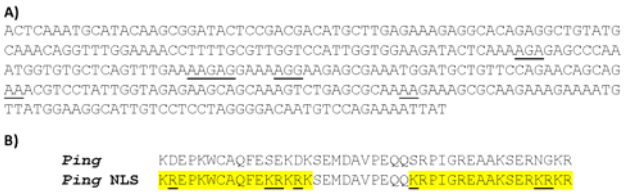
Notes and References

**Corresponding author email: nathanh@usca.edu*

1. Craig NL. 2002. *Mobile DNA II*. ASM Press, Washington, D.C.
2. Feschotte C and Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41**: 331-368.
3. Munoz-Lopez M and Garcia-Perez JL. 2010. DNA transposons: Nature and applications in genomics. *Curr Genomics* **11**: 115-128.
4. Rice PA and Baker TA. 2001. Comparative architecture of transposase and integrase complexes. *Nat Struct Biol* **8**: 302-307.
5. Oliver KR, McComb JA and Greene WK. 2013. Transposable elements: Powerful contributors to angiosperm evolution and diversity. *Genome Biol Evol* **5**: 1886-1901.
6. Lazarow K, Du ML, Weimer R and Kunze R. 2012. A hyperactive transposase of the maize transposable element *Activator* (*Ac*). *Genetics* **191**: 747-U146.
7. Tadege M, Wen JQ, He J, Tu HD, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M, Ratet P and Mysore KS. 2008. Large-scale insertional mutagenesis using the *Tnt1* retrotransposon in the model legume *Medicago truncatula*. *Plant J* **54**: 335-347.
8. Hancock CN, Zhang F, Floyd K, Richardson AO, LaFayette P, Tucker D, Wessler SR and Parrott WA. 2011. The rice miniature inverted repeat transposable element *mPing* is an effective insertional mutagen in soybean. *Plant Physiol* **157**: 552-562.
9. Hirochika H. 2001. Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr Opin Plant Biol* **4**: 118-122.
10. Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajardo DS, Wroclawska E, Tseung CW, Lai JS, Hunter CT, Avigne WT, Baier J, Messing J, Hannah LC, Koch KE, Becraft PW, Larkins BA and McCarty DR. 2007. Sequence-indexed mutations in maize using the *UniformMu* transposon-tagging population. *Bmc Genomics* **8**: 116.
11. Zhao DY, Ferguson AA and Jiang N. 2016. What makes up plant genomes: The vanishing line between transposable elements and genes. *Bba-Gene Regul Mech* **1859**: 366-380.

12. Zhang XY, Jiang N, Feschotte C and Wessler SR. 2004. *PIF*- and *Pong*-like transposable elements: Distribution, evolution and relationship with *Tourist*-like miniature inverted-repeat transposable elements. *Genetics* **166**: 971-986.
13. Hancock CN, Zhang F and Wessler SR. 2010. Transposition of the *Tourist*-MITE *mPing* in yeast: An assay that retains key features of catalysis by the Class 2 *PIF/Harbinger* superfamily. *Mobile DNA* **1**: 5.
14. Yang GJ, Zhang F, Hancock CN and Wessler SR. 2007. Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *P Natl Acad Sci USA* **104**: 10962-10967.
15. Sinzelle L, Kapitonov VV, Grzela DP, Jursch T, Jurka J, Izsvak Z and Ivics Z. 2008. Transposition of a reconstructed *Harbinger* element in human cells and functional homology with two transposon-derived cellular genes. *P Natl Acad Sci USA* **105**: 4715-4720.
16. Jiang N, Bao ZR, Zhang XY, Hirochika H, Eddy SR, McCouch SR and Wessler SR. 2003. An active DNA transposon family in rice. *Nature* **421**: 163-167.
17. Naito K, Cho E, Yang GJ, Campbell MA, Yano K, Okumoto Y, Tanisaka T and Wessler SR. 2006. Dramatic amplification of a rice transposable element during recent domestication. *P Natl Acad Sci USA* **103**: 17620-17625.
18. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T and Wessler SR. 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* **461**: 1130-1134.
19. Kikuchi K, Terauchi K, Wada M and Hirano HY. 2003. The plant MITE *mPing* is mobilized in anther culture. *Nature* **421**: 167-170.
20. Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H and Tanisaka T. 2003. Mobilization of a transposon in the rice genome. *Nature* **421**: 170-172.
21. Yasuda K, Ito M, Sugita T, Tsukiyama T, Saito H, Naito K, Teraishi M, Tanisaka T and Okumoto Y. 2013. Utilization of transposable element *mPing* as a novel genetic tool for modification of the stress response in rice. *Mol Breeding* **32**: 505-516.
22. Myers EW and Miller W. 1988. Optimal alignments in linear-space. *Comput Appl Biosci* **4**: 11-17.
23. Villiers BRM, Stein V and Hollfelder F. 2010. USER friendly DNA recombination (USERec): A simple and flexible near homology-independent method for gene library construction. *Protein Eng Des Sel* **23**: 1-8.
24. Katzen F. 2007. Gateway recombinational cloning: a biological operating system. *Expert Opin Drug Dis* **2**: 571-589.
25. Weijers D, Franke-van Dijk M, Vencken RJ, Quint A, Hooykaas P and Offringa R. 2001. An Arabidopsis Minute-like phenotype caused by a semi-dominant mutation in a *RIBOSOMAL PROTEIN S5* gene. *Development* **128**: 4289-4299.
26. Earley KW, Haag JR, Pontes O, Opper K, Juehne T, Song KM and Pikaard CS. 2006. Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J* **45**: 616-629.
27. Clough SJ and Bent AF. 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**: 735-743.
28. Gilbert DM, Bridges MC, Strother AE, Burckhalter CE, Burnette JM and Hancock CN. 2015. Precise repair of *mPing* excision sites is facilitated by target site duplication derived microhomology. *Mobile DNA* **6**: 15.
29. Dingwall C and Laskey RA. 1991. Nuclear targeting sequences - a consensus. *Trends Biochem Sci* **16**: 478-481.
30. Chiera JM, Bouchard RA, Dorsey SL, Park E, Buenrostro-Nava MT, Ling PP and Finer JJ. 2007. Isolation of two highly active soybean (*Glycine max* (L.) Merr.) promoters and their characterization using a new automated image collection and analysis system. *Plant Cell Reports* **26**: 1501-1509.
31. Naito K, Mondee Y, Yasuda K, Saito H and Okumoto Y. 2014. *mPing*: The bursting transposon. *Breeding Sci* **64**: 109-114.
32. Thompson JD, Higgins DG and Gibson TJ. 1994. Clustal-W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-4680.
33. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, Martin O, Kuznetsov D and Falquet L. 2007. MyHits: Improvements to an interactive resource for analyzing protein sequences. *Nucleic Acids Res* **35**: W433-W437.
34. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P and Drummond A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647-1649.

Supplemental Material



Supplemental Figure 1. Sequence changes resulting in a strong NLS. A) Nucleotide sequence of *Ping* ORF1 Fragment 3-4 NLS used to make the NLS versions of ORF1. Underlined sequences indicate bases that were changed from the *Ping* ORF1 sequence. B) Amino acid sequence of the strong NLS compared to the original *Ping* ORF1. Altered residues are underlined and the NLS sequence is highlighted in yellow.