Development of mPing-based activation tags for crop insertional mutagenesis

Alexander Johnson
University of Georgia, Athens

Edward Mcassey
University of Georgia, Athens

Stephanie Diaz
University of South Carolina - Aiken

Jacod Reagin
University of South Carolina - Aiken

Priscilla S. Redd
University of South Carolina - Aiken

See next page for additional authors
Follow this and additional works at: https://digitalcommons.unl.edu/agronomyfacpub

Part of the Agricultural Science Commons, Agriculture Commons, Agronomy and Crop Sciences Commons, Botany Commons, Horticulture Commons, Other Plant Sciences Commons, and the Plant Biology Commons

Johnson, Alexander; Mcassey, Edward; Diaz, Stephanie; Reagin, Jacod; Redd, Priscilla S.; Parrilla, Daymond R.; Nguyen, Hanh; Stec, Adrian; McDaniel, Lauren A.L.; Clemente, Thomas E.; Stupar, Robert M.; Parrott, Wayne A.; and Hancock, C. Nathan, "Development of mPing-based activation tags for crop insertional mutagenesis" (2021). Agronomy & Horticulture -- Faculty Publications. 1433. https://digitalcommons.unl.edu/agronomyfacpub/1433

This Article is brought to you for free and open access by the Agronomy and Horticulture Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Agronomy & Horticulture -- Faculty Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Authors
Alexander Johnson, Edward Mcassey, Stephanie Diaz, Jacod Reagin, Priscilla S. Redd, Daymond R.
Parrilla, Hanh Nguyen, Adrian Stec, Lauren A.L. McDaniel, Thomas E. Clemente, Robert M. Stupar, Wayne
A. Parrott, and C. Nathan Hancock
Development of mPing-based activation tags for crop insertional mutagenesis

Alexander Johnson1 | Edward Mcassey1 | Stephanie Diaz2 | Jacob Reagin2 | Priscilla S. Redd2 | Daymond R. Parrilla2 | Hanh Nguyen3 | Adrian Stec4 | Lauren A. L. McDaniel1 | Thomas E. Clemente3 | Robert M. Stupar4 | Wayne A. Parrott1 | C. Nathan Hancock2

1Institute of Plant Breeding, Genetics & Genomics/Center for Applied Genetic Technologies, University of Georgia, Athens, GA, USA
2Department of Biology and Geology, University of South Carolina Aiken, Aiken, SC, USA
3Department of Agronomy and Horticulture/Center for Plant Science Innovation, University of Nebraska, Lincoln, NE, USA
4Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, USA

Correspondence
C. Nathan Hancock, Department of Biology and Geology, University of South Carolina Aiken, Aiken, South Carolina 29801, USA.
Email: nathanh@usca.edu

Present address
Edward Mcassey, School of Life Sciences, University of Hawai’i at Mānoa, Honolulu, HI, USA
Stephanie Diaz, Department of Biochemistry, Purdue University, West Lafayette, IN, USA
Daymond R. Parrilla, Department of Molecular and Comparative Pathobiology, Johns Hopkins School of Medicine, Baltimore, MD, USA

Funding information
This study was supported by NSF Plant Genome Research Project Grant # 0820769, 1127083, and 1444581.

Abstract
Modern plant breeding increasingly relies on genomic information to guide crop improvement. Although some genes are characterized, additional tools are needed to effectively identify and characterize genes associated with crop traits. To address this need, the mPing element from rice was modified to serve as an activation tag to induce expression of nearby genes. Embedding promoter sequences in mPing resulted in a decrease in overall transposition rate; however, this effect was negated by using a hyperactive version of mPing called mmPing20. Transgenic soybean events carrying mPing-based activation tags and the appropriate transposase expression cassettes showed evidence of transposition. Expression analysis of a line that contained a heritable insertion of the mmPing20F activation tag indicated that the activation tag induced overexpression of the nearby soybean genes. This represents a significant advance in gene discovery technology as activation tags have the potential to induce more phenotypes than the original mPing element, improving the overall effectiveness of the mutagenesis system.

KEYWORDS
activation tagging, gene discovery, soybean, transposable element
Genetic improvement in crop species is facilitated by identification of genes that control important agricultural traits (Thomson et al., 2010). Using the model plant Arabidopsis, multiple T-DNA-based insertion mutagenesis populations have been effectively used for gene discovery in both phenotypic screens and large-scale reverse genetic approaches (Alonso et al., 2003; Kearsley, 2005; Maiti et al., 1997; Maiti et al., 1997; McElroy et al., 1997). The first advantage of insertional mutagenesis over other mutagenesis techniques is that the inserted sequence, often called a tag, can be used to anchor PCR-based strategies for identifying its genomic location (McElroy et al., 1997; Parinov et al., 1999). The second advantage is the ability to include sequences that interact with neighboring genes in the tag. These advanced tags contain promoter elements to induce overexpression of nearby genes [activation tags] (Dong & Von Arnim, 2003; Kakimoto, 1996; Weigel et al., 2000; Wilson et al., 1996), reporter genes with splice adaptors to produce fusion proteins [gene-trap tags] (Skarnes, 1990; Sundaresan et al., 1995), or reporter genes with minimal promoters for reporting the expression patterns of nearby promoter sequences [enhancer-trap tags] (Sundaresan et al., 1995; Wilson et al., 1990). Each type of tag provides additional resources for gene function analysis and are especially important for understanding the function of essential and redundant genes that would not otherwise produce loss-of-function phenotypes. These advanced tags are critical for gene discovery in polyploid or other highly duplicated genomes including wheat and soybean.

A significant factor contributing to the success of the Arabidopsis insertional mutagenesis projects is the availability of inexpensive high-throughput transformation techniques (Bechtold & Bouchez, 1995; Clough & Bent, 1998). However, high-throughput T-DNA tagging is not feasible for most crop species because their transformation protocols are relatively time and labor intensive. Given this limitation, a number of insertional mutagenesis programs have been developed around either native or heterologous transposable elements with varying amounts of success (Aarts et al., 1995; Cui et al., 2013; Hancock et al., 2011; Mathieu et al., 2009; McCarty et al., 2005; Meissner et al., 2000). The outcomes of these programs have been hindered by the inherent transposition characteristics of the elements including preferential transposition to linked sites, low transposition rates, and tissue culture activation.

Our goal is to develop a transposon-based activation tagging system that overcomes these limitations. Activation tags designed around the Activator (Ac)/Dissociation (Ds) (Fladung & Polak, 2012; Fridborg et al., 1999; Schaffer et al., 1998; Suzuki et al., 2003; Wilson et al., 1996) and Enhancer/Suppressor (Marsch-Martinez, 2011) transposon systems engineered with constitutive promoters or enhancer sequences clearly demonstrate that modified transposons retain their mobility and also induce gene expression. These approaches have been used to clone a variety of genes including TINY (Wilson et al., 1996), LATE ELONGATED HYPOCOTYL (Schaffer et al., 1998), SHORT INTERNODES (Fridborg et al., 1999), high phenolic compound1-1 Dominant (Schneider et al., 2005), and a strictosidine synthase (Mathieu et al., 2009). The Ac/Ds-based elements preferentially transpose into linked sites in many species, which requires many transformation events or labor-intensive screening measures to saturate the genome with tags (Bancroft & Dean, 1993; Nakagawa et al., 2000; Qu et al., 2008; Suzuki et al., 2001; Vollbrecht et al., 2010). In an effort to solve some of the limitations of current transposon-based activation tagging systems, we designed an activation tagging system around the mPing element from rice due to its favorable transposition behavior.

One of the most promising transposon-based mutagenesis systems relies on the mPing transposon from Oryza sativa (Jiang et al., 2003; Kikuchi et al., 2003; Nakazaki et al., 2003). This element has been shown to exhibit high rates of transposition in select rice cultivars (Naito et al., 2006) and can produce heritable mutations in rice (Naito et al., 2009; Nakazaki et al., 2003; Teraishi et al., 1999). When mPing was transferred to Glycine max along with the Ping transposase proteins required for mobilization, it produced heritable insertions without tissue culture treatment (Hancock et al., 2011). Subsequent generations of these plants revealed heritable mutant phenotypes that are being analyzed. Characterization of mPing transposition behavior in Oryza sativa, Arabidopsis thaliana, Saccharomyces cerevisiae, and Glycine max has shown that it transposes to unlinked sites, preferentially inserts near genes, and avoids insertion into GC-rich regions, making it an attractive candidate element to distribute activation tags across the genome (Hancock et al., 2010, 2011; Naito et al., 2009; Yang et al., 2007).

Here, we report the first efforts to combine the favorable transposition behavior of mPing with the activation tagging strategy. Before implementing such advanced mPing-based tags into genetically recalcitrant crop species, it is important to evaluate mPing for its suitability as an engineered tag. To this end, mPing-based activation constructs were developed and tested in both yeast and soybean. These experiments gave insight into the biology of mPing transposition and lead to a functional mPing-based activation tag platform that can be used for gene discovery in important crops.
100 (Earley et al., 2006) with the Rps5a promoter sequence by ligation and then Gateway cloning in the ORF1 Shuffle 1 NLS sequence. The Pong TPase (L418A, L420A) expression construct was then cloned into the Pmel site. The pWL89A mPing, pWL89A mmPing20, pWL89A NL60, pWL89A 2xE, pWL89A 4xE, pWL89A mmPing20F, pWL89A mmPing20, pWMD23, and pEarleyGate103 mmPing20F plasmids are available through Addgene (#140006-140007, #145787-145795).

2.2 Yeast assays

Transposition assays on 100 mm plates were performed using the previously described pWL89a, pAG413 ORF1 Shuffle 1 NLS, and pAG415 Pong TPase L418A, L420A plasmids in the CB101 yeast strain [MATa ade2∆::hphMX4 his3A1 leu2∆0 met15∆0 ura3∆0 lys2∆::ADE2+] (Gilbert et al., 2015; Hancock et al., 2010; Payero et al., 2016). One-way ANOVA followed by Tukey’s multiple comparison test was performed using GraphPad Prism version 8.4.3 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com.

2.3 Error prone PCR

Apex Master Mix PCR reactions (50µl) were supplemented with MnCl2, MgCl2, dCTP, and dTTP to conduct manganese error-prone PCR as described (Cadwell & Joyce, 1992). The mPing template was amplified with mPing TTA For and mPing TTA Rev primers (Table S3) in two rounds of mutagenic PCR to generate a library. The library was cloned into the ADE2 gene by cotransforming it with HpaI digested pWL89a plasmid into yeast. A total of 112 clones were screened for transposition to identify the mmPing20 hyperactive clone.

2.4 Soybean transformation

Stable transformation of cv ‘Thorne’ was performed using the cotagode (Zhang et al., 1999) or half-seed method (Curtin et al., 2011; Paz et al., 2006). Transformation was verified by PCR and Southern blot analysis.

2.5 DNA extraction

DNA was extracted from young soybean leaves using an extraction buffer composed of 100 mM Tris pH 8, 20 mM EDTA, 1.42 M NaCl, 2% CTAB, 2% PVP40, and 4 mM DIECA. The leaf tissue was ground in a bead mill with 400 µl of extraction buffer and incubated at 55°C for 30 min. The solution was combined with 400 µl chloroform, mixed by inversion, centrifuged, and the aqueous phase was transferred to a new tube and precipitated by adding 300 µl isopropyl alcohol followed by centrifugation. The DNA pellet was washed with 70% ethanol and resuspended in TE buffer.

2.6 Library preparation

To generate tagged libraries, 1 µg of genomic DNA was fragmented in a 20-µl reaction consisting of 2 µl of 10X fragmentase buffer (New England Biolabs) and 2 µl fragmentase (New England Biolabs) and brought up to 20 µl in nuclease free water and incubated at 37°C for 20 min before stopping with 5 µl of 0.5M EDTA. Fragmented DNA was purified using a Zymo clean and concentrator column (Zymo Research) following the manufacturer’s instructions and eluted into 11 µl of TLE (10 mM Tris, 0.1 mM EDTA pH 8). A 3 µl aliquot of the purified DNA fragments was blunted in a 20 µl reaction containing 2 µl buffer 2.1 (New England Biolabs), 1 µl 2 mM dNTPs, 0.2 µl T4 polymerase (New England Biolabs), and 13.8 µl nuclease-free water and incubated at 12°C for 15 min. Blunted fragments were purified using a Zymo clean and concentrator column (Zymo Research) following the manufacturer’s instructions (eluted into 8 µl of TLE) before A-tailing in a reaction containing 2 µl of GoTaq buffer (Promega), 0.2 µl 10 mM ATP, 0.8 µl GoTaq polymerase (Promega), and 7 µl of the cleaned blunted fragments (70°C for 20 min). The A-tailed fragments were size selected using Mag- Bind® RxnPure Plus magnetic beads (Omega Bio-Tek) following the manufacturer’s instructions, except using 80% ethanol for the bead wash steps. Y-yoke adapters were prepared according to (Glenn et al., 2019) and were ligated to the a-tailed fragments “on-bead” by adding 2.5 µl of 5 µM y-yoke adapter, 2.5 µl of 10X ligation buffer (Promega), 17.5 µl nuclease-free water, and 2.5 µl of T4-ligase (Promega) [Table S3] (Glenn et al., 2019). The ligation reaction was incubated at room temperature for 3 hr and purified by adding 25 µl of a 20% PEG 2.5 M NaCl solution, mixing, and incubation for 10 min at room temperature prior to magnetic separation as previously described. The ligation was eluted from the beads into 25µl TLE. Adapter ligated fragments were enriched for those containing mPing using a PCR primer which was reverse complementary to the mPing 5’ end with a tail which facilitates subsequent multiplexing and binding to the Illumina flow cell. PCR was performed in 25 µl reactions containing 12.5 µl KAPA HiFi Hotstart ReadyMix [2X] (Kapa biosystems), 1.25 µl 5 µM mPing fusion primer (Table S3), 1.25 µl 5 µM iTru7 barcoded primer (Glenn et al., 2019), and 10 µl of the cleaned ligation. The PCR was performed in a thermocycler at 98°C for 1 min followed by 24 cycles of 98°C for 20 s, 60°C for 15 s, 72°C for 1 min, and a final extension of 5 min. The second index was added to the amplified libraries through a second PCR containing 12.5 µl KAPA HiFi Hotstart ReadyMix (Kapa biosystems), 1.25 µl 5 µM p7 primer (Table S3), 1.25 µl 5 µM iTru5 barcoded primer (Glenn et al., 2019), and 10 µl of the mPing-enrichment amplicon. The thermocycler conditions were the same as the mPing-enrichment PCR, except
with 20 amplification cycles. The final amplified library was purified using Mag-Bind® RxnPure Plus magnetic beads (Omega Bio-Tek) following the manufacturer’s instructions, except using a bead-solution:sample ratio of 0.7:1 v:v. The cleaned libraries were quantified using the KAPA qPCR library quantification kit (Kapa biosystems), pooled in equimolar amounts, and sequenced on an Illumina Miseq with paired-end 300-bp reads.

2.7 | Bioinformatic analysis

Reads were quality trimmed using Trimmomatic (Bolger et al., 2014) with a sliding window of 10 bp with an average quality score of 20 on the phred 33 scale. The ends of the reads were also trimmed with a quality threshold of 10. Next, the reads were filtered based on whether they contained the mPing sequence and the target site duplication sequence TTA or TAA. The mPing sequence was clipped from the filtered reads using the "headcrop" function of Trimmomatic (Bolger et al., 2014), leaving the soybean sequence that flanked the insertion site. The clipped reads were assembled using CAP3 (Huang & Madan, 1999) set to a minimum read overlap of 16 bp with a 70% identity threshold to generate contigs. The contigs were used as queries against the soybean genome version Wm82.a2.v1 (https://phytozome.jgi.doe.gov) using BLASTn (Camacho et al., 2009) to identify the insertion position. The coordinates of the insertions were then cross-referenced to the coordinates of annotated gene models from soybean (https://phytozome.jgi.doe.gov) using the BEDTools function "closest" (Quinlan & Hall, 2010).

2.8 | RNA-seq

RNA was extracted from 5-cm meristem tips, including immature trifoliolates, from month-old greenhouse-grown plants. Tissue was ground in liquid nitrogen before using Trizol Reagent to purify the RNA according to the manufacturer’s recommendations. Frozen RNA samples from two plants harboring the mmPing20F insertion, an untransformed 'Thorne' plant, and a transformed 'Thorne' plant that only contained the pWMD23 plasmid (ORF1 and TPase expression) were shipped to Novogene for standard Illumina RNA-seq analysis. The latter two samples served as the control group. Differential expression analysis was conducted using the R package DEseq2 with default parameters (Love et al., 2014). P-values for differential expression of genes in a 200kb window surrounding the mmPing20F insertion were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). Fold changes were calculated as normalized expression values in plants harboring the insertion divided by the normalized value of the control plants. Expression differences of genes in this 200kb window surrounding the insertion were considered differentially expressed if they had a log2 (fold change) higher than 1 and their adjusted p-value was lower than 0.05.

3 | RESULTS

3.1 | Development of mPing-based activation tags

Initially, two strategies were pursued to engineer mPing-based activation tags. The first approach added the ends of the mPing element, including the required terminal inverted repeat sequences (TIRs), to the Agrobacterium tumefaciens Octopine synthase (OCS)-enhancer sequence. This element called NL60 is 100 bp smaller than the original 430-bp mPing element (Figure 1a). The second strategy was to insert multiple copies of the Cassava vein mosaic virus (CMV) enhancer in the center of a complete mPing element. This latter element (2xE) is 932 bp, while the version with four copies of the enhancer (4xE) is 1,425 bp (Figure 1a). To analyze the transposition frequency of these elements, a previously established yeast-based assay that has been shown to correlate with mPing transposition in plants was employed (Hancock et al., 2010; Payero et al., 2016). The transposition assays demonstrated that all three of these elements transposed significantly less frequently than the native mPing element (Figure 1b). The NL60 element showed very low transposition, indicating that the TIRs alone are not sufficient for robust transposition. However, the finding that the 2xE and 4xE activation tags also transposed at significantly lower rates suggests that other important sequences could be disrupted and/or that increasing element size reduces transposition frequency.

3.2 | A modified mPing element resulted in increased transposition frequency

Transposition frequency is an important parameter in mutagenesis efficiency; therefore, mitigation of the low transposition frequency of our activation tags was addressed by minimizing their size and identifying hyperactive versions of mPing. Previous research showed that some native transposable elements are not optimized for transposition and genetic variants can be identified with hyperactive mobility (Yang et al., 2009). A library of 112 PCR-mutated versions of mPing, designated mmPing, was screened using the yeast transposition assay. While 12 of the mmPing elements showed lower transposition, 6 were identified with higher transposition frequency. One of these, the mmPing20 element, transposed at a significantly higher frequency than mPing (Figure 1b). This hyperactive element was found to have seven mutations relative to the original mPing element (T162C, T258A, T287C, T300A, T304A, T310A, and G372A). The identification of the hyperactive mmPing20 element suggests that mPing contains sequences that inhibit transposition complex formation.

The mmPing20 element was used as the basis for two novel activation tags containing enhancer sequences from the promoters of the figwort mosaic virus [FMV, 207 bp] (Maiti, Ghosh, et al., 1997; Maiti, Gowda, et al., 1997) and the soybean β-conglycinin gene [182 bp] (Chen et al., 1988). The transposition of these activation tags, called mmPing20F and mmPing20B, respectively (Figure 1a),
The results of the assays indicate that the `mmPing20F` and `mmPing20B` activation tags transpose at a lower rate than `mPing` and `mmPing20`, but at higher frequencies than the `2xE` and `4xE` elements. The finding that `mmPing20F` transposes significantly better than `mmPing20B` even though they are similar in size (640 bp and 622 bp) suggests that the enhancer sequences can affect the efficiency of functional transposition complex formation.
3.3 | mPing-based activation tags can be mobilized in soybean

In order to test the mobilization of these activation tags in plants, a set of seven independent transgenic soybean lines with an mmPing20F activation tag construct (pEarleyGate 103 mmPing20F, Figure 2a) was developed. Transformants containing the pEarleyGate 103 mmPing20F construct were crossed with five transgenic lines that carry expression cassettes of ORF1 and TPase that reside in the vector pWMD23 (Figure 2b). Transposition of mmPing20F is only possible in the presence of ORF1 and TPase proteins, so a total of 20 transgenic stacks with both constructs were generated (Figure 3a). These were tested for transposition using PCR with primers flanking the mmPing20F element within the vector backbone (Figure 2a). Eighteen (90%) of the transgene stack combinations produced both a 960-bp and a 317-bp PCR product, indicating that somatic mmPing20F excision was occurring in the F1 generation (Figure 3a). Subsequent genotyping via PCR in the F2 generation from 10 lineages that displayed transposition in the F1 generation with mmPing20F flanking primers was performed (Figure 3b). This led us to select the 16-28-3 line because it produced some plants that only had the 317-bp PCR product, consistent with germinal excision of mmPing20F (Figure 3b). Progeny from line 16-28-3 were sown to identify homozygous lineages for pWMD23 and segregating for a single locus containing the pEarleyGate 103 mmPing20F transgenes (i.e., 16-28-3-14-4). A second line was identified that had lost the pEarleyGate 103 mmPing20F allele but was homozygous for the mmPing20F element (16-28-3-14-2). Together, this inheritance pattern indicates that a germinal transposition event occurred in the 16-28-3-14-2 line. These results along with evidence for element mobility in transgenic soybean containing the 2xE activation tag in two other soybean lines (Figures S1–S3) indicate that mPing-based activation tags are mobile in the soybean genome.

**FIGURE 3** Development and analysis of the mmPing20F soybean population. Diagram showing the crosses that were made between the transgenic lines carrying the mmPing20F construct and the pWMD23 (TPase) construct (a). Colors indicate the amount of transposition detected in the F1 generation from each respective cross. PCR analysis of the 13 plants from the F2 generation of the 16-28 line (b). In the PCR with mmPing20F flanking primers, the upper band (960 bp) indicates that the element is still located in the transgene, while the lower band (317 bp) indicates mmPing20F has excised. Pong TPase primers indicate if the TPase expression construct pWMD23 is present.
## 3.4 | Insertion site analysis

A high-throughput sequencing approach was employed to identify *mmPing20F* insertion sites over three generations of the transgenic events. An average of 280,267 clean reads with a standard deviation of 129,024 were generated for 16 plants derived from 16-28-3 (Table S1). Variation in read depth did not likely contribute to variation in *mmPing20F* reads retrieved (Figure S4), suggesting that this read depth was sufficient for the unbiased discovery of *mPing* insertions.

We identified 11 *mmPing20F* insertions in this population, and each of the mappable contigs were found to be located within 4 kb of annotated genes (Figure 4). Contigs that did not align to the soybean reference genome were composed of reads that matched the transformation vector, representing *mmPing20F* elements that did not mobilize. Seven of the eight plants harboring *mmPing20F* insertions had the same insertion at position 2785626 on chromosome 8 (Figure 5). This insertion was 3,792 bp upstream of the nearest gene, Glyma.08G035000, and 5,558 bp downstream from the next closest gene, Glyma.08G035100. This insertion was present in the progeny of 16-28-3-14 and 16-28-3-15, suggesting that they were inherited through their most recent common ancestor, 16-28-3 (Figure 5). However, the insertion at Chr08:2785626 was not detected in the parental lines 16-28-3-14 and 16-28-3-15 or the grandparent 16-28-3 using contig or single-read alignments (Figure S5). Our sequencing strategy could have missed this Chr08:2785626 insertion in these generations due to the limitations of sampling a single leaf, but it is likely that there was only a single insertion event. None of the other insertions were found in multiple progeny suggesting that they were either somatic insertions specific to the tissue used to generate the DNA preparation or novel germline insertions specific to that lineage (Figure 5).

### Table: Insertion site analysis

| Insertion site | Distance to closest gene | Nearest annotated gene | Exon | Intron | 3'UTR |
|----------------|--------------------------|------------------------|------|--------|-------|
| Chr02:46956366 | 3058                     | Glyma.02G289900        | -    | -      | -     |
| Chr05:1781052  | 0                        | Glyma.05G019900        | +    | -      | -     |
| Chr05:34959800 | -3646                    | Glyma.05G157000        | -    | -      | -     |
| Chr06:5418454 | 0                        | Glyma.06G070900        | +    | -      | -     |
| Chr06:9476608 | 0                        | Glyma.06G116400        | +    | -      | +     |
| Chr08:2785626 | -3792                    | Glyma.08G035000        | -    | -      | -     |
| Chr08:35502159 | 0                      | Glyma.08G271700        | +    | -      | +     |
| Chr09:6491470 | 0                        | Glyma.09G066300        | -    | +      | -     |
| Chr13:24754819 | 3292                    | Glyma.13G135200        | -    | -      | -     |
| Chr14:5513222 | -109                     | Glyma.14G066800        | -    | -      | -     |
| Chr20:25674286 | 173                     | Glyma.20G072600        | -    | -      | -     |

**FIGURE 4** Genomic features associated with *mmPing20F* insertions identified by sequencing. The insertion site corresponds to the base immediately flanking the *mmPing20F* tag. Distances of zero indicate the insertion is within the annotated gene model. Negative distances denote upstream insertions. Shaded cells with the "+" sign indicate the insertion is in the gene feature.
DISCUSSION

Activation tagging mutagenesis is designed to generate miss-expression mutations. The data presented demonstrates that the mPing-based activation tags developed are capable of transposition (Figures 1 and 3) and altering native soybean gene expression levels (Figure 6). While we did not evaluate expression in multiple tissues across time, we anticipate that because of the nature of the FMV enhancer, induction of overexpression by mmPing20F will be constitutive. This indicates that mPing is a suitable vector for the delivery of enhancer sequences.

A requirement for genome-wide saturation with transposon tags is that the element must maintain mobility. Compared to other transposable elements, relatively little is known about the transposition of the mPing element beyond the fact that the ORF1 and TPase proteins are required (Hancock et al., 2010; Yang et al., 2007). This study provides the first report of how modification of mPing changes its transposition capacity. The NL60 element, with only the 60 bp ends of mPing, had very low transposition rates (Figure 1), indicating that although the TIR sequences are sufficient for mobility, additional internal regions promote transposition activity. This finding is consistent with the results observed for a Stowaway-like MITE, where...
sub-TIR and internal regions were shown to promote transposition (Yang et al., 2009). One speculation is that these transposition promoting sequences may be involved in recruitment of the transposase proteins and subsequent formation of the transposition complex, similar to the role of Transposon Gene A binding sites in the transposition of Suppressor-mutator elements (Raina et al., 1998).

The result showing that the 2xE and 4xE elements, consisting of a large enhancer sequence inserted into mPing, also had decreased transposition suggests that increases in the size of the mPing element results in decreased transposition rates (Figure 1). This is consistent with the 430-bp mPing element showing considerably higher activity than the autonomous 5341-bp Ping element in rice (Kikuchi et al., 2003; Naito et al., 2004). This effect of size has also been observed for other Type II transposable elements (Tosi & Beverley, 2000; Way & Kleckner, 1985) and is likely one of the reasons that MITEs achieve higher copy numbers than their corresponding larger autonomous elements. Thus, the overall size of the element must be considered as additional mPing-based tags are constructed.

We identified both germinal and somatic insertions by introducing mPing-based activation tags into soybean. The majority of insertions were specific to individual plants, suggesting that they are somatic insertions that only affect a limited number of cells. However, we did identify a heritable mutation in our relatively small population that must have occurred in germinal tissue. Analysis of additional generations will be needed to calculate the frequency of novel germine insertions. This initial population suggests that germine insertion rate may be relatively low, consistent with a previous mPing population that showed a germine insertion frequency of about one per generation (Hancock et al., 2011). Thus, future efforts will need to focus on mechanisms to increase the transposition activity. Our goal is to increase transposition rates in soybean to those comparable to transposon tagging in maize, where populations mutagenized with an Enhancer/Suppressor transposon-based activation tag showed germinal excision rates ranging from 20% to 60% (Davies et al., 2019).

Interestingly, we did not detect the insertion at position 2785626 on chromosome 8 in plant 16-28-3 or its progeny, but this insertion was detected in two separate families derived from this event (progeny from 16-28-3-14 and from 16-28-3-15 contained the insertion). The fact that this insertion was not detected in 16-28-3 is not surprising, as the DNA collected for sequencing was sampled early in development, and the transposition may have occurred later and in a different tissue. It is surprising that we did not detect the insertion in the 16-28-3-14 or 16-28-3-15 plants, given that their progeny were positive for the insertion. One explanation for this is that the tissue that gave rise to the leaves sampled for sequencing in 16-28-3-14 and 16-28-3-15 had an early somatic transposition reverting the chromosome 8 locus back to wild type, but the germline insertion remained giving rise to the seeds that contained the insertion in the next generation. We did not take multiple tissue samples from these plants to test this possibility. This pattern is supported in previous transposon display analysis of native mPing elements where germline bands present in the T0 generation were either absent or weaker in the T1, but the subsequent T2 plants had strong bands (Hancock et al., 2011). It is also worth noting that these plants tested positive for ORF1 and TPose that are required for mobility.

Based on the behavior of the original mPing element, we anticipated approximately 50% of the insertions would fall within 2.5 kb of a gene (Hancock et al., 2011). We observed that 7 of the 11 mapped insertions (64%) were within 2.5 kb of an annotated gene and the remaining 4 were less than 3.8 kb away (Figure 4). This is relevant, given that our results with mmping2OF (FMV enhancer) showed that a gene as far away as 5.5 kb showed a significant increase in expression (Figure 6). We suspect that the degree of gene induction is dependent on the enhancer sequence used and its chromosomal context. Also, a larger population of mmping2OF mutagenized plants will need to be analyzed before we can determine the proximity limits for affecting gene expression. Nonetheless, these results are consistent with reports from T-DNA activation tagging with a cauliflower mosaic virus 35S enhancer sequence that induced overexpression at distances of up to 3.6 kb in Arabidopsis (Weigel et al., 2000) and 13.1 kb in tobacco (Liu et al., 2015).

The mmping2OF activation-tag induced upregulation of Glyma.08G035000 which is related to the Arabidopsis TINY2. TINY2 belongs to the APETALA2/ethylene response factor transcription factor family. There are three TINY homologs in Arabidopsis, and triple knockouts grow larger than wild type, while overexpression lines display stunted growth (Xie et al., 2019). There are two additional soybean genes with high sequence similarity to Glyma.08G035000, and phylogenetic analysis revealed four other related soybean genes cluster closer with AtTINY1 and AtTINY2 demonstrating that Glyma.08G035000 is only part of a larger TINY-like gene family (Figure S7). Our soybean plants overexpressing Glyma.08G035000 did not display any reduction in stature or other obvious visual phenotypes under normal growing conditions. This may be attributed to subfunctionalization of the Glyma.08G035000 gene as well as the difference in magnitude of upregulation by mmping2OF [6.2-fold compared to the ~200-fold increase in TINY-overexpression Arabidopsis lines] (Xie et al., 2019). AtTINY recognizes the DRE promoter element (A/GCCGAC) and negatively regulates brassinosteroid-mediated growth through suppression of genes that respond to brassinosteroid application (Xie et al., 2019). The number of downregulated genes in TINY-like overexpression lines relative to wild type was smaller (27 genes) in soybean lines compared to that of Arabidopsis (2,247 genes) (Xie et al., 2019). Additionally, we did not detect significant enrichment for DRE elements in the promoter regions of downregulated genes in our study, suggesting that Glyma.08G035000 has different target genes than AtTINY2, or the magnitude of Glyma.08G035000 upregulation was not sufficient to suppress its target genes.

Results from our relatively small population show promise that mPing-based tags offer less up front labor to generate mutant populations than Ac/Ds-based tags. In most species, Ac/Ds transposition largely occurs in localized chromosomal regions surrounding the transgene integration site with fewer transpositions occurring on different chromosomes (Bancroft & Dean, 1993; Nakagawa et al., 2000; Vollbrecht et al., 2010). To achieve insertions on every chromosome of a species of interest, Ac/Ds based
tags are often screened for unlinked insertions using selectable or phenotypic markers (Qu et al., 2008; Suzuki et al., 2001). This requires researchers to handle and analyze a large number of plants to identify unlinked insertions. For example, a rice Ac/Ds activation tagging project screened over 3,000 plants produced from 37 transformation events to find insertions that reached every rice chromosome (Qu et al., 2008). The exception is the case of soybean where Ac/Ds-based tags were not observed preferentially inserting in linked loci (Singh, 2012). However, mPing transposes to unlinked loci in both Arabidopsis and soybean (Hancock et al., 2011; Yang et al., 2007), so this system potentially offers less labor to saturate the genome with insertions across a wider range of species. Using a transformatant with a single-copy vector integration, Yang et al., 2007 found mPing insertions on all five Arabidopsis chromosomes with nine T2 individuals. Similarly in soybean, mPing insertions spanned 18 of the 20 soybean chromosomes by analyzing 15 plants (Hancock et al., 2011), and we identified mmPing20F insertions on nearly half of the soybean chromosomes in only 8 F3 lines.

Ac/Ds-based systems have an advantage over mPing in that it has been reported to produce more germline insertions. Novel germinal insertions were discovered at a rate of 56% (number of unique germinal insertions/total number of progeny screened) while mPing in soybean yielded 22% (Hancock et al., 2011; Qu et al., 2008). Since lengthening the mPing element reduces transposition frequency (Figure 1), it was not surprising that our germinal mmPing20F insertion discovery rate was lower than previous reports of mPing. As a means to ameliorate the relatively low transposition rates, screening for additional versions of the element with enhanced transposition rates is being pursued. Given that different enhancer sequences can variably affect transposition frequency (Figure 1), gaining insight on how these sequences contribute to transposition complex formation may also allow for the design of elements with enhanced mobility. It is also possible that some of the novel mutations that we introduced to form mmPing20F have inhibitory effects on transposition activity. Future evaluation of each of the seven mutations would be useful for identifying the mechanisms that control transposition.

Moreover, by incorporating a visual reporter platform that can track germinal transposition events would allow for large numbers of heritable mutations to be identified from a massive field population. Finally, efforts to develop a system to restrict ORF1 and TPase expression in the gametes are underway. This would stabilize mPing-based tags in the somatic tissues, allowing for simpler analysis of insertions, and strengthen its usefulness as a mutational tagging system for crop plants.

5 | CONCLUSION

Identification of the hyperactive mmPing20F element allowed for development of activation tags suitable for plant mutagenesis. The mmPing20F element was shown to be heritably mobilized in soybean, resulting in overexpression of the adjacent soybean genes. Transfer of mPing-based activation tagging technology to additional plant species should be relatively straightforward as our plasmids should be acceptable for most dicot species and development of monocot-compatible constructs is underway. This represents a technological breakthrough in that this mPing-based system will allow for production of mutagenized populations that can facilitate gene discovery in highly duplicated or polyploid genomes.

ACCESSION NUMBERS
Vectors used in this article can be found under accession numbers 140006-140007 and 145787-145795 on AddGene (https://www.addgene.org/).

LARGE DATASETS
Sequencing data from transposon mapping and RNA-seq from this article can be found in the GenBank data libraries under BioProject PRJNA605018.

ACKNOWLEDGMENTS
We would like to thank Gary Stacey and Minviluz Stacey (University of Missouri – Columbia) for their critical feedback on this project. We would also like to thank Theresa McManimon for performing the BertMN01 transformation. Funding was provided by NSF Plant Genome Research Program Grants # 0820769, 1127083, and 1444581.

AUTHOR CONTRIBUTIONS
C.N.H., W.P., T.E.C., and R.M.S. designed the experiments. D.R.P. developed and screened the mutagenized mPing library. A.S, C.N.H., P.S.R., and H.N. developed the plasmid constructs. S.D., J.R., and P.S.R. performed yeast transposition assays. S.D. and J.R. performed PCR analysis of the soybean plants. H.N. and A.S generated transgenic plants and performed PCR and southern blots. S.D. purified the RNA. L.M. constructed sequencing libraries and validated their quality. E.M. sequenced and mapped the mmPing20F insertions. A.J. analyzed the mmPing20F insertion sites and RNA seq data. C.N.H and A.J. wrote the manuscript. W.P., T.E.C., and E.M. critically edited the manuscript.

ORCID
Alexander Johnson https://orcid.org/0000-0002-4090-1765
Edward Mcassey https://orcid.org/0000-0003-1568-3361
Stephanie Diaz https://orcid.org/0000-0002-0871-0563
Jacob Reagin https://orcid.org/0000-0003-1392-212X
Robert M. Stupar https://orcid.org/0000-0002-8836-2924
Wayne A. Parrott https://orcid.org/0000-0001-7847-1134
C. Nathan Hancock https://orcid.org/0000-0002-0875-8495

REFERENCES
Aarts, M. G., Corzaan, P., Stiekema, W. J., & Pereira, A. (1995). A two-element Enhancer-Inhibitor transposon system in Arabidopsis
Meissner, R., Chague, V., Zhu, Q., Emmanuel, E., Elkind, Y., & Levy, A. A. (2000). A high throughput system for transposon tagging and promoter trapping in tomato. The Plant Journal, 22, 265–274. https://doi.org/10.1046/j.1365-313x.2000.00735.x

Naito, K., Cho, E., Yang, G. J., Campbell, M. A., Yano, K., Okumoto, Y., Tanisaka, T., & Wessler, S. R. (2006). Dramatic amplification of a rice transposable element during recent domestication. Proceedings of the National Academy of Sciences of the United States of America, 103, 17620–17625. https://doi.org/10.1073/pnas.0605421103

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C. N., Richardson, A. O., Okumoto, Y., Tanisaka, T., & Wessler, S. R. (2009). Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature, 461, 1130–1134. https://doi.org/10.1038/nature08479

Nakagawa, Y., Machida, C., Machida, Y., & Toriyama, K. (2000). Frequency and pattern of transposition of the maize transposable element Ds in transgenic rice plants. Plant and Cell Physiology, 41, 733–742. https://doi.org/10.1093/pcp/41.6.733

Nakazaki, T., Okumoto, Y., Horibata, A., Yamahira, S., Teraishi, M., Nishida, H., Inoue, H., & Tanisaka, T. (2003). Mobilization of a transposon in the rice genome. Nature, 421, 170–172. https://doi.org/10.1038/nature01219

Parinov, S., Sevugan, M., Ye, D., Yang, W. C., Kumaran, M., & Sundaresan, V. (1999). Analysis of flanking sequences from Dissociation insertion lines: A database for reverse genetics in Arabidopsis. The Plant Cell, 11, 2263–2270.

Payero, L., Outten, G., Burckhalter, C., & Hancock, C. N. (2016). Alteration of the Ping and Pong ORF1 proteins allows for hyperactive transposition of mPing. Journal of the South Carolina Academy of Science, 14, 3.

Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M., & Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient Agrobacterium-mediated soybean transformation. Plant Cell Reports, 25, 206–213. https://doi.org/10.1007/s00299-005-0048-7

Qu, S., Desai, A., Wing, R., & Sundaresan, V. (2008). A versatile transposon-based activation tag vector system for functional genomics in cereals and other monocot plants. Plant Physiology, 146, 189–199. https://doi.org/10.1104/pp.107.111427

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841–842. https://doi.org/10.1093/bioinformatics/btp033

Rains, R., Schlappi, M., Karunanandaa, B., Elhofy, A., & Fedoroff, N. (1998). Concerted formation of macromolecular Suppressor-mutator transposition complexes. Proceedings of the National Academy of Sciences of the United States of America, 95, 8526–8531. https://doi.org/10.1073/pnas.95.15.8526

Schaffer, R., Ramsay, N., Samach, A., Corden, S., Putterill, J., Carré, I. A., & Coupland, G. (1998). The late elongated hypocotyl mutation of Arabidopsis disrupts circadian rhythms and the photoperiodic control of flowering. Cell, 93, 1219–1229. https://doi.org/10.1016/S0092-8674(00)81465-8

Schneider, A., Kirch, T., Gigolashvili, T., Mock, H.-P., Sonnewald, U., Simon, R., Flügge, U.-I., & Werr, W. (2005). A transposon-based activation-tagging population in Arabidopsis thaliana (TAMARA) and its application in the identification of dominant developmental and metabolic mutations. FEBS Letters, 579, 4622–4628.

Singh, M. (2012). Monitoring Ds transposition in the soybean genome. University of Nebraska-Lincoln, DigitalCommons.

Skarnes, W. C. (1990). Entrapment Vectors - A new tool for mammalian genetics. Bio-Technology, 8, 827–831. https://doi.org/10.1038/nbt0990-827

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J. D. G., Dean, C., Ma, H., & Martienssen, R. (1995). Patterns of gene-action in plant development revealed by enhancer trap and gene trap transposable elements. Genes & Development, 9, 1797–1810. https://doi.org/10.1101/gad.9.14.1797

Suzuki, Y., Uemura, S., Saito, Y., Murofushi, O., Schmitz, G., Theres, K., & Yamaguchi, I. (2001). A novel transposon tagging element for obtaining gain-of-function mutants based on a self-stabilizing Ac derivative. Plant Molecular Biology, 45, 123–131.

Teraishi, M., Okumoto, Y., Hirochika, H., Horibata, A., Yamagata, H., & Tanisaka, T. (1999). Identification of a mutable slender glume gene in rice (Oryza sativa L.). Molecular & General Genetics, 261, 487–494. https://doi.org/10.1007/s004380050992

Thomson, M. J., Ismail, A. M., McCouch, S. R., & Mackill, D. J. (2010). Marker assisted breeding. In A. Pareek, S. K. Sopory, & H. J. Bohnert (Eds.), Abiotic stress adaptation in plants (Vol. 4, pp. 451–469). Springer.

Tosi, L. R. O., & Beverley, S. M. (2000). Cis and trans factors affecting Mos1 mariner evolution and transposition in vitro, and its potential for functional genomics. Nucleic Acids Research, 28, 784–790. https://doi.org/10.1093/nar/28.3.784

Vollbrecht, E., Duvick, J., Schares, J. P., Ahern, K. R., Deewatthanawong, P., Xu, L., Conrad, L. J., Kikuchi, K., Kubinec, T. A., & Hall, B. D. (2020). Genome-wide distribution of transposed Dissoociation elements in maize. The Plant Cell, 22, 1667–1685.

Way, J. C., & Kleckner, N. (1985). Transposition of plasmid-borne Tn10 elements does not exhibit simple length-dependence. Genetics, 111, 705–713.

Weigel, D., Ahn, H. J., Blazquez, M. A., Borevitz, J. O., Christensen, S. K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malanchurivil, E. J., Neff, M. M., Nguyen, J. T., Sato, S., Wang, Z. Y., Xia, Y. J., Dixon, R. A., Harrison, M. J., Lamb, C. J., Yanofsky, M. F., & Chory, J. (2000). Activation tagging in Arabidopsis. Plant Physiology, 122, 1003–1013. https://doi.org/10.1100/pp.122.4.1003

Wilson, C., Bellen, H. J., & Gehring, W. J. (1990). Position effects on eukaryotic gene-expression. Annual Review of Cell Biology, 6, 679–714. https://doi.org/10.1146/annurev.cb.06.110190.003335

Wilson, K., Long, D., Swinburne, J., & Coupland, G. (1996). A dissociation insertion causes a semidominant mutation that increases expression of TINY, an Arabidopsis gene related to APETALA2. The Plant Cell, 8, 659–671.

Xie, Z., Nolan, T., Jiang, H., Tang, B., Zhang, M., Li, Z., & Yin, Y. (2019). The AP2/ERF transcription factor TINY modulates brassinosteroid-regulated plant growth and drought responses in Arabidopsis. The Plant Cell, 31, 1788–1806. https://doi.org/10.1105/tpc.18.00918

Yang, G. J., Nagel, D. H., Feschtotte, C., Hancock, C. N., & Wessler, S. R. (2009). Tuned for transposition: Molecular determinants underlying the hyperactivity of a stowaway MITE. Science, 325, 1391–1394. https://doi.org/10.1126/science.1175688

Yang, G. J., Zhang, F., Hancock, C. N., & Wessler, S. R. (2007). Transposition of the rice miniature inverted repeat transposable element mPing in Arabisopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America, 104, 10962–10967. https://doi.org/10.1073/pnas.0702080104

Zhang, Z., Xing, A., Stawiski, P., & Clemente, T. (1999). The use of glucosinate as a selective agent in Agrobacterium-mediated transformation of soybean. Plant Cell, Tissue and Organ Culture, 56, 37–46.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Johnson A, Mcassey E, Diaz S, et al. Development of mPing-based activation tags for crop insertional mutagenesis. Plant Direct. 2021;00:1–12. https://doi.org/10.1007/pld3.300