Alternative Transposition Generates New Chimeric Genes and Segmental Duplications at the Maize *p1* Locus

Dafang Wang,* Chuanhe Yu,* Tao Zuo,* Jianbo Zhang,* David F. Weber,[†] and Thomas Peterson^{*,†,1}

*Department of Genetics, Development and Cell Biology and [†]Department of Agronomy, Iowa State University, Ames, Iowa 50011-3260, and [†]School of Biological Sciences, Illinois State University, Normal, Illinois 61790-4120

ABSTRACT The maize *Ac/Ds* transposon family was the first transposable element system identified and characterized by Barbara McClintock. *Ac/Ds* transposons belong to the *hAT* family of class II DNA transposons. We and others have shown that *Ac/Ds* elements can undergo a process of alternative transposition in which the *Ac/Ds* transposase acts on the termini of two separate, nearby transposons. Because these termini are present in different elements, alternative transposition can generate a variety of genome alterations such as inversions, duplications, deletions, and translocations. Moreover, *Ac/Ds* elements transpose preferentially into genic regions, suggesting that structural changes arising from alternative transposition may potentially generate chimeric genes at the rearrangement breakpoints. Here we identified and characterized 11 independent cases of gene fusion induced by *Ac* alternative transposition. In each case, a functional chimeric gene was created by fusion of two linked, paralogous genes; moreover, each event was associated with duplication of the ~70-kb segment located between the two paralogs. An extant gene in the maize B73 genome that contains an internal duplication apparently generated by an alternative transposition event was also identified. Our study demonstrates that alternative transposition-induced duplications may be a source for spontaneous creation of diverse genome structures and novel genes in maize.

KEYWORDS *Ac/Ds* alternative transposition; translocation; segmental duplication; chimeric gene; exon shuffling

THE complex architecture of the modern maize genome was shaped by ancient whole-genome duplication and subsequent diploidization events (Messing *et al.* 2004; Haberer *et al.* 2005). Transposable elements proliferated during the diploidization process (Gaut *et al.* 2000) and today comprise ~85% of the maize genome (Schnable *et al.* 2009). For example, multiple insertions of high-copy-number families of class I retrotransposons greatly expanded the intergenic regions of the maize genome. Comparisons of maize and its close relative sorghum indicate that gene number and colinearity are largely conserved (Hulbert *et al.* 1990), whereas the intergenic regions differ greatly due to frequent retrotransposon insertions in maize. A good example is the *adh1* region

of maize inbred B73. Nineteen nested LTR retroelements and two solo LTRs together make up 74% of the intergenic sequences between *adh1* and the nearest adjacent gene (Tikhonov *et al.* 1999). Transposition and proliferation of retroelements thus can largely explain genome size variations, the C-value paradox, and discordant evolutionary history among different grass species (SanMiguel and Vitte 2009). Class II DNA transposons generally do not amplify to such high copy numbers as the class I retroelements, but they can induce structural changes in various ways. For example, the maize *Ac/Ds* transposon family, class II transposable elements, can undergo a process of alternative transposition that can generate large-scale structural changes such as duplications, translocation, inversions, and deletions (Zhang *et al.* 2006; Yu *et al.* 2011). Moreover, a recent study reported evidence that alternative transposition of transposon *dhAT-Zm1* generated three tandem duplications during the evolution of the maize B73 genome (Zhang *et al.* 2013).

In addition to generating structural variations, class II elements and some families of class I retroelements often

Copyright © 2015 by the Genetics Society of America

doi: 10.1534/genetics.115.178210

Manuscript received May 15, 2015; accepted for publication September 6, 2015; published Early Online September 30, 2015.

Supporting information is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178210/-/DC1.

¹Corresponding author: 2258 Molecular Biology, Iowa State University, Ames, IA 50011-3260. E-mail: thomasp@iastate.edu

insert near gene regulatory and coding sequences where they may potentially induce ectopic gene expression or generate chimeric genes in grass genomes (Bennetzen 2005). In rice, Wang *et al.* (2006) found that >30% of retrogenes had recruited flanking exons at insertion sites and formed chimeric genes. The class II transposons *Pack-MULE* in rice (Jiang *et al.* 2004) and *helitrons* in maize (Morgante *et al.* 2005) were also reported to generate chimeric transcripts by producing dispersed duplications that engage genic sequences from different genes. In maize, Zhang *et al.* (2006) reported that functional chimeric genes could be formed by deletions induced by Reversed-Ends Transposition (RET) of the class II *Ac/Ds* transposon system.

The autonomous *Ac* element is 4565 bp in length, encodes a single transposase (TPase) gene, and contains TPase-binding sites located in Terminal Inverted Repeats (TIR) and subterminal regions (Kunze *et al.* 1987; Kunze and Starlinger 1989; Starlinger *et al.* 1989; Kunze and Weil 2002). *Fractured Ac* (*fAc*) and nonautonomous *Ds* elements retain one or both functional transposon termini, respectively, but lack TPase-coding capability; they can be nonautonomously transposed and participate in alternative transposition reactions (Fedoroff *et al.* 1983; Kunze and Starlinger 1989; Starlinger *et al.* 1989; Varagona and Wessler 1990; Zhang and Peterson 1999; Du *et al.* 2011).

The study described here follows a recent report that demonstrated that RET can generate *de novo* segmental duplications at the *p1* locus in maize (Zhang *et al.* 2013). We hypothesized that the same mechanism of *Ac* alternative transposition can shuffle exons and generate chimeric genes and thus can serve as a novel mechanism to generate new genes. Here we show that (1) RET of reversed *Ac/fAc* ends in the maize *p1* gene can induce an interchromosomal reciprocal translocation (*p1-vvD103*) with disrupted *p1* function; (2) among the progeny of *p1-vvD103*, subsequent RET events produced a series of 11 new alleles (named as *P1P2-1–P1P2-14*) with ~70-kb segmental duplications and restored *p* function; and (3) an extant gene in the maize B73 genome (AC234515.1) contains an internal duplication that is consistent with the structure produced by an alternative transposition event. The *P1P2* duplications shuffle exons from different genes to generate fusion transcripts with the phenotypes resembling the *p1* gene. Our study demonstrates that RET-induced segmental duplication is a mechanism to produce both genome structural variation and novel maize genes.

Materials and Methods

Maize stocks and screen

P1-ovov454 was heterozygous with 4Co63 (*p1-ww*). The *p1-ovov454* plants were crossed with *Ac* tester lines *p1-ww*; *r-m3::Ds*. Ears were screened for multikernel sectors with colorless pericarp and a 1:1 ratio of spotted to nonspotted kernels. The colorless spotted kernels were planted and

analyzed, yielding the allele *p1-vvD103*. The *p1-vvD103* plants were also crossed to the *p1-ww*; *r-m3::Ds* tester line. Red sectors with purple spots were phenotypically screened and selected as candidates of *P1P2* alleles.

The *Ac* tester line is an *r1* loss-of-function line by *Ds* insertion. The introduction of active *Ac* in the genome can somatically excise *Ds* and recover *r1* function, which produces purple sectors on the aleurone (Kermicle 1980). For *p1-vvD103/p1-ww r-m3::Ds* ears, we observed normal frequency of purple sectors, indicating that *Ac* in *p1-vvD103* allele is active.

Molecular biological methods

Leaf tissue from young plant was collected and ground in liquid nitrogen. Total DNA was prepared by using a modified cetyltrimethylammonium bromide extraction protocol (Porebski *et al.* 1997). HotMaster Taq polymerase from Eppendorf (Hamburg, Germany) was used in the PCR reaction. The PCR samples were heated at 94° for 2 min, followed by 35 cycles of 94° for 20 sec, 60° for 30 sec, and 65° for 2 min. Another cycle at 65° is extended for 8 min. PCR primers sequences are listed in [Supporting Information, Table S1](#).

Genomic Southern blot was performed according to the protocol by Sambrook *et al.* (1989); the washing stringency used is 0.5% SDS, 0.5× SSC at 60°.

Pericarp samples for RT-PCR analysis were collected from kernels at 15–20 days after pollination and frozen in liquid nitrogen. Total RNA was extracted by an RNeasy Plant mini kit (Qiagen, Valencia, CA) and treated with DNase I (Qiagen) to remove residual genomic DNA. The PCR products were purified by a Gel/PCR DNA fragment extraction kit (IBI Scientific, Peosta, IA). Sanger sequencing was performed by the DNA Facility at Iowa State University.

Data availability

Maize stocks are available upon request. Reference genome sequence data for maize inbred lines B73 (Schnable *et al.* 2009) and Mo17 (Xin *et al.* 2013) have been reported. Sequences of maize genomic MITEs are from Chen *et al.* 2014. [Table S1](#) contains sequences of oligonucleotide primers used in PCR experiments. [Table S2](#) contains cytogenetic data for presence of reciprocal translocations associated with the *P1P2* alleles. [File S1](#) contains sequences at the *p1-vv-D103* chromosome 1-10 translocation junction. [File S2](#) contains sequences of *P1P2* fusion junctions. [File S3](#) contains sequence of the *p2* transcript detected in allele *P1P2-3*. [File S4](#) shows the alignment of MITE SQ225117735 contained in the duplications in maize gene AC234515.1 with representative genomic MITE SE220500133. [File S5](#) shows an alignment of the 3' TIR of MITE SQ225117735 contained in the duplications in maize gene AC234515.1 with the 5' TIR of a complementary genomic MITE SQ225005122. [File S6](#) shows examples of maize B73 genomic MITEs containing 5' TIRs with perfect complementarity to the 3' TIR of the duplication-associated MITE SQ225117735. [File S7](#) shows an alignment of *p1* and *p2* sequences including the sites of *Ac* insertion derived from

Figure 1 RET of *Ac/fAc* produces reciprocal chromosome translocation and segmental duplications. Black and blue lines indicate maize chromosome 1 and 10 sequences, respectively. Red lines with arrowheads indicate *Ac* and *fAc* elements; open and solid red arrowheads indicate 3' and 5' *Ac* termini, respectively. Black and gray boxes indicate exons of maize *p1* and *p2* genes, respectively. This legend applies to all figures. Photographs of kernels of plants heterozygous for each allele are presented alongside their schematic structures. (A) Upper (black line): chromosome 1 containing *p1* and *p2* loci with *Ac/fAc* elements in progenitor allele *p1-ovov454*. Lower (blue line): chromosome 10 with transposition target site indicated by vertical orange line. (B) Following chromosome replication, *Ac* and *fAc* termini are excised from *p1* locus and inserted into target site on chromosome 10. (C) Translocation structure of the *p1-vvD103* allele. Asterisk marks the 5.9-kb duplication at the chromosome 1-10 translocation junction in *p1-vvD103*. (C') Following DNA replication, the reciprocal translocation chromosomes contain two identical sister chromatids. The ~70-kb region of the ensuing duplication is indicated by gray-shaded arrows. (D) The reverse-oriented *Ac* and *fAc* termini excise from the *p1* locus of the upper chromatid and insert into the target site (green vertical line) of *p2* gene on the sister chromatid. (E) Schematic structure of the *P1P2* allele with associated chromosomes. Upper line indicates translocation chromosome 1-10 containing ~70-kb direct duplications indicated as gray-shaded arrows. Lower line indicates reciprocal translocation chromosome 10-1.

standard transposition in *p1* (Athma *et al.* 1992) and alternative transposition in *p2* (this study).

Results

P1-vvD103, an interchromosomal translocation allele induced by reversed-*Ac*-ends transposition

In this study we use a pair of paralogous genes, *p1* and *p2*, as markers to phenotypically track the structural changes at the maize *p1* and *p2* loci. Both genes encode R2R3 Myb-like transcription factors that regulate the biosynthesis of red phlobaphene pigments (Lechelt *et al.* 1989; Grotewold *et al.* 1991, 1994). The *p1* and *p2* genes share >98% identity in the ORF sequences, but exhibit distinct expression patterns due to different promoter regions. *P1* is expressed in pericarp, cob, tassel glumes, and silk, and *p2* is expressed in anther and silk (Zhang *et al.* 2000). Alleles of *p1* are commonly identified by a suffix that indicates their expression in pericarp and cob: *P1-rr* specifies red pigmentation in both pericarp and cob, *p1-ww* has no pigmentation in either pericarp or cob, and

p1-vv exhibits variegated pigmentation in both pericarp and cob (Grotewold *et al.* 1991; Athma *et al.* 1992).

Our studies were initiated with the allele *p1-ovov454*, which conditions orange-variegated pericarp and cob (Figure 1). The *p1-ovov454* allele contains a *p1* gene in which a full-length *Ac* and a 2039-bp fractured *Ac* (*fAc*) are inserted in the *p1* second intron (Figure 1A) (Yu *et al.* 2011). The 5' end of *Ac* and the 3' end of *fAc* are oriented toward each other and are separated by 822 bp; in this configuration the *Ac* 5' end and the *fAc* 3' end can undergo RET to generate rearrangements such as deletions and inversions (Yu *et al.* 2011, 2012). To identify new RET-induced rearrangements, we screened a population of *p1-ovov454* ears for those with multi-kernel sectors exhibiting loss of *p1* function. One such sector gave rise to an allele (*p1-vvD103*) that specified colorless pericarp with infrequent red sectors (Figure 1C). In crosses with the *Ac* tester stock *p1-ww rm3::Ds*, the *p1-vvD103* allele exhibits normal *Ac* activity (see *Materials and Methods* for details of *Ac* activity test). Interestingly, the *p1-vvD103/p1-ww* plants exhibit irregular seed spacing on the ear and 50% pollen

Figure 2 The *p1-vvD103* allele is associated with a chromosome 1-10 reciprocal translocation. (A) Schematic structures of *p1-vvD103* reciprocal translocation chromosomes 1-10 (upper) and 10-1 (lower). All symbols as in Figure 1. Positions of primers D5, Ac5, P3, and D3 are indicated by the short horizontal arrows. (B) Gel analysis of PCR amplification by primers Ac5 + D5. Lanes 1–4: *P1P2-2*, *p1-vvD103*, *p1-ovov454*; *p1-ww[4Co63]*. Both *p1p2* and *p1-vvD103* alleles show a band at ~5 kb. (C) Gel analysis of PCR amplification by primers D3 + P3. Lanes 2–4: *p1-vvD103*, *p1-ovov454*; *p1-ww[4Co63]*. The *p1-vvD103* allele shows a band at 2.5 kb. (D) PCR analysis of genomic DNA from oat-maize addition lines using primers D3 + D5 to identify reciprocal translocation chromosome. Lane 1, oat; lane 2, maize *p1-ww[4Co63]*; lanes 3–12, oat-maize addition lines with maize chromosomes 1–10, respectively; lane 13, maize *p1-ovov454*. Lane 12, oat-maize addition line containing maize chromosome 10 shows a band of the same size (800 bp) as produced by maize *p1-ww[4Co63]* and allele *p1-ovov454*. (E) Maize sporocyte nucleus (pachytene stage) of *p1-vvD103/p1-ww[4Co63]*. The short and long arms of chromosome 1 and chromosome 10 are designated 1S, 1O, 1L, and 10L. “10C” indicates the centromere of chromosome 10. The breakpoint of desynapsis in chromosome 10 is in the short arm near the end, and the breakpoint in chromosome 1 is in the short arm near the middle.

abortion, indicative of male and female semisterility. These results are consistent with the occurrence of a large inversion or an interchromosomal translocation.

To identify the rearrangement(s) involved, we cloned the sequences flanking the *Ac* and *fAc* junctions by inverse PCR. The sequences flanking the *Ac/fAc* junctions were used in BLASTN searches against B73 RefGen_v2 (Lawrence *et al.* 2004) and were mapped to a locus on the short arm of chromosome 10. Once the target locus was identified, new primers were designed and used in PCR with *p1-vvD103* as template to confirm the validity of both *Ac* and *fAc* junctions (Figure 2, A–C). In addition, primers from the chromosome 10 target locus were used in PCR with DNA from oat-maize chromosome addition lines as templates (Ananiev *et al.* 1997; Kynast *et al.* 2002); the results (Figure 2D) confirmed that the junction sequences were derived from chromosome 10. Finally, we confirmed the presence of a chromosome 1-10 translocation by cytogenetic analyses of meiotic pachytene-stage nuclei from sporocytes of plants of genotype *p1-vvD103/p1-ww*. An example is shown in Figure 2E, which clearly shows a quadrivalent association between the short arms of chromosome 1 and 10. The chromosome 1-10 translocation breakpoint in *p1-vvD103* occurs in *p1* intron 2, and hence places *p1* exon 3 on a different chromosome from

exons 1 and 2; this structure is consistent with the disruption in *p1* function and semisterility observed with the *p1-vvD103* allele.

We propose that the translocation in *p1-vvD103* was generated by RET as shown in Figure 1, A–C. In this model, *Ac* TPase utilizes the reverse-oriented termini of *fAc* and *Ac* in progenitor allele *p1-ovov454* as transposition substrates, followed by insertion of these transposon ends into the chromosome 10S target site. As shown in Figure 1B, the *Ac* 5' terminus is joined to the distal end of 10S, while the 3' terminus of *fAc* is ligated to the proximal side, thereby maintaining the monocentric condition of each chromosome. In addition, we detected a 5.9-kb segmental duplication (labeled by the asterisk in Figure 1C; sequences are listed in File S1) at the translocation junction of *Ac* 5' terminus in *p1-vvD103*. The 5.9-kb segment includes 1601 nucleotides from the *p1* locus (822 bp of *p1* intron 2 plus 779 bp of *fAc*) and 4331 bp sequences duplicated from the translocation junction on chromosome 10. We propose that this duplication was generated by repair of a transposition-induced double-strand break (DSB) through template switching followed by nonhomologous end joining (NHEJ), as explained further in the Discussion.

***P1P2*, the *p1* and *p2* fusion allele induced by reversed-*Ac*-ends transposition**

The progeny ears of *p1-vvD103/p1-ww* exhibited occasional kernels with small red pericarp stripes and sectors. This observation was surprising because the *p1* gene exons 1 and 2 were separated from exon 3 due to the chromosome 1-10 translocation in *p1-vvD103*. In a larger population of *p1-vvD103* ears, we obtained a number of multi-kernel sectors exhibiting distinct red pigmentation. The new phenotype highly resembles that of *p1-ovov454* (Figure 1), suggesting that the structural changes likely involve *p1* and/or *p2* loci, by which we name them as “*P1P2*.”

Like the *p1-vvD103/p1-ww* plants, the *P1P2/p1-ww* plants show irregular seed spacing on the ear and 50% pollen abortion, resulting from male and female semisterility. We also observed a ring of four in diakinesis cells indicative of quadrivalent associations in reciprocal translocation heterozygotes (Table S2). The translocation junction could also be amplified by PCR with primers Ac5 and D5 in *P1P2* alleles in the same way as for *p1-vvD103* (Figure 2B). Collectively, these results indicate that the newly obtained red pericarp phenotype was produced in spite of the chromosome 1-10 translocation. To characterize *Ac* activity, plants carrying *P1P2* were crossed with the *Ac* tester stock *p1-ww r-m3::Ds* (see Materials and Methods for details of the *Ac* activity test). Interestingly, the *P1P2* alleles exhibited developmentally delayed *Ds* excision, distinct from *p1-vvD103*. This phenotype is associated with increased *Ac* copy number resulting in the characteristic *Ac* negative-dosage effect (McClintock 1950, 1952; Heinlein *et al.* 1994; Kunze *et al.* 1995; Kunze and Weil 2002).

To explain the formation of the *P1P2* alleles from *p1-vvD103*, we propose a model of RET during DNA replication

each allele. Lanes in B and C: J—*p1-wv[4Co63]*; 454—*P1-ovov454*; D—*p-vvD103*. Lane 7f contains DNA from a derivative of *P1P2-7* in which Ac has transposed and a newly formed *fAc* is present; hence its PCR patterns are somewhat different from those of lane 7.

followed by insertion of the excised transposon ends into the *p2* locus of the sister chromatid (Figure 1, C–E). This mechanism doubles Ac copy number, maintains the intact translocation junctions, and generates a duplication of sequences located between the insertion and excision sites (~70 kb in size). The duplication junction links the *p1* sequences from the excision site to the *p2* sequences at the insertion site, creating a transcriptional fusion of exons 1 and 2 of the *p1* gene with exon 3 of the *p2* gene. The *P1P2* fusion alleles would retain the 5' promoter region of the *p1* gene and hence should be transcribed and expressed in kernel pericarp.

To test this hypothesis, we performed PCR on genomic DNA of independently-derived *P1P2* alleles using one primer (Ac5) on Ac and three primers spaced along intron 2 and exon 3 of the *p2* gene; these primers should amplify the predicted *p1::p2* duplication junction (Figure 3A). The results confirm the prediction: strong product bands of various sizes are observed in 12 independent *P1P2* alleles (Figure 3B, lanes 1–14), while little or no products are formed in the negative control *p1-wv[4Co63]* (“J”), the grandparent allele *p1-ovov454* (454), or the immediate progenitor *p1-vvD103* (“D”). Further evidence is provided by genomic Southern blot using the probe fragment *pp1*, which detects both *p1* and *p2* genes. The results from a subset of *P1P2* alleles are shown in Figure 3C. In addition to bands derived from the intact *p1* gene (10 kb), the *p1* fragment linked to the chromosome 10-1 translocation (6.5 kb) and the allele from inbred 4Co63 (~6 kb), the *P1P2* alleles exhibit novel bands ranging from 7.9 to 3.9 kb. Importantly, the sizes of PCR products (Figure 3B) and Southern blot bands (Figure 3C) are consistent with their origin from the *P1P2* duplication junction depicted in Figure 3A. Additional tests by semiquantitative PCR and genomic DNA gel blot indicate the presence of duplications in these alleles (Figure S1).

Finally, the PCR products spanning the duplication junctions in each of 11 *P1P2* alleles were cloned and sequenced. The sequences flanking the 5' end of Ac are all derived from

breakpoints in the *p2* gene (summarized in File S2). These results confirm that, in each case, the 5' end of Ac is inserted into the *p2* gene at the positions shown in Figure 4.

Chimeric *P1P2* alleles are expressed in maize pericarp

To explain the pericarp pigmentation phenotype observed in the *P1P2* alleles, we tested the expression of chimeric fusion alleles in the developing pericarp, where *p1* is expressed. We dissected developing kernel pericarp (15–20 days after pollination) from progenitor allele *p1-vvD103* and 10 of 11 *P1P2* fusion alleles, prepared total RNA, and performed RT-PCR (Materials and Methods). Using primers located on exon 1 of *p1* and exon 3 of *p2*, we detected a 620-bp PCR band in all *P1P2* fusion alleles (Figure 5). The size of the 620-bp band is consistent with its origin from a chimeric transcript containing exons 1 and 2 from the *p1* gene and exon 3 from *p2*. These results confirm that the fusion alleles are transcribed in the pericarp.

To assess the potential functionality of the resulting protein, we aligned the deduced *P1P2* protein sequence with that of the Myb-like transcription factor P1 (*Zea mays*) AAL24047.1 (Figure S2). The chimeric protein shares 95% identity with maize P1 protein and contains both the R2R3 Myb DNA-binding region and the putative transcription activation domain (Essers *et al.* 2000; Sidorenko *et al.* 2000; Zhang *et al.* 2000; Kunze and Weil 2002). Considering the high similarity of the P1 and predicted *P1P2* proteins, we consider it likely that *P1P2* is capable of activating the phlobaphene biosynthetic pathway and hence is responsible for the kernel pericarp pigmentation observed in *P1P2* fusion alleles.

In *P1P2-1*, *P1P2-2*, and *P1P2-3*, the Ac insertion site (and hence the *p1/p2* junction) is located upstream of *p2* intron 2, leading to a precursor transcript with a tandem duplication of intron 2 and exon 2 (the first from *p1* and the second from *p2*). We tested for potential alternative splicing of the *p1-p2* transcript in these three fusion alleles by performing RT-PCR

Figure 4 Structures of chimeric *P1P2* fusion genes. (A) Exon/intron structure of *p2* gene with *Ac* insertion sites specific to each *P1P2* allele indicated by numbered red triangles. Arrowhead marked "ATG" indicates translation start codon; stop sign indicates stop codon. (B) Fine structures of *P1P2* fusion genes. Each chimeric allele contains *p1* exons 1 and 2 and partial intron 2, fused via *Ac* with the *p2* gene. Sequences shown are from the *Ac/p2* fusion junctions; sequences from *p2* exons 1, 2, and 3 are shown in gray boxes.

using flanking primers P4 and P5. This test yields a single 620-bp band in these alleles, as for the other fusion alleles, indicating that only one copy of exon 2 remains in the final transcript; this result was confirmed by sequencing the 620-bp RT-PCR product (data not shown). Because *p1* and *p2* exon 2 sequences differ at two SNPs, we examined the signal peaks in the Chromas sequencing profiles to determine which exon 2 is retained. As shown in Figure S3, only *p1* signals were detected in the *P1P2-1* and *P1P2-2* alleles, indicating that *p2* exon 2 is always removed; i.e., alternative splicing does not occur. In contrast, allele *P1P2-3* exhibits mixed *p1/p2* sequence peaks. However, further investigation showed that the proximal duplicated copy of *p2* contains an additional composite insertion (Zhang *et al.* 2014) that drives expression of *p2* in pericarp (File S3). Therefore, the *p2* signals seen in *P1P2-3* may be derived from the ectopically expressed *p2* gene. In summary, two alleles tested (*P1P2-1* and *P1P2-2*) show no evidence of alternative splicing, while for a third allele (*P1P2-3*) the evidence is inconclusive.

Extant maize gene contains duplicated exons derived by alternative transposition

A previous report provided evidence that the RET mechanism has contributed to maize B73 genome evolution by inducing tandem duplications (Zhang *et al.* 2013). To test whether RET events have also been involved in formation of chimeric genes, we developed and applied bioinformatics methods (T. Zuo, unpublished results) to search the maize B73 reference genome sequence (Schnable *et al.* 2009) for genes containing duplications possibly derived by alternative transposition events. One such case (gene AC234515.1; Figure 6, A and B) contains 12

exons; exons 10 and 11 are nearly identical duplications of exons 8 and 9. The duplicated exons are contained within tandem 1054- to 1056-bp duplications that are 99% identical. The same duplicated structure of gene AC234515.1 is also found in the reference genome sequence of another maize inbred line, Mo17 (Xin *et al.* 2013); the duplication in Mo17 is 99% identical (2106/2110) with that of B73.

Interestingly, two *PIF-Harbinger* MITE transposons are found precisely at the duplication endpoints; these MITEs belong to the DTH_Zem5 family (Chen *et al.* 2014, indexed with the accession nos. SQ225117734 and SQ225117735; alignment with the representative MITE SE220500133 is shown in File S4). The duplicated MITEs are 100% identical in sequence, but are <95% similar to other MITEs in the genome. This suggests that these MITEs were generated in a relatively recent transposition or duplication event. However, both MITEs contain two nucleotide substitutions at the 5' TIR and one substitution at the 3' TIR. These changes severely reduce the complementarity of the 14-bp TIRs and would most likely render these MITEs unable to undergo standard transposition. Thus, it is unlikely that these MITEs were generated by two independent standard transposition and insertion events. However, these MITEs and their associated duplications could result from RET, as this mechanism does not require intact termini from the same element. For example, a maize *fAc* element, which contains only the 3' terminus of *Ac*, is capable of undergoing RET in cooperation with a 5' *Ac* termini provided by a nearby *Ac* element (Zhang *et al.* 2009).

We propose that the duplicated structure of gene AC234515.1 was formed by RET during DNA replication

Figure 5 The chimeric *P1P2* alleles are expressed in maize pericarp. (A) Diagram of predicted structure of *P1P2* transcripts, indicating positions of primers P4 and P5. (B) Gel analysis of RT-PCR products produced by primer P4 from *p1* exon 1 and P5 from *p2* exon 3 (620 bp), with the internal control of a *tubulin* gene fragment (322 bp). Total RNA extracted from developing kernel pericarp was used as template for RT-PCR. Depending on whether plants were homozygous or heterozygous at the fusion locus, different intensities of 620-bp bands were obtained for different samples of *P1P2* alleles.

(Figure 6, C–E) as follows: The progenitor gene may consist of 10 exons and contain a *PIF-Harbinger* MITE element inserted into intron 9, as well as a second MITE located downstream of the last exon. Subsequent to its initial insertion, the 5' and 3' termini of the intron 9 MITE underwent the mutations noted above. Because the 3' TIR has only a single mismatch, it could participate in RET with the 5' terminus of a downstream MITE with a complementary TIR. These two MITE termini were excised from the chromatid in an RET event (Figure 6D). The excised MITE termini reinserted into intron 7 of the progenitor gene on the sister chromatid, generating the 1054- to 1056-bp direct segmental duplications of exons 8 and 9 and intron 9 including the MITE. Reinsertion after RET generated the same TSD sequences as those flanking the initial MITE insertions, probably due to the strong preferential insertion at TA sequences during the transposition of *PIF/Harbinger* (Bureau and Wessler 1992; Jiang and Wessler 2001).

Our model presupposes the existence of an intact 5' end from a *PIF-Harbinger* element of the DTH_Zem5 family located downstream of gene AC234515.1. We could not identify this type of MITE terminus within the 1-Mbp distance 3' of AC234515.1 in the maize B73 sequence; possibly it was lost sometime after the RET event due to excision or segregation. However, we did identify other maize genomic MITEs that contain 5' TIRs with perfect complementarity to the 3' TIR of the duplication-associated MITEs (File S5 and File S6). These MITEs with complementary 5' TIR were also flanked by 3-bp TSD (File S6), indicating their competence for transposition. Together, these findings are entirely consistent with the proposed mechanism of duplication via RET involving complementary 5' and 3' TIRs derived from two different MITEs.

In addition, other potential mechanisms of duplication (reviewed in Hastings *et al.* 2009) are inconsistent with the structure and sequence of the duplication in AC234515.1. One well-known mechanism for generating genomic duplications is unequal crossing over, also known as Non-Allelic Homologous Recombination (NAHR). In NAHR, displaced crossing over between two independent MITE insertions in introns 7 and 9 could produce a duplication; however, this crossover mechanism would also result in a third MITE located exactly at the beginning of the duplications (Figure S4). However, there is no MITE present at this position in gene AC234515.1 in either B73 or Mo17. Excision of the

predicted third MITE from its position after the duplication is also unlikely due to the noncomplementary 5' and 3' termini, which would have existed prior to the duplication. Moreover, the fact that the two MITEs in the duplication are 100% identical is entirely consistent with a recent RET event, but is highly unlikely to result from NAHR. If the duplication were generated by NAHR, the two MITEs would most likely be polymorphic since they would have originated from two independent insertions into the progenitor gene, and a crossover between them would have produced a hybrid element. We also discount the possibilities of other known mechanisms such as NHEJ, fork stalling and template switching, and microhomology-mediated break-induced replication. Each of these mechanisms would be expected to occur at random genomic positions, whereas in this case the MITEs are located precisely at the duplication endpoints, strongly suggesting that the MITEs played a causative role in the duplication mechanism. In addition, there is no evidence of microhomologies, ectopic sequences, or filler DNA as would be expected from these repair mechanisms. Taken together, the evidence supports RET as the mechanism that generated the duplications in gene AC234515.1 and is inconsistent with other duplication models.

Discussion

Previous reports have demonstrated the ability of *Ac/Ds* alternative transposition events to generate genome rearrangements in maize, including deletions, duplications, inversions, and translocations (Zhang and Peterson 2004; Zhang *et al.* 2006, 2009, 2013; Huang and Dooner 2008). At each rearrangement breakpoint there is the potential for coding sequences of two different genes to be fused to form a novel chimeric gene. Here we show that such events can occur to generate functional fusion genes. We obtained genome rearrangements using visual screens for changes in maize kernel pericarp pigmentation, a trait controlled by the maize *p1* gene. Starting with a pigmented maize stock containing a *p1* allele (*p1-ovov454*) with *Ac/fAc* insertions in *p1*, we isolated a loss-of-function allele (*p1-vvD103*) with a chromosome 1-10 reciprocal translocation. This translocation breaks the *p1* gene in intron 2, placing the *p1* promoter and exons 1 and 2 on a different chromosome from exon 3. From this stock, we then isolated and characterized a series of 11

Figure 6 RET model for generation of duplications in maize gene AC234515.1. (A) Structure of maize gene AC234515.1_FGT002 downloaded from the Maize Genome Database (<http://www.maizegdb.org>). (B) AC234515.1 contains two directly duplicated segments of 1054 and 1056 bp (gray-shaded arrows), including exons 8, 9, 10, 11, intron sequences, and duplicated *PIF-Harbinger* MITE elements. The MITEs (blue lines) have degraded 5'-end sequences (truncated solid arrowheads) and intact 3' ends (complete open arrowheads). Both MITEs are flanked by TSDs ("TAA") shown as black vertical lines. (C) Prior to RET, the progenitor gene may consist of 10 exons (labeled here as 1–9, plus 12). An intact 5' end (blue triangle) from a *PIF-Harbinger* MITE element is predicted to be located downstream of exon 12. (D) The reverse-oriented 3' and 5' MITE termini (circled in cyan color) flanking exon 12 undergo RET beginning with excision of the termini. The intertransposon segment containing exon 12 is circularized and presumably lost. The excised termini reinsert into intron 7 on the sister chromatid, generating TSDs indicated by vertical orange lines. (E) The expected gene structure generated by RET. The gene contains identical duplicated segments (gray-shaded arrows) containing exons 8 and 9 and MITE elements located precisely at the duplication endpoints. The 5' MITE end located downstream of exon 12 may be subsequently excised or segregated from the AC234515.1 gene.

gain-of-function alleles in which *p1*-coding sequences were joined to a paralog (*p2*) located ~70 kb proximal to *p1*. The fusion junctions occur at various sites in the *p2* gene, but all appear to be capable of producing a chimeric gene product encoded by *p1* exons 1 and 2 and *p2* exon 3. Due to the high sequence similarity of *p1* and *p2*, we propose that their fusion products retain the capacity to regulate flavonoid biosynthesis. Transcript analysis shows that the chimeric genes are expressed in developing kernel pericarp, as would be expected given the genes' location downstream of the *p1* promoter.

We propose that these 11 cases were generated by a type of alternative transposition known as RET (Zhang and Peterson 2004) in which an *Ac* end excised from the *p1* gene on one chromatid is inserted into the *p2* gene on the sister chromatid. This event produces a duplication of ~70 kb representing the DNA between *p1* and *p2*. This model is consistent with previous reports of RET-induced duplications (Zhang *et al.* 2013) and is supported by PCR and Southern blot data.

In an independent study, Zhang *et al.* (2006) documented a series of chimeric *p-oo* alleles (orange pericarp, orange cob)

Figure 7 Template switching followed by NHEJ explains the 5.9-kb duplication at the *p1-vvD103* translocation junction. (A) Replicated chromosomes 1 (black) and 10 (blue) in progenitor *p1-ovov454* allele are shown; chromosome 1 contains *p1* and *p2* loci with *Ac/fAc* insertions indicated in red. The segments to be duplicated are shaded in pink (1601 nt from *p1* locus) and blue (4331 nt from chromosome 10). (B) Excision of *fAc* and *Ac* termini by *Ac* transposase generates DSBs and circularizes the small inter-transposon segment. *Ac* transposase also cuts at the insertion target site (orange vertical line) on chromosome 10, which will join with the *fAc* end to generate the chromosome 10-1 translocation. (C) Repair of DSBs. The DSB at the *Ac* 5' end primes repair replication using the sister chromatid as template, copying *p1* and partial *fAc* sequences. Similarly, the DSB at the distal side of chromosome 10 uses sister chromatid as replication template. (D) The two extended chromatids are joined by NHEJ using the 3-bp microhomology site "GTA," producing a chromosome 1-10 translocation junction containing a 5.9-kb duplication composed of 1601 nt from chromosome 1 and 4331 nt from chromosome 10. (E) Final structure of translocation chromosomes in *p1-vvD103* allele.

also generated by RET events, except that the transposon excised from *p1* reinserted into the proximal *p2* gene of the same chromatid. This results in a deletion of the ~70-kb inter-genic segment and formation of a chimeric gene consisting of

the promoter, exon 1 and exon 2 of *p2* fused to exon 3, and downstream enhancer of *p1* (Sidorenko *et al.* 1999). Although the *p2* gene is normally not expressed in pericarp (Zhang *et al.* 2000), *p-oo* transcripts were detected in kernel pericarp, suggesting that the downstream enhancer of *p1* activates the *p2* promoter. An interesting observation is that the intensity of pigmentation of different *p-oo* alleles was inversely correlated with both the size of the chimeric intron and the distance to the downstream enhancer. In contrast, the pigmentation intensities of the fusion alleles described here do not correlate with intron size (Figure S5). This suggests that splicing efficiency does not appreciably differ for these introns ranging in size from 7.9 to 11.9 kb. Thus, the observed variation in *p-oo* expression is likely due to varying distances between the *p1* 3' enhancer and the *p2* promoter.

The distribution of duplication junctions in *p2* (and hence the *Ac* insertion sites during RET) are distinctly nonrandom: 4 of 11 insertions are located at the 5' region of *p2* gene, including the first two exons, and 6 insertions are located at the 3' region of *p2* intron 2. Only one *P1P2* allele has a breakpoint near the center of intron 2, and the 5' half of intron 2 is devoid of insertions. This pattern suggests that the 5' region of *p2* and 3' region of *p2* intron 2 are *Ac* insertion hot spots and that the 5' half of intron 2 is a cold zone. A similar pattern of nonrandom *Ac* insertion was observed in 21 *p1-vv* alleles derived by intragenic transposition of *Ac* (Athma *et al.* 1992): 15 were clustered in the 5' region of the *p1* gene, including exon 1 and exon 2, and 6 were inserted in the 3' region of *p1* intron 2. To compare the *Ac* insertion preferences between *p1* and *p2*, we aligned the *p1* and *p2* genomic sequences and plotted the insertion sites of *Ac* in *p1-vv* by standard transposition and *p2* by RET (Figure S6 and File S7). The results indicate that the target site preferences of *Ac* insertion during standard and alternative transposition are similar. The areas of frequent *Ac* insertions in both *p1* and *p2* exhibit high sequence similarity, while the regions devoid of *Ac* insertions are more diverse. The conservation of hot spots and dissimilarity of cold zones suggest that insertion site preference is largely a result of preferential targeting to hot spots, rather than avoidance of cold zones.

The chimeric genes isolated here were derived from a loss-of-function allele (*p1-vvD103*) borne on a translocation chromosome, which was itself derived from progenitor *P1-ovov454* by alternative transposition. Previous research has demonstrated that RET followed by insertion of the excised transposon ends into a different chromosome can generate translocations (Zhang *et al.* 2009). However, in the *p1-vvD103* allele described here, the translocation junctions contain an additional 5.9-kb duplication. Sequences of the duplication suggest that it was formed by DSB repair involving template switching followed by NHEJ as described for the origin of an *Ac-immobilized* allele in maize (Conrad and Brutnell 2005). As shown in the model in Figure 7, two pairs of DSBs were initiated by *Ac* TPases during alternative transposition: one pair at the *p1* locus where the alternative transposon excised and the other pair at chromosome 10 where the excised transposon ends inserted. Following

the excision event, the 3' terminus of the alternative transposon was ligated to the proximal side of the chromosome 10 break precisely at the translocation site. While at the 5' DSB, the broken end invaded and copied the sister chromatid for 1601 bp, resulting in an extension of *p1* and partial *fAc* sequences. Similarly, the break at the distal side of chromosome 10 used the sister chromatid as template for replication of 4331 bp from chromosome 10. The two extended fragments annealed at the 3-bp microhomology site "GTA" and were covalently joined by the NHEJ repair mechanism (Lehman 1997).

In summary, we characterized 11 *de novo* segmental duplications and associated functional fusion genes induced by alternative transposition. We also identified an extant gene AC234515.1 in the maize B73 genome that contains an internal duplication apparently generated by an alternative transposition event. These results indicate that *Ac* (and possibly other TEs that transpose via a cut-and-paste mechanism) can generate duplications and functional chimeric genes in the maize genome. The pattern of genetic diversity in maize is attributable to mechanisms such as mutation, recombination, artificial selection, gene conversion, and homology-based gap repair (Gaut *et al.* 2000). Our results demonstrate another possible mechanism for the formation of chimeric genes and further illustrate the potential role of TEs in generating genetic diversity.

Acknowledgments

We thank Meixia Zhao for her great help in the bioinformatics searches of MITE sequences in the maize genome; Terry Olson for technical assistance in the molecular experiments; and Douglas Baker for field assistance. This research is supported by National Science Foundation award 0923826 (to T.P., D.F.W., and J.Z.).

Authors' contributions: D.W. conceived of the study, designed the experiments, carried out the molecular genetic studies, and drafted the manuscript. C.Y. conceived of the study, designed the experiments, and carried out the molecular genetic studies. T.Z. performed a genome search for duplication structures and drafted the manuscript. J.Z. conceived of the study and participated in the design of the experiments. D.F.W. participated in the design of the study and performed the cytogenetics analysis. T.P. conceived of the study, participated in its design and coordination, and drafted the manuscript. All authors read and approved the final manuscript.

Literature Cited

- Ananiev, E. V., O. RieraLizarazu, H. W. Rines, and R. L. Phillips, 1997 Oat-maize chromosome addition lines: a new system for mapping the maize genome. *Proc. Natl. Acad. Sci. USA* 94: 3524–3529.
- Athma, P., E. Grotewold, and T. Peterson, 1992 Insertional mutagenesis of the maize *p*-gene by intragenic transposition of *Ac*. *Genetics* 131: 199–209.
- Bennetzen, J. L., 2005 Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr. Opin. Genet. Dev.* 15: 621–627.
- Bureau, T. E., and S. R. Wessler, 1992 Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell* 4: 1283–1294.
- Chen, J., Q. Hu, Y. Zhang, C. Lu, and H. Kuang, 2014 P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res.* 42: D1176–D1181.
- Conrad, L. J., and T. P. Brutnell, 2005 *Ac*-Immobilized, a stable source of *Activator* transposase that mediates sporophytic and gametophytic excision of *Dissociation* elements in maize. *Genetics* 171: 1999–2012.
- Du, C., A. Hoffman, L. He, J. Caronna, and H. K. Dooner, 2011 The complete *Ac/Ds* transposon family of maize. *BMC Genomics* 12: 588.
- Essers, L., R. H. Adolphs, and R. Kunze, 2000 A highly conserved domain of the maize activator transposase is involved in dimerization. *Plant Cell* 12: 211–223.
- Fedoroff, N., S. Wessler, and M. Shure, 1983 Isolation of the transposable maize controlling elements *Ac* and *Ds*. *Cell* 35: 235–242.
- Gaut, B. S., M. Le Thierry d'Ennequin, A. S. Peek, and M. C. Sawkins, 2000 Maize as a model for the evolution of plant nuclear genomes. *Proc. Natl. Acad. Sci. USA* 97: 7008–7015.
- Grotewold, E., P. Athma, and T. Peterson, 1991 Alternatively spliced products of the maize *p*-gene encode proteins with homology to the DNA-binding domain of Myb-like transcription factors. *Proc. Natl. Acad. Sci. USA* 88: 4587–4591.
- Grotewold, E., B. J. Drummond, B. Bowen, and T. Peterson, 1994 The Myb-homologous *P*-gene controls phlobaphene pigmentation in maize floral organs by directly activating a flavonoid biosynthetic gene subset. *Cell* 76: 543–553.
- Haberer, G., S. Young, A. K. Bharti, H. Gundlach, C. Raymond *et al.*, 2005 Structure and architecture of the maize genome. *Plant Physiol.* 139: 1612–1624.
- Hastings, P. J., J. R. Lupski, S. M. Rosenberg, and G. Ira, 2009 Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10: 551–564.
- Heinlein, M., T. Brattig, and R. Kunze, 1994 In-vivo aggregation of maize activator (*Ac*) transposase in nuclei of maize endosperm and petunia protoplasts. *Plant J.* 5: 705–714.
- Huang, J. T., and H. K. Dooner, 2008 Macrotransposition and other complex chromosomal restructuring in maize by closely linked transposons in direct orientation. *Plant Cell* 20: 2019–2032.
- Hulbert, S. H., T. E. Richter, J. D. Axtell, and J. L. Bennetzen, 1990 Genetic-mapping and characterization of sorghum and related crops by means of maize DNA probes. *Proc. Natl. Acad. Sci. USA* 87: 4251–4255.
- Jiang, N., and S. R. Wessler, 2001 Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. *Plant Cell* 13: 2553–2564.
- Jiang, N., Z. R. Bao, X. Y. Zhang, S. R. Eddy, and S. R. Wessler, 2004 Pack-MULE transposable elements mediate gene evolution in plants. *Nature* 431: 569–573.
- Kermicle, J. L., 1980 Probing the component structure of a maize gene with transposable elements. *Science* 208: 1457–1459.
- Kunze, R., and P. Starlinger, 1989 The putative transposase of transposable element *Ac* from *Zea-mays* l interacts with subterminal sequences of *Ac*. *EMBO J.* 8: 3177–3185.
- Kunze, R., and C. F. Weil, 2002 The hAT and CACTA superfamilies of plant transposons, pp. 565–610 in *Mobile DNA II*, edited by N. Craig, R. Craigie, M. Gellert, and A. Lambowitz. ASM Press, Washington, DC.
- Kunze, R., U. Stochaj, J. Laufs, and P. Starlinger, 1987 Transcription of transposable element activator (*Ac*) of *Zea-mays*-l. *EMBO J.* 6: 1555–1563.

- Kunze, R., S. Kuhn, J. D. G. Jones, and S. R. Scofield, 1995 Somatic and germinal activities of maize activator (Ac) transposase mutants in transgenic tobacco. *Plant J.* 8: 45–54.
- Kynast, R. G., R. J. Okagaki, H. W. Rines, and R. L. Phillips, 2002 Maize individualized chromosome and derived radiation hybrid lines and their use in functional genomics. *Funct. Integr. Genomics* 2: 60–69.
- Lawrence, C. J., O. F. Dong, M. L. Polacco, T. E. Seigfried, and V. Brendel, 2004 MaizeGDB, the community database for maize genetics and genomics. *Nucleic Acids Res.* 32: D393–D397.
- Lechelt, C., T. Peterson, A. Laird, J. C. Chen, S. L. Dellaporta *et al.*, 1989 Isolation and molecular analysis of the maize *P*-locus. *Mol. Gen. Genet.* 219: 225–234.
- Lehman, I. R., 1997 Minireview prologue: eukaryotic DNA repair minireview series. *J. Biol. Chem.* 272: 23463.
- McClintock, B., 1950 Mutable loci in maize. *Carnegie Institution of Washington Yearbook* 49: 157–167 pp.
- McClintock, B., 1952 Mutable loci in maize. *Carnegie Institution of Washington Yearbook* 51: 212–219.
- Messing, J., A. K. Bharti, W. M. Karlowski, H. Gundlach, H. R. Kim *et al.*, 2004 Sequence composition and genome organization of maize. *Proc. Natl. Acad. Sci. USA* 101: 14349–14354.
- Morgante, M., S. Brunner, G. Pea, K. Fengler, A. Zuccolo *et al.*, 2005 Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat. Genet.* 37: 997–1002.
- Porebski, S., L. G. Bailey, and B. R. Baum, 1997 Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15: 8–15.
- Sambrook, J., E. F. Fritsch, and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*.
- SanMiguel, P., and C. Vitte, 2009 The LTR-retrotransposons of maize, pp. 307–327 in *Handbook of Maize: Genetics and Genomics*, edited by J. L. Bennetzen and S. Hake. Springer, New York.
- Schnable, P. S., D. Ware, R. S. Fulton, J. C. Stein, F. Wei *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.
- Sidorenko, L., X. G. Li, L. Tagliani, B. Bowen, and T. Peterson, 1999 Characterization of the regulatory elements of the maize *P-rr* gene by transient expression assays. *Plant Mol. Biol.* 39: 11–19.
- Sidorenko, L. V., X. G. Li, S. M. Cocciolone, S. Chopra, L. Tagliani *et al.*, 2000 Complex structure of a maize Myb gene promoter: functional analysis in transgenic plants. *Plant J.* 22: 471–482.
- Starlinger, P., G. Coupland, H. Fusswinkel, M. Heinlein, R. Kunze *et al.*, 1989 Studies on transposable element Ac in *Zea-mays*. Goldberg, R. (Ed.). *Ucla (University of California-Los Angeles) Symposia on Molecular and Cellular Biology New Series*, Vol. 92, The Molecular Basis of Plant Development, Steamboat Springs, CO, March 26–April 2, 1988. Alan R. Liss, Inc.: New York. Illus: 65–76.
- Tikhonov, A. P., P. J. SanMiguel, Y. Nakajima, N. M. Gorenstein, J. L. Bennetzen *et al.*, 1999 Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proc. Natl. Acad. Sci. USA* 96: 7409–7414.
- Varagona, M., and S. R. Wessler, 1990 Implications for the cis-requirements for *Ds* transposition based on the sequence of the *wxb4 Ds* element. *Mol. Gen. Genet.* 220: 414–418.
- Wang, W., H. Zheng, C. Fan, J. Li, J. Shi *et al.*, 2006 High rate of chimeric gene origination by retroposition in plant genomes. *Plant Cell* 18: 1791–1802.
- Xin, M., R. Yang, G. Li, H. Chen, J. Laurie *et al.*, 2013 Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. *Plant Cell* 25: 3212–3227.
- Yu, C., J. Zhang, and T. Peterson, 2011 Genome rearrangements in maize induced by alternative transposition of reversed Ac/Ds termini. *Genetics* 188: 59–67.
- Yu, C., F. Han, J. Zhang, J. Birchler, and T. Peterson, 2012 A transgenic system for generation of transposon Ac/Ds-induced chromosome rearrangements in rice. *Theor. Appl. Genet.* 125: 1449–1462.
- Zhang, J., and T. Peterson, 1999 Genome rearrangements by non-linear transposons in maize. *Genetics* 153: 1403–1410.
- Zhang, J. B., and T. Peterson, 2004 Transposition of reversed Ac element ends generates chromosome rearrangements in maize. *Genetics* 167: 1929–1937.
- Zhang, J., F. Zhang, and T. Peterson, 2006 Transposition of reversed Ac element ends generates novel chimeric genes in maize. *PLoS Genet.* 2: 1535–1540.
- Zhang, J., C. Yu, V. Pulletikurti, J. Lamb, T. Danilova *et al.*, 2009 Alternative Ac/Ds transposition induces major chromosomal rearrangements in maize. *Genes Dev.* 23: 755–765.
- Zhang, J., T. Zuo, and T. Peterson, 2013 Generation of tandem direct duplications by reversed-ends transposition of maize Ac elements. *PLoS Genet.* 9: e1003691.
- Zhang, J., T. Zuo, D. Wang, and T. Peterson, 2014 Transposition-mediated DNA re-replication in maize. *eLife* 3: e03724.
- Zhang, P. F., S. Chopra, and T. Peterson, 2000 A segmental gene duplication generated differentially expressed myb-homologous genes in maize. *Plant Cell* 12: 2311–2322.

Communicating editor: J. Sekelsky

GENETICS

Supporting Information

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.178210/-/DC1

Alternative Transposition Generates New Chimeric Genes and Segmental Duplications at the Maize *p1* Locus

Dafang Wang, Chuanhe Yu, Tao Zuo, Jianbo Zhang, David F. Weber, and Thomas Peterson

Figure S1 Evidence that the *P1P2* alleles contain tandem direct duplications. **A.** Schematic structures of duplications and *P1P2* fusion alleles on chromosome 1-10 and reciprocal translocation chromosome 10-1. Bracketed segments indicate *HindIII* fragments. Hatched boxes represent the positions of *p1* probe fragment 15. Black arrowheads indicate positions of PCR primers. **B.** Gel profiles of PCR products from *fAc/Ac* junctions. Primers *Ac3* + *P6* + *P7* were used in the same PCR reaction to amplify both junctions of *fAc* and *Ac* with their flanking sequences. 500 bp bands (upper row) were generated by primers *Ac3*+*P7* from *Ac* junctions; and 300 bp bands (lower row) were generated by primers *Ac3*+*P6* from *fAc* junctions. Alleles *p1-ovov454* (454) and *p1-vvD103* (D) have a single copy of each junction and exhibit similar intensities of the 500 bp and 350 bp bands. The *P1P2* alleles generally show higher intensity 500 bp bands compared to the 300 bp bands, indicating an increased copy number of *Ac* flanking sequences as predicted by the duplication model. **C.** DNA gel blot analysis of *P1P2* alleles digested with *HindIII* and hybridized with *p1* probe fragment 15. Compared to *p1-vvD103* (D), the *P1P2* alleles show a higher intensity of 8.3 kb bands compared to the 6.5 kb band, indicating an increased copy number of the *p1* upstream sequences as predicted by the duplication model.

Lanes in B and C: J: *p1-ww[4Co63]*; 454: *P1-ovov454*; D: *p1-vvD103*. Bands of *P1P2*-12 and *P1P2*-14 are from a separate gel.

gi 16507120 gb AAL24047.1 P1-rr-plp2	MGRTPCCEKVGLKRGRTAEEDQLLANYIAEHGEGSWRSLPKNAGLLRCG 50 MGRTPCCEKVGLKRGRTAEEDQLLANYIAEHGEGSWRSLPKNAGLLRCG 50 *****
	Myb DNA binding domain
gi 16507120 gb AAL24047.1 P1-rr-plp2	KSCRLRWINYLRADVVRGNISKEEEDIIKLHATLGNRWSLIASHLPGRG 100 KSCRLRWINYLRADVVRGNISKEEEDIIKLHATLGNRWSLIASHLPGRG 100 *****
gi 16507120 gb AAL24047.1 P1-rr-plp2	DNEIKNYWNSHLSRQIHTYRRKYTAGPDDTAIAIDMSKLQSADRRRGGR 150 DNEIKNYWNSHLSRQIHTYRRKYTAGPDDTAIAIDMSKLQSADRRRGGR 150 *****
gi 16507120 gb AAL24047.1 P1-rr-plp2	PGRPPKASASRTKQADADQPGGEAKGPAAAASSPRHSDVVPNGPNQPNSS 200 PGRPPKTSASRTKHSDADQPGGEAKG--AAASSPRHSDAVNPGPNQPNSS 198 *****:*****:*****:*****:*****
	Activation domain
gi 16507120 gb AAL24047.1 P1-rr-plp2	SGSTGTAEEEGPSSSEDASGPWVLEPIELGDLVWGEADSEMDALMPIPGG 250 SGSTGTAEEEGPSSSEDASGPWVLEPIELGDL-WGEADSEMDALMPIPGG 247 *****:*****:*****:*****:*****
gi 16507120 gb AAL24047.1 P1-rr-plp2	HDSAALEGLGAVGCEAQVDDLFDMWDGFAAHLWGGPEQDEHSAQLRQAA 300 HDSAAQLGLGAVGCEAQVDDLFDMWDGFAAHLWGGPEQDDHSAQLRQAA 297 *****:*****:*****:*****:*****
gi 16507120 gb AAL24047.1 P1-rr-plp2	EPL---VAAAAAATAARTPDDRELEAFETWLLSDSF 335 EPMEAAAVAAAAAATAACTPDDRELEAFETWLLSDSF 335 **:* *****

Figure S2 Alignment of deduced amino acid sequences of P1 and P1P2 proteins. The alignment of maize Myb-like transcription factor P1 (AAL24047.1; upper) with deduced P1P2 amino acid sequence (lower). Myb-like DNA binding domains and putative Activation domains are boxed.

Figure S3 Analysis of potential alternative splicing in *P1P2* fusion alleles. **A.** Gene structures and possible transcript structures for *P1P2-6*, *P1P2-1 or 2*, and *P1P2-3* alleles. We predicted one possible transcript for *P1P2-6* and similar alleles; and two possible forms for *P1P2-1, 2* and *3* resulting from alternative splicing. The alternatively-spliced forms differ in the origin of exon 2 from either *p1* or *p2*. Unlike the other *P1P2* fusion alleles, *P1P2-3* contains a Composite Insertion at *p2* locus [40]. The Composite Insertion carries a part of flanking gene GRMZM2G073064 (orange box) that is expressed in pericarp; transcription initiated from this promoter may produce an additional chimeric transcript resulting in *p2* expression in pericarp.

B. Results of PCR using primers P5+Ac8 (left panel), and primers P4+P5 (right panel). For allele *P1P2-3*, a PCR product was amplified by P5+Ac8, indicating that *p2* is expressed in the pericarp driven by GRMZM2G073064 promoter. The PCR product cannot be from genomic DNA contamination due to the absence of *p1* intron 1 sequences in this PCR product (Figure 5).

C. The Chromas sequence files of RT-PCR products for *P1P2-6*, *P1P2-1*, *P1P2-2*, and *P1P2-3* from the experiment shown in Figure 5. The PCR products were aligned to determine whether a mixture of alternative transcript forms were present in the RT-PCR products. The polymorphic sites between *p1* and *p2* sequences in exon 2 are enclosed in grey boxes. For *P1P2-6* that can generate only one spliced product, we observed a single peak at each polymorphism site, both of which are from *p1*. For *P1P2-1*

and *P1P2-2*, which possess two possible spliced forms, we also observed only *p1* peaks at the polymorphic sites, indicating that no alternative splicing occurred. For *P1P2-3*, a mixture of peaks is present, indicating the inclusion of both *p1* and *p2* sequences. However, we cannot determine whether the observed *p2* signal derives from alternative splicing of the *P1P2* gene or the ectopic expression of *p2* driven by the Composite Insertion.

Figure S4 The predicted structure of duplications induced by NAHR (Non-Allelic Homologous Recombination) or UCO

(Unequal Crossover). Prior to NAHR or UCO, the progenitor gene may consist of 10 exons (labeled here as 1-9, plus 12). Two copies of *PIF-Harbinger* MITE elements (in blue lines) are located in introns 7 and 9, and the segment to be duplicated is indicated by the grey shaded arrowhead. Both MITE elements contain sequence degraded 5' termini (solid, partial triangles) and intact 3' termini (open triangles). TSDs ("TAA") flanking the MITEs are shown as black vertical lines.

- A.** NAHR/UCO occurs between the proximal and distal MITE elements on sister chromatids. Crossing over is indicated by the grey dotted line.
- B.** The resulting duplication structure. Note that this NAHR/UCO mechanism predicts the presence of a third MITE element inserted at the duplication endpoint in intron 7.

Figure S5 Kernel pericarp pigmentation specified by *P1P2* alleles. Orange-red color is from *P1P2*-controlled phlobaphene pigments expressed in kernel pericarp. Dark purple solid color and sectors is from anthocyanin pigments expressed in kernel aleurone and controlled by *r1* and *c1* genes independently of *p1*-controlled phlobaphenes.

Figure S6 Comparison of *Ac* insertion sites from standard transposition in *p1* (ATHMA *et al.* 1992) and alternative transposition in *p2* (from RET; this study). The upper panel plots the insertions of *Ac* in *p1* by standard transposition (ATHMA *et al.* 1992) and the lower panel plots the insertions of *Ac* from RET in *P1P2* alleles. The middle panel shows the alignment of *p1* and *p2* sequences by MUSCLE with color-coded nucleotides at aligned sites by Jalview. The bracket indicates the deleted nucleotides in *p2* intron 2 compared to *p1*. The insertions from standard and alternative transposition in both *p1* and *p2* are in similar positions in conserved regions, suggesting similar insertion site preferences for both modes of transposition.

EXONS are in uppercase; introns are in lowercase, and the *gene embedded in the 4331nt duplications* are in italics.

D. Wang et al.

TGAAAACGAATTGCAGCAGATTGCTCGGTTTGGAGCTCTCCTGTGACCAATTCATGAGGTTTGGGCTACAAGTTGTATAGCAAAGTCAAAT
ACCAGATGAAGGTGAACAAGTTTCATGAGAGGAGTGAGGCTTGCTACTGTGTAAAATGTGGAGAAATGTGTGCTGCAA

File S2. Sequences of *P1P2* fusion junctions.

The PCR bands shown in Figure 3B were directly sequenced. *Ac* sequences are underlined.

>P1P2-1

NNNNNNNNNNNNNGATCGGGANNNNTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTCCCATCTACTTTTCATCCCTGGTC
GGTCCGTAACGTGCACTCTACTAAACGCTAGTGACGTACGTACTCCGTCCGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTTAGCCA
GCACAGCACACACTGGAAAGTGCAAGCTGTAGTGAGACCTGTGCGACTGCCAGCGTGTGTCCGCGCGGTCTGCGCCCGCACGGCCACC
AACTCCCTTGACGCACGCGCGACGCGACCGAGCTGCTAGCCGTGCGCAAGTAGTGCGAGCTCGCCGCCGGCCAGGGTCGCTAGCTCGATT
GGAGGGACAGCTGATACGACTCCGGTGTGGCGCGCATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGA
CGGCGGAGGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCTGGAGGTGCTGCCAAGAATGCAGGTAACCAA
GCCGCGCGCGCCATGCATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAGCTAGCTTCTTCTCGTCGCCGTGCTGTTCTTAGC
TAGTTAGGACGCGCATGCAGGCCTGCTCCGGTGGCGCAAGAGCTGCCGGCTCCGGTGGATCAACTACCTTCGGGGGGGAANNNN

> P1P2-2

NNNNNNNNNNGGGNTNNAATCGATCGGGANNNNNNTAACNAAATCGGTTATACGATAACGGTCGGTACGGGATTTCCCNCTCTACTTTCA
TCCTGCTCCCTTGACGCACGCGCGACGCGACCGAGCTGCTAGCCGTGCGCAAGTAGTGCGAGCTCGCCGCCGGCCAGGGTCGCTAGCTCG
ATTGGAGGGACAGCTGATACGACTCCGGTGTGGCGCGCATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGT
GGACGGCGGAGGAGGACCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCTGGAGGTGCTGCCAAGAATGCAGGTAACC
AAGCCGCGCGCGCCATGCATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAGCTAGCTTCTTCTCGTCGCCGTGCTGTTCTT
AGGTAGTTAGGACGCGCATGCAGGCTGCTCCGGTGGCGCAAGAGCTGCCGGTCCGGTGGATCAACTACCTCCGGCGGACGTCAAGAGG
GGGAACATCTCCAAGGAGGAAGAAGACATCATCAAGCTCCACGCCACCCTCGGGAACAGGTAAACAATAAGCGCGCCCTAATCTCAACGC
TGATCGACTGTGCATCCGACTAGAGAGTAGNACTACTACTTCTTCTTATGCATGGGAGTCAATGCACGCAGTCCCGAAAACTTGGT
ATACGNACTTCTCTTACACNANNNNNNGAAATCTAGCCCANCAATATCAACTTTGATCAAGACATTCATATAAAAAATACAACACNAAC
TTCGCATATATATTATGAAATATATTTATAANAACTTCTAAATAANTAAATGNTGATAGTACTATCAATATTNNTGATTGATGTTTTCTA
AATTGNGAGGNNNATGTATACTTATNTATCTTACTTNAACGNTTCNNACGNCANNNNTGNAGAAAAGGATTTCTGGCNAAGACTNNAAC
NNNNGCAANTNNANNAAGGGCTNCTCNNATGGAACTCNTTNTCTCNAANAAAAANNGNGNNTAACATGCCNGGNTCTATTGCNNNT
NCAATTACTGAAAACTTTAAANCAANNNTAANAANGNGATATAACNNGGNCCAANTNTNGAATNN

> P1P2-3

NNNNNNNNNNNNNCGATCGGGNNANACTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTCCCATCTACTTTTCATCCCTGG
ATACGACTCCGGTGTGGCGCGCATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGACGGCGGAGGAGGA
CCAGTTACTTGCCAACTACATTGCGGAGCACGGCGAGGGGTCTGGAGGTGCTGCCAAGAATGCAGGTAACCAAAGCCGGCGCGCC
ATGCATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAGCTAGCTTCTTCTCGTCGCCGTGCTGTTCTTAGCTAGTTAGGACGCG
CATGCAGGCTGCTCCGGTGGCGCAAGAGCTGCCGGTCCGGTGGATCAACTACCTCCGGGCGGACGTCAAGAGGGGGAACATCTCCAAGG
AGGAAGAANAACATCATCN

> P1P2-4

NNNGNNNNANCGATCGGGANAACTAACNNAATCGGTTATACNGATAACGGTCGGTACGGGATTTCCCATCTACTTTTCATCCCTGCATG
CATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAGCTAGCTTCTTCTCGTCGCCGTGCTGTTCTTAGCTAGTTAGGACGCGCATG
CAGGCTGCTCCGGTGGCGCAAGAGCTGCCGGTCCGGTGGATCAACTACCTCCGGGCGGACGTCAAGAGGGGGAACATCTCAAGGAGG
AAGAAAGACATCATN

> P1P2-6

NNNNNNNNNNNNNNATCGNTCGGGNNAACTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTCCCNCTACTTTTCATCCCT
GGCTTGATGTGGATGCAAAATGCACCCTATAAAATACAATTTCTATAAATTCAGTTAAATGTCACTATATATGCACTCCTACAAAAAGTTCTT
ATGTACCCACAAGGCTAAATACACCCTCTCAATTTACACTCTAGCTTCGCCACTGTGGTAATTGATGAGGTATAAAGCTAGTGGGGTGTGTTG
GTATATATGGACTAATTTTTAGTCCTTCAATTTATTCTATTTTCTAAATGTCAAATACGAAAACATAAATAGAGTTTAACTTTCCGTAT
TTTGCCAATTTAAGACTAACATAAAATAAATAGGTGGGATAAAATTAATCCTCAGAAACCAAACTCTCTAAGTTGATCCGTGCATATATA
TAGTCGTTACATTACATCTCTAAGACGAGAAGCAGCAGCACGAGGAAATGCTCGATCTAGGCATTTCTCGTGCTCAAAGATATGTGCAACTG
AAGGAAGATTAGAAAGTTACTACTGAGCAGTACTGTGATAGCGTCATGCATGCATATATACACACAAACACCTTCCACTCCATAAAATGCA
GTTTTTTTTGGTTACATCAGTTTTGTTAACCTTTTTNCCCAAGCAACTTATGTGATTTGACTGAACCTGGATAAAATATACTGGCCATAGT
ATTTATAATATAATACCAAAAAACAANNNCNNGGNNNTTATTAACAACGGGTTTTTATTAAATTAAGTATTTGGTTTGGCGTAAATTTA
GGGGAACCAAAACGATCAATGAGTAGCGTATATACATAGCTATATCTTATCTANTACAGTACAAAAAATCTTTTCAAAAAATATATAAAN
CAAGCACNCAAAAGCANGNNNTTCGTCGANAGCATCNNNTACCAACNTGACNACNCTATTCNAGGACNANNNANGNANCTTCNTCGAN
GNNNNNNNTTNTTTCNNCNNNTNCNNNGATGANNNGNNNGNNANTNNNNNGNNNNAANAANNANANGNNAAANNNNNNNNNNN
NNNNAANNNAANNCCNANNNNNNNNNNNNNNNNNNTNANNGNACNNNNNNAANANNNNNNNNTNNNGGNAGNNNNNNNNNN
NNNN

> P1P2-7

GATCGGGANAACTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTCCCATCTACTTTTCATCCCTGTCGATCTAGGCATTTCTC
GTGCTCAAAGATATGTGCAACTGAAGGAAGATTAGAAAGTTACTACTGAGCAGTACTGTGATAGCGTCATGCATGCATATATACACACAA
ACACCTTCCACTCCATAAAATGCAGTTTTTTTTTGTACATCAGTTTTGTTAACCTTTTTTCCCAAGCAACTTATGTGATTTTGAAGTGA

GATAAAAATATACTGGCCATAGTATTTATAATATAATACCCAAAAAACAATGTCATGGNTTTTATTAACAACGGGTTTTCTATAAATTAACCT
GATTTGGTTTGGCGTAAATTTAGGGGAACCAAAACGATCAATGAGTAGCGTATATACATAGCTATATCTTATCCTAGTACAGTACAAAAATC
TTTCAAAAAATATATAAACAGCACAAAAAGCATGCGATTCTGTCGACAGCATCGTCTCACCACATGACCACACTATTCAAGGACGATAGATG
GAGCTTCTTCGATGCTCAAACCTTTATTTCTACCATCTACCCAACTGATGATATGTACAGTAAATTAATTAATGCAAGGAANAACAAGATAGG
GAAAAAAACCCCAAACCTGATAGATTANAAGTCGCTGAANAATCCTCTAACAAAACCTGGCGAGCTATCAAACAGGACACGAGAGAGAATA
AATAATTATACAATAATCNCCTCGTGCAATGCANGCCACTGTAGCGCCGTAATGNNANANATTGNGCTATTNNNCCTACAACCCACATGGTG
TTGCNAGAGCTAGCGGNGCCACACACTCATGGNAAACATGGTTTGTGAAAGCAGCTTAATTAATTACTAGCTGANNGTGANAGTGAANAGG
ANCACNCNAATAAACGANNATCTGGNNNCANACTCCTNCNTCATTNAGCTAGCTAGCGCTGGNANCGAACCTCCNTCCNNNGGNNNNAN
NCTGNCTGCTGCCNNNNNNNNNGNNANGNNCNNNTCCNNNGNNNNNNNNNGNCNTTNNNTNGCTNCNNNNNCNNNCNNCNNNGNNCN
GNGCAAGNNNGCATNNNTANNNNNNNNNNNNNNNNNATCTANNNNNNNNGNCNNNNNNNGANNNNNNNGNNNNNNNNNNNNNNAN
NTNCCGCNNNNNNNNNNNNNTGNNNNNNNN

> P1P2-8

NNNNNTNNNNCGTTTCGTTTCGTTTTCGTTTTTACCTCGGGTTCGAAATCGATCGGGATAAACTAACAAAATCGGTTATACGATAACGGTCG
GTACGGGATTTTCCATCCTACTTTTCATCCCTGTATACATAGCTATATCTTATCCTAGTACAGTACAAAAATCTTTCAAAAAATATATAACAA
GCACAAAAAGCATGCGATTCTGTCGACAGCATCGTCTCACCACATGACCACACTATTCAAGGACGATAGATGGAGCTTCTTCGATGCTCAGAC
TTTATTCTACCATCTACCCAACTGATGATATGTACAGTAAATTAATTAATGCAAGGAAGAACAAGATAGGGAAAAAAACCCCAAACCTGAT
AGATTAGAAGTCGCTGAAGAAATCCTCTAACAAAACCTGGCGAGCTATCAAACAGGACACGAGAGAGAATAAATAATTATACAATAATCCCTC
GTGCAATGCATGCCACTGTAGCGCCGTAATGATAGATATTGCGCTATTGCCCTACAACCCACATGGTGTTGCGAGAGCTAGCGGTGCCACAC
ACTCATGGAAACATGGTTTGTGAAAGCAGCTTAATTAATTAC46TAGCTGACAGTGAGAGTGAGAGGACCACGCGAAATAAACGACAATCTG
GTCCAACACTCCCTCTTCATTCAAGCTAGCTAGCGCTGGGAGCGAACTCGCGTCGCATGGCAGAGCCTGCCTGCTGCCGAGCGGACAGGTT
ACGGGNNA

> P1P2-9

NTNNGGNNNNNNNNNGATCGGGANAACTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTTCCATCCTACTTTTCATCCCTGA
TTTCTACCATCTACCCAACTGATGATATGTACAGTAAATTAATTAATGCAAGGAAGAACAAGATAGGGAAAAAAACCCCAAACCTGATAGAT
TAGAAGTCGCTGAAGAAATCCTCTAACAAAACCTGGCGAGCTATCAAACAGGACACGAGAGAGAATAAATAATTATACAATAATCCCTCGTGC
AATGCATGCCACTGTAGCGCCGTAATGATAGATATTGCGCTATTGCCCTACAACCCACATGGTGTTGCGAGAGCTAGCGGTGCCACACTC
ATGGAAACATGGTTTGTGAAAGCAGCTTAATTAATTACTAGCTGACAGTGAGAGTGAGAGGACCACGCGAAATAAACGACAATCTGGTCCAA
CACTCCCTCTTCATTCAAGCTAGCTAGCGCTGGGAGCGAACTCGCGTCGCATGGCAGAGCCTGCCTGCTGCCGAGCGGACAGGTCACGGTCC
ATGTCCGTGGCGTTGGCGTCAGGTCCATTTGCTTTCGCTTTCGCTGCTGCCATCACCGCCGGCCGGTACTGTGCAAAGCGCGCATTGGCTAGCTTCC
CTGTTGCTTGATCTGATCATAGAGGGGTCCATGTCCGTGGCACACAGTGGGATTAATGACGGGCCACGCGGCCCTCCCTAATGATTTCCC
CCGCTTAACGTTGACACTGCGGACGGTGAGAGGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCCATCGTCCTTCGCGAACGAT
GCGTACATTTAGCGCCATATATAACACACGCGGACATGATTACTACGGCTATACGTCGTGAGCGGCGCATCAATGAGTGAGCTAGTTT
GTTGCACCAAATGTGCNATCAAGTTGCATCAGACAGTANAGTACTGTATACTCTCTATATATATATGTTATGTACTAGGCGAGGCTAANNA
AACACGATANANACNNNNNTTCNGNTNNTCATTACNAAGCTGCAGCTGCCANCCAGCAGCGATGATCACGTACNCTNNCNNNNNNNN
NNNNCNNNNNNNN

> P1P2-10

NNNNNNNNANCGATCGGGNNANNTAACNNNTCGGTTATACGATAACGGTCGGTACGGGATTTTCCATCCTACTTTTCATCCCTGAGTGAGA
GGACCACGCAGAAATNNNCGACAATCTGGTCCAACACTCCCTCTTCATTCAAGCTAGCTAGCGCTGGGAGCGAACTCGCGTCGCANGGCGAG
AGCCTGCCTGCTGCCGAGCGGACAGGTACGGTCCATGTCCGTGGCGTTGGCGTCAGGTCCATTTGCTTTGCTGCCTCGCCATCACCGCCGG
CCGGTACTGTGCAAAGCGCGCATTGNCTAGCTTCCCTGTTGCTTGATCTGATCATAGAGGGGTCCATGTCCGTGGCACACAGTGGGATTAAT
GACGGGGCCACGCGGCCCTCCCTAATGATTTCCCCCGCCTTAACGTTGACACTGCGGACGGTGAGAGGCGCGGCTGTGGACGTACGGCGA
GATCTATCGGGGGCCCATCGTCCTTCGCGAACGATGCGTACATTTAGCGCCATATATAACACACGCGCGGACATGANTTACTACNGGCCAT
ACGTGCTGAGCGGCGGATCAANNAGTGAGCTAGTTTGTGCACCAAATGTGATCAANTTGCATCAGACAGTAGAGTACTGTATACTCT
CTCTATATATATATGTTATATGTACTAGGCAAGCTAAGAAACACGATAGATACCACGCNTNCGGTTCTCATNANNANNTGCAGCTGCCAAC
CAGCAGCGATGATCACGTACGCTCNCNTCNGGNNNNNNNNAN

> P1P2-12

GGGANANNTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTTCCATCCTACTTTTCATCCCTGGTGGCGTTGGCGTCAGGTCCATT
TCGCTTTGCTGCCTCGCCATCACCGCCGGCCGGTACTGTGCAAAGCGCGCATTGGCTAGCTTCCCTGTTGCTTGATCTGATCATAGAGGGT
CCATGTCCGTGGCACACAGTGGGATTAATGACGGGGCCACGCGGCCCTCCCTAATGATTTCCCCCGCCTTAACGTTGACACTGCGGACGGTG
AGAGGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCATCGTCCTTCGCGAACGATGCGTACATTTAGCGCCATATATAACACAC
GCCGGCACATGATTACTACGGCTATACGTCTGAGCGGCGGATCAATGAGTGAGCTAGTTTGTGACCAAATGTGCGATCAAGTTGCATC
AGACAGTAGAGTACTGTATACTCTCTATATATATATGTTATATGTACTAGGCAAGCTAAGAAACACGATAGATACCACGCATTGCGTTCTCA
TTACGAAGCTGCAGCTGCCAACACGACGAGATGATCACGTACGCTCACCATCNGGCGTCTTGA

> P1P2-14

NNNNNNNNNNANCGATCGGGNNAACCTAACAAAATCGGTTATACGATAACGGTCGGTACGGGATTTTCCATCCTACTTTTCATCCCTGAT
CATAGAGGGTCCATGTCCGTGGCACACAGTGGGATTAATGACGGGCCACGCGGCCCTCCCTAATGATTTCCCCCGCCTTAACGTTGACAC
TGCGGACGGTGAGAGGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCCATCGTCCTTCGCGAACGATGCGTACATTTAGCGCCAT
ATATAACACACGCGGCGACATGATTACTACGGCTATACGTCTGAGCGGCGGATCAATGAGTGAGCTAGTTTGTGACCAAATGTGCG

ATCAAGTTGCATCAGACAGTAGAGTACTGTATACTCTCTATATATATATGTTATATGTACTAGGCAGGCTAAGAAACACGATAGATACCAC
GCATTGCGTTCTCATTACGAAGCTGCAGCTGCCCAACCAGCAGCGATGATCACGTACGCTCACCATCCNNNGTCCNNN

File S3. Sequence of the *p2* transcript detected in *P1P2-3* allele in kernel pericarp.

The *p2* expression is driven by the promoter of GRMZM2G073064, and includes sequences of *Ac*, *p2* exon1 and *p2* exon 2:

(*AcSequences*:)ATACGATAACGGTCGGTACGGGNNTTCCNTCCTACTTTCATCCCTG(*p2Exon1*:)GATACGACTCCGGTGTGGCGCGCGAT
GGGGAGGGCGCCGTGCTGCGAGAAGGTGGGGCTCAAGCGAGGGAGGTGGACGGCGGAGGAGGACCAGTTACTTGCCAACTACATTGCGG
AGCACGGCGAGGGTCCTGGAGGTCGCTGCCAAGATGCAG(*p2Exon2*:)GCCTGCTCCGGTGC GGCAAGAGCTGCCGGCTCCGGTGGATCA
ACTACCTTCGGGCGGAATTCA

File S4. The alignment of the MITE in the duplication with the representative MITE

```
SE220500133      ---GGCCTTGTTTCGTTTCTGCCGGATTGGTGGGTCGGAACGATTCTGACCGGATTGCTT  57
SQ225117734      TAAGATCTTGTTTCGTTTGTGTCGAATTGGTGGGTCGGAAAGATTCTAGCCGGATTGCTA  60
                  *  *****  **  **  *****  *****  *****  *****
                  *  *****  **  **  *****  *****  *****  *****

SE220500133      CTCTAATTTATATAAACTTTGATTAGCTGGAACGATTCCGGGTGCAATCCGACGAAAACG 117
SQ225117734      CTCTAATTTATATAAACTTTGATTAGCTGTAACGATTCCGGGTGCAATCCGACAAAACG 120
                  *****  *****  *****  *****  *****  *****

SE220500133      AACAAGGCC--- 126
SQ225117734      AACATGGCCTAA 132
                  ****  ****
```

The MITE sequences in the duplication of AC234515.1 (indexed as SQ225117734 by Chen, *et al.* 2014) were aligned with the representative sequences SE220500133 by ClustalW 2.1. The sequences are listed from 5’TIR to 3’TIR (TIR sequences are boxed), including the “TAA” (in bold) as the 3-bp target site duplications.

File S5. The alignment between the 3'TIR of the duplicated MITE (SQ225117735) with a compatible MITE 5'TIR in the genome.

```

SQ225005122      TAAGGCCATGTTTCGTTTCTGTCGGATTGGTGGGTCGGAACGATTCCGAGCCGG--ATTGC 58
SQ225117735      TTAGGCCATGTTTCGTTTGTGTCGGATTG--CACCCGGAATCGTTAC-AGCTAATCAAAGT 57
*  *****

```



```

SQ225005122      TTCTCTAATTTATATAAAC--TTTGATTACCGAAACGATTCCGGGTG--TAATCCGATG 114
SQ225117735      TTATATAAATTAGAGTAGCAATCCGGCTA---GGAATCTTTCCGACCCACCAATTCGACA 114
*  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

```



```

SQ225005122      TAACCGAACAAGACCTAA 132
SQ225117735      CAAACGAACAAGATCTTA 132
*  *****

```

An example (SQ225005122) of the MITE with a compatible 5' TIR with the 3' TIR of the duplicated MITE (SQ225117735) in the RET reaction is shown. SQ225005122 is shown from 5' to 3' and SQ225117735 is from 3' to 5'. The compatible TIR of RET is enclosed in black box. The 3-bp TSD sequences are in bold. The MITE sequences were indexed in Chen, *et al.* 2014. The alignment was calculated by ClustalW 2.1.

File S6. 5' TIRs with perfect complementarity to the 3' TIR of the duplication-associated MITE SQ225117735 are found in multiple MITEs in the maize B73 genome.

SQ225158225	TTA GGCCATGTTTCGTTT T TGTGTCAATCCATGGGGATTGAGGGGGAACAAATCCCCTATAA
SQ225163989	TTA GGCCATGTTTCGTTT T TGTGCCAATCCATAGGGATTGAGGGGGATTAAATCCCCTACAA
RC_SQ225030358	TTA GGCCATGTTTCGTTT T TGTGCCAATCCATAGGGATTGAGGGGGATTAAATCCCCTACAA
RC_SQ225082017	TTA GGCCATGTTTCGTTT T TGTGCCAATCCATAGGGATTGAGGGGGATTAAATCCCCTACAA
SQ225027317	TTA GGCCATGTTTCGTTT T TGTGCCAATCCATAGGGATTGAGGGGGATTAAATCCCCTACAA

SQ225158225	GTCAAATCCCCCTCAATCCCTTCCAATCCCCCTCAATCCCCATGGTTTGGGGATGAAAC
SQ225163989	GTCAA-ATCCTCTCCAATCCCTTCCAATCCCCCTCAATCCCTATGGTTTGGGATCAAAC
RC_SQ225030358	GTCAAATCCTCTCCAATCCCTTCCAATCCCCCTCAATCCCTATGGTTTGGGATCAAAC
RC_SQ225082017	GTCAAATCCTCTCCAATCCCTTCCAATCCCCCTCAATCCCTATGGTTTGGGATTAAAC
SQ225027317	GTCAAATCCTCTCCAATCCCTTCCAATCCCCCTCAATCCCTATGGTTTGGGATTAAAC

SQ225158225	GAACATGGCC TTA
SQ225163989	GAACATGGCC TTA
RC_SQ225030358	GAACATGGCC TTA
RC_SQ225082017	GAACATGGCC TTA
SQ225027317	GAACATGGCC TTA

Alignment of maize genomic MITEs with perfectly complementary 5' and 3' termini (enclosed in boxes). The perfect MITEs have

5' TIRs complementary to the 3' TIR of the duplication-associated MITE SQ225117735 and are flanked by 3-bp TSDs (in bold).

The MITE sequences are all indexed by Chen, *et al.* 2014 with SQ numbers.

File S7. Comparison of *Ac* insertion sites from standard transposition in *p1* (ATHMA *et al.* 1992) and alternative transposition in *p2* (this study)

Intergenic sequences and introns are shown in plain type; Exons are shown in bold type; *Ac* insertion sites in *p2* are shown in blue type; *Ac* insertion sites in *p1* are shown in red type.

```

p2      TCCCCCAGTCCCACGTAACGGACAGGTCGGTCCGTAACGTGCACTCTACTAAACGCTAG 480
p1      CTGTTTAAAGGGCTAAAATACAAACG-----TGCACCTCTGCACTCTACTAAGCGCTAG 295
      .** * *..:*.**.*.          * *. *****.*****

p2      TGCACGTACG----TACTCCGTCCGTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC 536
p1      TGTACGTACGTACGTACTCCGTCGGCTGCTATATTATGGCCGGCCGTGGCGTGCCCTCTC 355
      ** *****

p2      TAGCCAGCACAGCACACACTGGAAGTGCAAGCTGTAGTGAGACCTGTGCGACTGCCA 596
p1      TAGCCAGCACAGCACACACTGGAAGTGCAAGCTGTAGTGAGACCTGTGCGCGACTGCCA 415
      *****

p2      GCGTGTGTCCGCGCG-----GTCGTGCGCCCGCACGGCCACCAACT 637
p1      GCGTGTATCCGCGCGGCAAGGAGCGTAGCGCGCGGTGTCGCGCCCGCACGGCCACCAACT 475
      *****

p2      CCCTTGAGCGCACGCGCGCACGCGACCAGCTGCTAGCCGTGCGCAAGTAG---TGCGAGC 694
p1      CCCTTGAGCGCACGCGCGCGCGACCAGCTGCTAACCGTGCGCAAGTAGTAGTGCGACT 535
      *****

p2      TCGCCGCGGCCAGGGTCGCTAGCTCGATTGG--AGGGACCAGCTGATACGACTCCGGTGT 753
p1      TCGCCGCGGCCGGGATCGCTAGCTCGATCGATCGCGGGACCATACGACTCCGGTGT 595
      *****

p2      G----- 754
p1      GGCCAGCGGCGCGCGCGGGAACGCACGTGCTGCGAGCGAGGCGAGGGCAGACGCTAG 655
      *

p2      -----GCGCGCGATGGGGAGGGCGCCGTGCTGCGAGAAGGTGGGG 794
p1      CTGTTGCCGGGAGCTAGCCGCGCGCGATGGGGAGGACGCCGTGCTGCGAGAAGGTGGGG 715
      *****

p2      CTCAAGCGAGGGAGGTGGACGGCGGAGGAGGACCAGTTACTTGCCAACCTACATTGCGGAG 854
p1      CTCAAGCGAGGAGGTGGACGGCGGAGGAGGACCAGTTACTTGCCAACCTACATTGCGGAG 775
      *****

p2      CACGGCGAGGGGTCCTGGAGGTCGCTGCCCAAGAATGCAGG-TAACCAAAGCCGGCCGCG 913
p1      CACGGCGAGGGGTCCTGGAGGTCGCTGCCCAAGAATGCAGGTAAACCAAAGCGGCCGCG 835
      *****

p2      CGCATGCATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAGCTAGCTTCTTC 973
p1      CGCCATGCATCGCCACGTAGCATCAATCTCCGATCCATGCATATATGAG----CTTCTTC 891
      *****

p2      TTCGTCGCCGTGCTCGTCTTCTTAGCTAGTTAGGACGCGCATGCAGGCTGCTCCGGTGCGG 1033
p1      TTCGTCGCCGTGCTCGTCTTCTTAGCTAGTTAGGACGCGCATGCAGGCTGCTCCGGTGCGG 951
      *****

p2      CAAGAGCTGCCGGCTCCGGTGGATCAACTACCTCCGGGCGGACGTCAAGAGGGGGAACAT 1093
p1      CAAGAGCTGCCGGCTCCGGTGGATCAACTACCTCCGGGCGGACGTCAAGAGGGGGAACAT 1011
      *****

p2      CTCCAAGGAGGAAGAAGACATCATCATCAAGCTCCACGCCACCCCTCGGGAACAGGTAACA 1153
p1      CTCCAAGGAGGAAGAAGACATCATCATCAAGCTCCACGCCACCCCTCGGGAACAGGTAACA 1071
      *****

p2      ATAAGCGCGCCCTAATCTCAACGCTGATCGACTGTGCATCCGACTAGAGAGTAGTACTAC 1213
p1      ATAAGCGCGCCCTAATCTCAACGCTGAT-CTGTGCATCCGACTAGAGAGTAGTACTAC 1130
      *****

p2      TACTACTTCCTTCTTTATGCATGG-----GAGTCAATGCACGCAGTCCCGAAAACTTG 1268
p1      TACTACTTCCTTCTTTATGCATGGGATGGGAGTCAATGCACGCAGTCCCGAAAACTTG 1190
      *****

```

p2 GTATACGTACTTCCTCCTTCACACGAAGAACGGAAATCTAGTCCAACAATATCAACTTTG 1328
p1 GTATACGTACTTCCTCCTTCACACGAAGAACGGAAATCTAGTCCAACAATATCAACTTTG 1250

p2 ATCAAGACATTTCATATAAAATACAACCTACTAAGTTCGCATATATATATGAAATATATT 1388
p1 ATCAAGCATTCATAT-----ATATTATGAAATATATT 1283

p2 TTATAAGAACTTCCATAAATATATAAATGTTGATAGTACTATAAATATAGTTGATT--G 1446
p1 TTATAAGAACTTCCATAAATATATAAATGTTGATAGTACTATAAATATAGTTGATTGAG 1343

p2 AGAGTTTCTTAAATTGTGAGGTACCATGTATACTT--ATATATACTACTCAAACGCTTC 1504
p1 AGAGTTTCTTAAATTGTGAGGTACCATGTATACTTTTATATATACTACTTAAACGCTTC 1403

p2 TCGACGTACAAATTTGGAGAAAAGGATTTCTGGCAGAAGACTAGAACAGTCAGCACAGAC 1564
p1 TCGACGTACAAATTTGGGGAAGGATCTCTGGCACAAGACTAGAACAGTCAGCACAGAC 1463

p2 AATAAAGGGCTACTCCGATGGAATCTCGTTTTCATTCATAGGAG-----AGCAGTGTAG 1619
p1 AATAAAGGGCTACTCCGATGGAATCTCGTTTTCATTCATAGGAGTATAGAGTAGTGTAG 1523

p2 TAACAATTGCCTGCTCTATATGCTCTCTCAATTACTGAAATCTTAGAGCAAGTATAATAA 1679
p1 TAACAATTGCCTGCTCTATATGCTCTCTCAATTACTGAAATCTTAGAGCAA----- 1574

p2 ATGGATATAAGCCGGCTCAATATTGAGATTGATGAGAGAGAGAGCAGAATCGGACTGCAA 1739
p1 -----ATATTGAGATTGATGAGAGAGAGAGCAGAATCGGACTGCAA 1615

p2 ACTTAAAGCCGGGGCTTAGACACAGAATAAAACCTTTTGTG-TGAGAAAGACATGCAAA 1798
p1 ACTTAAAGTCGGGGCTTAGACACAGAATAAAACCTTTTGTGAGAAAGACATGCAAA 1675

p2 TCATAAATTAACAGTGAAGAGCTAACTACTATACGGATAG----- 1838
p1 TTATAAATTAATAGTGAAGAGCTAACCTACACGGATAGGCTAATGTAATCATAAATT 1735
* *****

p2 -----GCTAAGAAAGTAGACTATAATGTTTCCTTAC 1867
p1 AATAGTAAAGAGCTAACTACTACAGATAGACTAAGAAAGTAGACTATAAGGTTTCCTTAC 1795

p2 TTACAACCAGCGGTTAGCTATATTATTAACCTTGATCTTATAATTTTCGATGGCTGGCCCG 1927
p1 TTATAACCAGCGGTTGACTATATTATTAACCTTGATCTTATAATTTTCGATAGCTGCCCCG 1855
*** *****

p2 ATCGGTTGAGAGTTAAGACACCGTATACCATACACGTACACAGTACTAAATAAAGAATTA 1987
p1 ATCGGTTGAGAGTTAAGACACCGTATACCATACACGTACACAGTACTAAATAAAGAATTA 1915

p2 AATGATGTAGGGGCACACGCTGCTGGAATTGAGGCGCGCAGCAGCAATAGTAGGCTAA 2047
p1 AATGATGTAGGGGCACACGCTGCTGGAATTGAGGCGCGCAGCAGCAATAGTAGGCTAA 1972

p2 TGTCTTTTTTT-CTATGCTTGGTATGGGTTCCAATCGGCTTTATTATCACTGTTTAGCTA 2106
p1 TGTCTTTTTTTCTATGCATGGTATGGGTTCCGATCGGCTT---TATCACTGTTTAGCTA 2029

p2 CTGTAGTGATAATGTCCAAAAGATGGTTTCATCAGCGAAAGACAAAAGAGAGAGAAGAG 2166
p1 CTGTAGTGATAATGTCCAAAAGATGGTTTCATCAGCGAAAGACAAAAGAGAGAGAAGAG 2089

p2 AGAAGAGAGAAGAGAGAAGAGAGAG--AGAAAGAAAGACTCCTATAAAGAGTAGTGTG 2223
p1 AGAAGAGAGAAGAGAGAAGAGAGAAGAGAGAGAGAGAGACTCCTATAAAGAGTAGTGTG 2149

p2 CTATAGTGCTAGAATGGTGCATATTACTACAGTGCATATATGTGAGAAAGACCGGAGCT 2283
p1 CTATAGTGCTAGAATGGTGCATATTACTACAGTGCATATATGTGAGAAAGAGCCGGAGCT 2209

p2 ATTGAGCTAATACAATCTATAGGGCTAAAT---AGAGACATGGCTG-----AGG 2330

p1 ATTGAGCTAATACAATCTATAGGGCCTAAATTAATAGAGACATGGCTGAAAACAATATGG 2269
***** : **

p2 CCATCTGGTCCCTCCGATCGATGTGCTGTAGTATGTTTATGTACACGCGGGCTCTACGCA 2390
p1 CCATCTGGTCCCTCCGATCGATGTGCTGTAGTATGTTTATGTACACGCGGGCTCTACGCA 2329

p2 CGCTTGCTCTAGCATTGAGCAGTGCTACTTTACTACACT--ATCAGTCCGGTCCTTGTC 2447
p1 CGCTTGCTCTAGCATTGAGCAGTGCTACTTTACTACACTACTATCAGTCCGGTCCTTGTC 2389

p2 AGTCACTGGCTCAGCATGTGTCTCTGAAGTTGCCCGTACGTGTGGCAGCAGCCAAAAGGG 2507
p1 AGTCACTGGCTCAGCATGTGTCTCTGAAGTTGCCCGTACGTGTGGCAGCAGCCAAAAGGG 2449

p2 CGTGCTGCGTGTGAAGTTTACTGTCCAGTCAGATGCCTGTGATTTCCGAGGCCACGTGC 2567
p1 CGTGCTGCGTGTGAAGTTTACTGTCCAGTCAGATGCCTGTGATTTCCGAGGCCACGTGC 2509

p2 ACGAGCATGCATGACGCCAACATGGTATGATAGCCCTCTTTGTTTAGGCTTTTTTTTCGG 2627
p1 ACGAGCATGCATGGCGCCAACATGGCATGATG-TGATGTTTGTTCGGCCTTTTTTTTAG 2568
*****.*****.*****.* *****.* ** *****.*

p2 TTTATGGCCATCAAAATCTGTACAACTGTCAAACACTCAGCTTTTTCAGCTCGCTTCTA 2687
p1 ATTTTGGCTACCAAAATCTGGTACAGACAGTCAAATATTCA-ATTTTGTAGTCCACTTCT- 2626
*: ** : ** * ***** ** : ** : ** * * * . * * * * * . *

p2 AAAAAATCGATTTGGTAAAAATCAACAACAAAGTCG-CATGAACACAAAATTGGTCGG 2746
p1 AAAAAATCGATTTGGTAAAAATCATCCAACAAAATCAACATAAACACAAAATCGGTCGG 2686
*.*****:*****.*.***.***** *****

p2 GTCGTTGCAATATAGTAGGATTCGGTCACTTTCTAGATCATGAACCATATGTACACATTC 2806
p1 -ACGTTGCAATA-GTAGGAATCTATCACTTTCTAGATTGTTAACCTATAGACACATTA 2743
:***** *****:* *****.* *****.* *****.*

p2 ATCTTTCTCCGCACATAATCC-CACAACTGCTCAGATTATCTCCACAGCCAAATCTCT-G 2864
p1 ATCTTCTCCGCACATAATCCACAACTACTCATATTCTCTCTGCAGCCAAATCTTTTA 2803
***** *****.***** **.* *****.***** *

p2 AAAAGCTGGTTAGAAAAAGTTCAACCAAACAGTCCCGATGATGCCCGCCACTTCTTATG 2924
p1 AAAAATCGGTAAGAAAAAGATGAACCAAACAGGCTCGATGATGCTGCCACTTCTTATG 2863
****.*****:*****:* ***** * ***** *****

p2 GCTTCGCATATGTGGCAACGGGGCGAACTAGGTTTCCTTGATGTGGATGCAAATGCACC 2984
p1 ACTTCGCATATGTGGCAAGAGGGGCG-AAATAGGTTGCTAACTGTGGATGCAAATGCACC 2922
.******.***** **.******:.* *****

p2 CTATAAAATACAATTTCTATAAATTCAGTTAAAAATGTCACATATATGCACCTCTACAAA 3044
p1 CTGTAAATACAATTTCTATAAATTCAGTACAATTTCACTATATATGCACCTCTAGAAA 2982
.****.*.*** ***** *

p2 AAAGTTCTTATGTACCCACAAGGCTAAATACACCTCTCAATTCTACACTCTAGCTTCGC 3104
p1 AAG-TTCTTATCTACCCACAAGACTAAATACACCTCTCAAT--TATAATCTGGCTTCGC 3039
.* ***.***** **.*.*****

p2 CACTGTGGTAATTGATGAGGTATAAGCTAGTGGGGTGTGGTATATATGGACTAATTT 3164
p1 TATTATG--TGGTAATGAGGTCTAAAAGTAG-GGGGTGTGGTATATATGGACTAATTT 3096
* *.** :.*.*****.*****.*** *****:*****

p2 TAGTCCTTCATTTTATTTCTATTTTAGTTTCTAAATGTCAAATACGAAACTAAAATAG 3224
p1 T-----AGTTTCTGTATT----- 3109
* *****.*:***

p2 AGTTTTAACTTCCGTATTTTGCCAATTTAAAGACTAACATAAAATAAAATAGGTGGGAT 3284
p1 -----TGACAATTTAAGACTAAAATAAAATAAAATAGATAGGCT 3149
.****.*****.*****.*.*.*

p2 AAAAAATTAATCCTCAGAAACCAACACTCTCTAAGTTGATCCGTCATATATATAGTCGTT 3344
p1 AAAAAATTAATCTTTAGAAACCAACACTCTCTAAGCTGATCCGTCATATAT---TCGTT 3205
***** * ***** ***** *****

p2 ACATTACATCTCTAAGACGAGAAGCAGCAGCAGGAGAAATGTCGATCTAGGCACCTTC 3404
p1 ACATTTTCATCTCTAAGACGAAAAGCAGCAGCAGGAGAAAT---GATCTAGGCACCTTC 3261

```

*****;*****.*****
p2 TCGTGC TCAAAGATATGTGCAACTGAAGGAAGATTAGAAAGTTACTACTGAGCAGTACTG 3464
p1 TCGTGC TCAAAGATTTTGTGCAACTAAGGGAAGATTAGAAAGCTACTACTGAGCAGTACTG 3321
*****;*****.*.***** *****

p2 TGATAGCGTCATGCATGCATATATATACACACAAACACCTTCCACTCCATAAAATGCAGT 3524
p1 TGATAGCGTCATGCGTGCATATA-----CACACACACCTTCCACTCCATAAAATGCAGG 3375
*****.***** ****.*****

p2 TTTTTTTTGTACATCAGTTTGTGTTAACCT-TTTTTTCCCAAGCAACTTATGTGATTT 3583
p1 GTTTTTTTTGCTCCATCAGTTTGTGTTAACCTATTTTTTCCCAAGCAACTTCTTTGATTT 3435
***** *.***** *****.* *****

p2 TGACTG----- 3589
p1 TGACCAGTGGCGGGACGAAAGTCCAGGTGAGGCCAATCCGTAATAAAGAAATCACAAGAA 3495
****.

p2 -----A 3590
p1 AAAATGCTAAACTATAAGTGGTTTTTTTGGTAATAGCATGAAATTAAGTTGTAATATCATG 3555
.

p2 ACTTGGATAAAAAATAT----- 3606
p1 ACTATTATAAAATAATGTGGATATCGAAAATCTTTTAGCGACGCCAAATCATCGATAGA 3615
***; *****;:**

p2 -----
p1 AAATTCAATCTACAATTAGAAGAAAAAGTATTAGGTAAGCATCAATGTAAGTAAAAGACA 3675

p2 -----
p1 AAACATATAGAAAAACATCGAAAAACACCATTAAACCATCAATCTTGTGGTTGTGGCAG 3735

p2 -----
p1 CTTGCACTCTAGTGCGGGGCCGTGGCCGAGTGGAGGCAGGAAGACGAGCAGCAAGCTAGC 3795

p2 -----ACTGGCCAT----- 3615
p1 GACAGTGTCTGACAGCTGGCGATGACGGTTGGGTGGGCGGCCAACGATGATGAACCTGAC 3855
.***** **

p2 -----
p1 GGTGGACGTC AATTGGACAGCGGGCACAACGCTAGGACGAGCGCTAAGCGGGGGTTGGA 3915

p2 -----
p1 TGCTACAAACGTGCGGTTGATGCCTTTCTCATCTGTGTTTTATTTCGATATAGGTCGTGA 3975

p2 -----
p1 CCTAAGATGATGGGCTAGATCTTGTGGGTTAGTCGATGGTATGATGTGGTGGACCCGGTG 4035

p2 -----AGTATTTATAATA----- 3628
p1 TTTATCTAATTTAGTCGCATATACCTTATATTTTTTATTTTATAGTATATATAAAAGGTA 4095
*****;*****;*

p2 -----
p1 GGTATAAACAGTTC CAAAAATTAAGGTGGCGCCATGGCCTACTTTGCCACATTGTGGAT 4155

p2 -----TAATACCC 3636
p1 CCGCCCCTGATTTTGACTGAACTTGGATAAAATATACTGGTCATAGTATTTATAATACCC 4215
*****

p2 AAAAAACAATGTCATGGCATTATTAACAACGGGTTTTCATTAAATTAACTGATTGG 3696
p1 AAAAAACAATGTCATGGCATTATTAACAACGGGTTTTCATTAAATTAACTGATTGG 4275
*****

p2 TTGCGCGTAAATTTAGGGGAACCAAAACGATCAATGAGTAGCGTATACATAGCTATAT 3756
p1 TTGCGCGTAAATTTAGGAGAACCAATATGATCAATGAGTAGCG-----TATAT 4323
*****.* ***** *****

```

p2 CTTATCCTAGTACAGTACAAAAATCTTTCAAAAAATATATAAACAGCACACAAAAGCA 3816
p1 CTTA---CAGTACAG---AAAAATCTTTGAAAAAATATATAAACAGCACACAAAAGCA 4376
**** ***** *****: : : : : *****

p2 TGC GATTCGTCGACAGCATCGTCTCACCACATGACCACACTATTCAAGGACGATAGATGG 3876
p1 TGC GATTCGTCGACAGCATCGTCTCACCAGTCTCACCACACCATTCAAGGACGATG---G 4432
***** . * ***** ***** *

p2 AGCTTCTTCGATGCTCAGACTTTATTTCTACCATCTACCCAAACTGATGATATGTACAGT 3936
p1 AGCTTCTTCGATGCTCAGACTTTATTCTACCATCTACCAACCCAAACTGATATGTACAGT 4492
***** . ***** *****

p2 AAATTAATTAATGCAAGGAAGAACAAGATAGGAAAAAAACCCCAACTGATAGATTAG 3996
p1 AAAT-----GGAAGGAAGAACAAGATAG-AGAAAAAAACCCCAACTGATAGATTAG 4543
**** * ***** . *****

p2 AAGTCGCTGAAGAAATCCTCTAACAAAACCTGGCGAGCTATCAACAGGACACGAGAGAGA 4056
p1 AAGTCGCTGAAGAAATCATCTAACAAAACCTGGCGAGCTATCAACAGGACACGGGAGAGA 4603
***** . ***** *

p2 ATAAATAATTATACAATAATCCCTCGTGCAATGCATGCCACTGTAGCGCCG-----TAAT 4111
p1 ATAGATGATTAAACAATAATCCCTCGTGCAATGCATGCCACTGTAGCGCCGTAATATAAT 4663
*** . ** . *** . ***** *****

p2 GATAGATATTGCGCTATTGCCCTTACAAAC----- 4140
p1 GATAGATATG-CGCTATTGCTCCTACAACCTACAACTAGGGATGAAAACGGTCGGTAACGG 4722
***** *****

p2 -----
p1 TCGGTAAATACCTCTACCGTTTTCATTTTCATATTTAACTTGCGGGACGGAAACGAAAA 4782

p2 -----
p1 CGGGATATACCGGTAAACGAAACGAACGGGATAAATACGGTAATCGAAAACCGATACGAT 4842

p2 -----
p1 CCGGTGCGGTTAAAGTCGAAATCGGACGGGAACCGGTATTTTGTTCGGTAAATCACAC 4902

p2 -----
p1 ATGAAAACATATATTCAAACTTAAAAACAAATATAAAAAATGTAAACACAAAGCTTAA 4962

p2 -----
p1 TTAAACATAGATAAAATCCATATAAATCTGGAGCACACATAGTTTAAATGTAGCACATAAG 5022

p2 -----
p1 TGATAAGTCTTGGGCTCTTGGCTAACATAAGAAGCCATATAAGTCTACTAGCACACATGA 5082

p2 -----
p1 CACAATATAAAGTTTAAACACATATTCAATACTACTTGCTCACATCTGGATCACTTAGC 5142

p2 -----
p1 ATGCATAAACTATTACAACCAAGGCTCATCTGTCAACAAACATAAGACACATTGCTCATG 5202

p2 -----
p1 GAGAGGAGCCACTTGCTACATCTTCATATTCTTAGAAAAATCTATTGCGTCTTCATCCT 5262

p2 -----
p1 GTTAATACACAAAAATAAGTCAGTTTTGGATAAATAAATACATATAGAAGAACATGAATT 5322

p2 -----
p1 GATATGCAGGGAGTATAAATAAATACATATAGGAGAACATGAATCTGTGAACCTAACACGG 5382

p2
p1

CTGGGAGCTAGGCAGCTAGCAGCTAGCGCCTAACAGCTGGGAGCCTAACAGCTAGCAGCT 5442

p2
p1

AGCAGCCAATCAAAACAAGGCGACAAGGCGCATGCAGTGAGATCAAAAATCTGTTAATGC 5502

p2
p1

CAGCCATGCAGGGAGTATAACACGGCTGGGCAGCAAGGCGCATGCATCAAAACAAGGCGA 5562

p2
p1

CAGCAACAGCCCATGCATCAAAACAGTAGTGAATAATAGCAAATTAATAGCCCATGCAC 5622

p2
p1

GAAGTAAATAAATCTTTAAATACCTCATCCATATGATTCTCATGATTTGTTGCAGCAG 5682

p2
p1

CAATAACAGAGTCTAGCACCTCGAGATCACCAATCATTTGTTGGAAAATATGTAGCACCTT 5742

p2
p1

GAATGACACAAATATGCATCAATATAAGTAAAATAATTGTTGAATAACTATAAATTGGAA 5802

p2
p1

CTTCATATAACATATATGCATTCACCTTTTCTAGATGCTGCTACCCAATCTTTTGTGCA 5862

p2
p1

TATCAAAGCTTCAACAATCTCCGAACCAAGACGATTGCGGTAAGGATCAACAACACGACC 5922

p2
p1

ACCAGCACTGAACGCAGACTCAGAAGCAACAGTTGACACTTGTATTGCTAGCACATCCCT 5982

p2
p1

TGCAATTGGGTGAGAATAGGATATTCTGCAACCCCTCCCTCCACCATGATAAAATATC 6042

p2
p1

AAACTGACCACATATGCTTCAAAAGGGGTTCAGACATATATTTATCCAATTCATTTGACTC 6102

p2
p1

TACTTGATCATAATCCTTCAACTCATGCAAAATAGTTTGAATTCATCATCTTCATTTTC 6162

p2
p1

CATCAAGGTATCATCCATACATCATTAGTAGTGTCTTTGTCTTTGGAGCTGAAGGACT 6222

p2
p1

ACAAC TAGAATAGAATTGATACAATTTTCTAATGACCCTAACAAAGTCATCTACATGAAC 6282

p2
p1

TTTGTATGAATACCATGAAATTTTTCATATAGAACTCAATCAATATTTTCTTGTACCT 6342

p2
p1

AGGGTCAAGGAAGCATGCTACAGCTAGTGCAATATTAGACACTTTCCAATATTTCTCAA 6402

p2
p1

CTTTTCACTCATTGCAACGGCCATTCTCCTAATGACAAATTTTCATGAACACACCATTG 6462

p2

p1 GTCAATCAAATCCTTTATCTCACAGAAACCTTTGTAAATAAATTTGCAGTGGAATATTG 6522

p2
p1 -----
AGTACCAGATAGGAGTTCAGTGAGATCAAAAACTTCTTCAAACACTTAAAAAGAGTTAA 6582

p2
p1 -----
TGCCATCTTCCACTCCTCGGCTTTAGGACAAATTGCATCGTACCTACAATAATTGACATT 6642

p2
p1 -----
TGATTAATTGAGAATTATAATGATGACATGTACAACAATTGAGACAAACATACCTGCGA 6702

p2
p1 -----
GGATCACTTGTTTAAAGCCTTATTAGTGCAGGCTAGTAATTAGTTAGTAATTAGTTAAGC 6762

p2
p1 -----
TAGTAATTAGTTAAGCTGCTTTCACAAACCATGTTTCCATGACTGTGTGGCACCGCTAGC 6822

p2
p1 -----CCACATGGTGTTCGAGAGCTAGCGGTGCCACA 4173
TCTCGCAACACCATGTGGGTGTAGTCCACATGGTGTTCGAGAGCTAGCGGTGCCACA 6882

p2 CACTCATGGAACATGGTTTGTGAAAGCAGCTTAATTAATTACTAGCTGACAGTGAGAG- 4232
p1 CAGTCATGGAACATGGTTTGTGAAAGCAGCTTAATAATTACTAGCTAGCTGTGAGAGA 6942
** ***** . * *****

p2 -----TGAGAGGACCACGCGAAATAAACGACAATCT-- 4263
p1 GTCGTGAGGAGTACGCGAGAGAGTGTGAGAGGACCACGCGAAAGAAACGACAATAATCT 7002
***** . :

p2 -----GGTCCAACACTCCCTCTTCATTCAAGCTAGCTA 4296
p1 CCGGTTTCGTTTTCGTTACCGATCTCGAGGTCCAACACTCGCTCTTCATTCAAGCTAGCTA 7062

p2 GCGCTGGGAGCGAACTCGCGTCGCATGGCAGAGCCTGCCTGCTGCCGAGCGGACAGGTCA 4356
p1 GCGCTGGGAGCGAACTCGCGTCGCATGGCAGAG-CTGCCTGCTGCCGAGCGGACAGGTCA 7121

p2 CGGTCCAT-----GTCCTGGCGTTG---GCGTCAGGTCCATTTCGCTTT 4398
p1 CCGTCTTCTTGACGCACGCACAGTCCATGGCGTTGGTTGGCGTCCTGTCCATTTCGCTTT 7181
***** :

p2 GCTGCCTCGCCATCACCGCCGCGCGTACTGTGCAAAGCGCGCATTTGGCTAGCTTCCCTG 4458
p1 GCTGCCTCGCCATCACCGCCGCGCGTACTGTGCAAAGCGCGCATTTGGCTAGCTTCCCTG 7241

p2 TTGCCTTGATCTGATCATAGAGGGGTCCATGTCCGTGGCAGACAGTGGGATTAATGACGG 4518
p1 TTGCCTTGATCTGATCATAGAGGGGTCCCTGTCCGTGGCAGACAGTGGGATTAATGACGG 7301
***** . *****

p2 GCCACGCCGCGCCTCCCTAATGATTTCCCCCGCCTTAACGTTGACACTGCGGACGGTGA 4578
p1 -GCCACGCCGCGCCTCCCTAATGATTTCCCCCGCCTTAACGTTGACACTGCGGACGGTGA 7360

p2 GAGGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCCATCGTCCTTCGCGAACGA 4638
p1 GAGGCGCGGCTGTGGACGTACGGCGAGATCTATCGGGGGCCCATCGTCCTTCGCGAACGA 7420

p2 TGCGTACATTTAGCGCCATATATAACACACGCCGGCACATGATTACTACGGCCTATA-- 4696
p1 TGCGTACATTTAGCGCCATATATAACACACGCCGGCACATGATTACTACGGCCTATATA 7480

p2 ----CGTCGTGAGCGGCGCGATCAATGAGTGAGCTAGTTTGTGACCAAAATGTGCGATC 4752
p1 CGCGCGTCGTGAGCGGCGCGATCAATGAGTGAGCTAGTTTGTGACCAAAATGTGCGATC 7540

p2 AAGTTGCATCAGA-----CAGTAGAGTACTGTATACTCTCTATATATATAT----- 4800
p1 AAGTTGCATCAGAGATCAGACAGTAGAGTACTGTATACTCTCTATATATATATATCTCCCT 7600

```

*****
p2 -----
p1 CCGTTTCTTTTATTGTGCTGGATAGTGTAAATTTTACACTATCCTGCGACAAATAAAA 7660

p2 -----GTTATATGTACTAGGCAGGCTAAGAAACACGATAGATACCACGCATTCGG 4850
p1 AGAAACGGAGGGAGTATGTACTAGGCAGGCTAAGAAACACGATAGATACCACGCATTCGG 7720
    * : *****

p2 TTCTCATACGAAGCTGCAGCT-----GCCCAACCAGCAGCGATGATCACGTACGCT 4902
p1 TTCTCATACGAAGCTGCAGCTGCCCAAGTGCCCAACCAGCAGCGATGATCACGTACGCT 7780
    *****

p2 CACCATCTGCGTCCTTGCGGTTTAAATTAATTACGTATGTATCCGCATCCGCATGCAGG 4962
p1 CACCATCTGCGTCCTTGCGGTTTAAATTAATTACGTATGTATCCGCATCCGCATGCAGG 7840
    *****

p2 TGGTCCCTGATCGCCAGCCACCTCCCCGGCCGAACAGACAACGAGATTAAGAACTACTGG 5022
p1 TGGTCCCTGATCGCCAGCCACCTCCCCGGCCGAACAGACAACGAGATCAAGAACTACTGG 7900
    *****

p2 AACTCGCACCTCAGCCGGCAGATCCACACGTACCGCCGGAATACACCGCCGGGCGCGAC 5082
p1 AACTCGCACCTCAGCCGGCAGATCCACACGTACCGCCGGAATACACCGCCGGGCGCTGAC 7960
    *****

p2 GATACCGCCATCGCCATCGACATGAGCAAGCTGCAGAGCGCCGACAGGCGGCGCGGCGGC 5142
p1 GACACCGCCATCGCCATCGACATGAGCAAGCTGCAGAGCGCCGACAGGCGGCGCGGCGGC 8020
    ** *****

p2 AGGACCCCGGGCGGCGCCGAAGACTAGCGCCAGCAGGACCAAGCATTTCGGACGCCGAT 5202
p1 AGGACCCCGGGCGGCGCCGAAGGCTAGCGCCAGCAGGACCAAGCAGGCGGACGCCGAT 8080
    *****

p2 CAGCCCGGCGGCGAGGCGAAAGG-----CGCGGCGGCGTCGAGCCCGGCGACAGCGAC 5256
p1 CAGCCCGGCGGCGAGGCGAAAGGCCCGGCGGCGGCGTCGAGCCCGGCGACAGCGAC 8140
    *****

p2 GCGGTGAACCCGGGCCCCGAACCAGCCCAACAGCAGCAGCGGCAGCACGGGCACGGCCGAG 5316
p1 GTGGTGAACCCGGGCCCCGAACCAGCCCAACAGCAGCAGCGGCAGCACGGGCACGGCCGAG 8200
    * *****

p2 GAGGAAGGGCCCAGCAGCGAGGACGCGAGCGGGCCGTGGGTGCTGGAGCCGATAGAGCTC 5376
p1 GAGGAGGGGGCCCAGCAGCGAGGACGCGAGCGGGCCGTGGGTGCTGGAGCCGATAGAGCTC 8260
    *****

p2 GGGGACCT---CTGGGGGAGGCCAGCAGAGATGGACGCCCTGATGCCTATCGGGCCC 5433
p1 GGGGACCTAGTCTGGGGGAGGCCAGCAGAGATGGACGCCCTGATGCCTATCGGGCCC 8320
    *****

p2 GGCGGCCACGACTCGGCTGCCCTCCAAGGGCTTGGCGCGGTCGGCGGCGAGGCCAGGTG 5493
p1 GGCGGCCACGACTCGGCTGCCCTCGAAGGGCTTGGCGCGGTCGGCTGCGAGGCCAGGTG 8380
    *****

p2 GACGACCTGTTTCGACATGGACTGGGATGGCTTCGCGGCCATCTGTGGGGCGGGCCGGAG 5553
p1 GACGACCTGTTTCGACATGGACTGGGATGGCTTCGCGGCCATCTGTGGGGCGGGCCGGAG 8440
    *****

p2 CAGGACGACCACAGCGCGCAGCTGCGGCAGGCCGCCGAGCCGATGGAAGCGGCTGCTGTT 5613
p1 CAGGACGAGCACAGCGCGCAGCTGCGGCAGGCCGCCGAGCCGCTGGAAG-----TT 8491
    *****

p2 GCCGCTGCCGCTGCCGCTGCGACGGCGGCTGCACCCCGGACGATCGCGAGCTGGAGGCG 5673
p1 GCTGCTGCTGCTGCTGCTGCGACGGCGGCCGCACCCCGGACGATCGCGAGCTGGAGGCG 8551
    ** *****

p2 TTCGAGACTTGGCTCCTGTCCGACTCGTTCTGACGGCTCCGGTCATCGGACCGATCAGAC 5733
p1 TTCGAGACTTGGCTCCTGTCCGACTCGTTCTGACGGCTCCGGTCACCGGACCGATCAGAC 8611
    *****

p2 AGACAGACAGACCAACCAAGGTG-GCCCGGCCATATGGTCGACGCCG-CTAGTAGGCGTT 5791
p1 AGACCAATAATTGGGTCACGTGTGCTCGCTCGCTCGCTCGCTCGCTCGCTGGGTCTT 8671
    ****.:.:. * . . * * * * * . : * * * . * * * * . * * *

```


p2 **G-CTCGTGTGTACAGTTTTTT-----TTTCTTT-----** 5818
 p1 **GGTTTCAGATGGCCAAATAATTGGGAAAAAATTCTACGCGGCAGGGCCGTAAAGCCACCA** 8731
 * ** . ** . ** . : * : ** : * * * :

p2 **-----TTTATTGTAGAT-----GTTGTTCTTAG-----** 5842
 p1 **CCGTGCGCTCCTGATGTCGATGCCGCGTGGAGCTCTTGCGTATCTAACGCTCCCAC** 8791
 * * . : * * . * * * * : * : * * * * .

p2 **-----CTCGTGCTGACG-----TGCTGTTTCG-----TGATGAC** 5872
 p1 **GACAATCACCTTCCAGACGGCTCGAATTACATACGACAGGATCGGCTCCGCTCTACTCC** 8851
 * * * * : * * * * . : * : * * * * * * * . *

p2 **GTA**CTG---GCATGTAGAATAG-----AGAGCAGAGATCGTATGGATTGGCGTG-- 5919
 p1 **GTTCTGTTTCGCTTCTGCTTTAGGTGCGTGCCTAGCAGATGGTGAGGCGGCGTCGCGCGGC** 8911
 ** : * * * * * : * : * . : : * * * * * . * * * . * * * * *

p2 **-----TAAAGGTAGTTCT-----CCCGCC** 5938
 p1 **CCTCCCACGGCTCGCCGGCCGCGCTACGGGGCCTGCTGCAGCAGCCCTCCTCCACGCC** 8971
 * * . * * . * * * * * * * * * * : * * * *

p2 **GGTGAAG-----TGAAGACA**CTGTCCGATT-----TTGTG----- 5969
 p1 **TGTAAAAGAGCTTTGTATTTACCTGTTTGTGTTGTGCTTTTGTGCAATGGAATAACAATG** 9031
 * * . * * . * * : * : * * * * * : * : * * * * *