Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice

Jinfeng Chen\textsuperscript{1,2,3}, Lu Lu\textsuperscript{2,3}, Jazmine Benjamin\textsuperscript{4}, Stephanie Diaz\textsuperscript{4}, C. Nathan Hancock\textsuperscript{4}, Jason E. Stajich\textsuperscript{1,3} & Susan R. Wessler\textsuperscript{2,3}

Transposable elements (TEs) shape genome evolution through periodic bursts of amplification. In this study prior knowledge of the mPing/Ping/Pong TE family is exploited to track their copy numbers and distribution in genome sequences from 3,000 accessions of domesticated Oryza sativa (rice) and the wild progenitor Oryza rufipogon. We find that mPing bursts are restricted to recent domestication and is likely due to the accumulation of two TE components, Ping16A and Ping16A\_Stow, that appear to be critical for mPing hyperactivity. Ping16A is a variant of the autonomous element with reduced activity as shown in a yeast transposition assay. Transposition of Ping16A into a Stowaway element generated Ping16A\_Stow, the only Ping locus shared by all bursting accessions, and shown here to correlate with high mPing copies. Finally, we show that sustained activity of the mPing/Ping family in domesticated rice produced the components necessary for mPing bursts, not the loss of epigenetic regulation.

\textsuperscript{1} Department of Microbiology and Plant Pathology, University of California, Riverside, CA 92521, USA. \textsuperscript{2} Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA. \textsuperscript{3} Institute for Integrative Genome Biology, University of California, Riverside, CA 92521, USA. \textsuperscript{4} Department of Biology and Geology, University of South Carolina Aiken, Aiken, SC 29801, USA. Correspondence and requests for materials should be addressed to J.E.S. (email: jason.stajich@ucr.edu) or to S.R.W. (email: susan.wessler@ucr.edu)
Eukaryotic genomes are populated with transposable elements (TEs), many attaining copy numbers of hundreds to thousands of elements by rapid amplification, called a TE burst. For a TE to successfully burst, it must be able to increase its copy number without killing its host or being silenced by host surveillance. However, because the vast majority of TE bursts have been inferred after the fact—via computational analysis of whole-genome sequence—the stealth features they require for success have remained largely undiscovered. Revealing these features requires the identification of a TE in the midst of a burst. This was accomplished for the miniature inverted-repeat TE (MITE) mPing from rice. MITEs are non-autonomous DNA transposons that are the most common TE associated with the noncoding regions of plant genes. To understand how MITEs attain high copy numbers, a computational approach was used to identify mPing, and its source of transposase, encoded by the related autonomous Pong element (Fig. 1a).

Ongoing bursts of mPing were discovered in four temperate japonica accessions: EG4, HEG4, A119, and A123, whose genomes were sequenced, and insertion sites and epigenetic landmarks were determined. These analyses uncovered two features of successful mPing bursts. First, although mPing, like other DNA TEs, prefers generic insertions, de novo insertions in exons were only 14% of expected for random insertions, thus minimizing harm to the host. Exon avoidance arises from mPing’s extended 9-bp adenine-thymine (AT)-rich insertion preference coupled with rice exon sequences that are significantly more guanine-cytosine (GC) rich than rice introns (51% vs. 37%). Second, because mPing does not share coding sequences with Pong (Fig. 1a), increases in its copy number and host recognition of its sequences does not silence Pong genes, thus allowing the continuous production of the proteins necessary to sustain the burst for decades.

The contributions of two other genetic components to the success of the bursts could not be assessed previously and are a focus of this study. First, all Pong elements in the four bursting accessions contain a single nucleotide polymorphism (SNP) at position 16 (+ 16 G/A) that distinguishes mPing and Pong sequences (Fig. 1a). The second genetic component is a single Pong locus (called Ping16A_Stow) that is the only Pong locus shared by all bursting accessions. Comparative sequence analysis of two of the four bursting accessions (A123 and A119) revealed that they were derived by self or sibling pollination about a century ago from a common ancestor that had not yet undergone Ping or mPing amplification. Significantly, this common ancestor had only a single Pong locus, which was Ping16A_Stow.

To understand the origin of these genetic components and their possible role in the burst, we analyzed the presence, sequence, and copy numbers of Ping and mPing elements in the genomes of 3000 domesticated rice accessions and 48 genomes of their wild progenitor, Oryza rufipogon. Rice has been divided into five major subgroups (indica, aus/boro, aromatic, temperate japonica, and tropical japonica) that are thought to have originated from distinct populations of the wild progenitor O. rufipogon that arose prior to domestication. Rice genomes are very stable: all analyzed genomes are composed of 12 chromosomes and rice subgroups share high sequence identity (> 98.9%). However, the genomes also exhibit extensive presence—absence variation both within (5%) and between (10%–19%) subgroups, with TEs representing more than half of this variation. In addition, significant gene flow from japonica to indica and aus has been noted previously, reflecting the more ancient origin of japonica.

Knowledge of the relationships between the major subgroups of rice and the populations of O. rufipogon have been utilized in this study to better understand the identity and origin of the components necessary for mPing bursts. Of particular interest was whether (i) mPing bursts could be detected in other accessions of wild and/or domesticated rice, (ii) the + 16 G/A Ping SNP and Ping16A_Stow could be detected in wild rice or first appeared in domesticated rice, and (iii) the presence of + 16 G/A Ping SNP and Ping16A_Stow correlated with higher mPing copy numbers.

Finally, another potential player that may be implicated in mPing bursts, Pong, a related transposase-encoding element, is a focus of this study (Fig. 1a). The Pong element is the closest relative of Pong and there are at least five identical copies found in the genome of all rice accessions analyzed to date. Pong encoded proteins catalyzed the transposition of mPing in rice cell culture and in transposition assays in Arabidopsis thaliana and yeast. However, Pong elements do not catalyze mPing transposition in planta because all Pong copies are effectively silenced and its sequences are associated with heterochromatin. Here we are able to address questions regarding the origin and stability of Pong silencing before and after domestication.

Our analysis show that mPing copy number has burst only in a few domesticated accessions and is associated with the acquisition of two variants of the transposase loci, Ping16A and Ping16A_Stow. The proportion of accessions with Ping16A has increased in domesticated rice while the original Ping (Ping16G) has been dramatically reduced. A transposition event of Ping16A into a Stowaway element created Ping16A_Stow whose presence correlates with accessions that have high mPing copies. We reject the hypothesis that a loss of global epigenetic regulation has occurred as no other TEs have amplified, indicating that these new Pong loci are the primary driver of the observed mPing burst in domesticated rice.

Results

Detection of mPing, Pong, and Pong element. Insertion sites and copy numbers for mPing, Pong, and Pong were identified from genome sequences of 3000 rice accessions using RelocaTE (see Methods). The paired-end DNA libraries had an average insert size of ~500 bp and were sequenced to a depth of 14-fold genome coverage, which allowed clear distinction between mPing, Pong, and Pong elements (Fig. 1a). Sequence analyses identified a total of 27,535 mPing, 262 Pongs, and 12,748 Pongs (Figs. 1b–d and Supplementary Data 1). Copy numbers of mPing, Pong, and Pong elements in each genome were also estimated using a read depth method (see Methods). Outputs from the RelocaTE and read depth methods were well correlated (Pearson’s correlation, R = 0.97, P < 2.2e–16 for mPing; R = 0.82, P < 2.2e–16 for Pong; R = 0.66, P < 2.2e–16 for Pong; Supplementary Figure 1) indicating that both methods to estimate approximate mPing, Pong, and Pong copy numbers in the 3000 rice accessions were robust.

Insertion sites and copy numbers for mPing, Pong, and Pong were also identified for 48 O. rufipogon accessions, but only the read depth method was used because of the limited insert size of the libraries (Supplementary Data 2). In total, 195 mPing, 25 Pongs, and 125 Pongs were estimated to be present in the 48 O. rufipogon accessions (Supplementary Data 2, Figs. 1e–g, and Supplementary Figure 2).

Copy number variation of mPing and Pong elements. None of the 3000 rice accessions analyzed in this study have more mPing elements than the 231–503 copies found in the four temperate japonica accessions (HEG4, EG4, A119, A123) in the midst of mPing bursts. Of the 3000 rice accessions, 2780 (92.7%) contain mPing, with an average of about 9 elements per accession (Fig. 1b). Temperate japonica accessions do, however, have
significantly more mPing elements (~30.5/accession) than tropical japonica (~2.6/accession), indica (~8.2/accession), or aus/boro (~3.8/accession) (one-way analysis of variance (ANOVA) with Tukey’s honest significant difference (HSD) test, adjusted P-value < 2e−16; Supplementary Table 1 and Supplementary Figure 3). All O. rufipogon accessions have mPing elements with copy numbers ranging from 1 to 11 (mean = 4.06, standard deviation = 2.39; Fig. 1e and Supplementary Figure 2). Prior studies identified four subtypes of mPing elements (mPingA-D) in domesticated rice (Supplementary Figure 4), representing four
distinct deletion derivatives of Ping. Two of the four subtypes (mPingA,B) were previously detected in O. rufipogon accessions21,22. Here we detected all four subtypes of mPing elements in O. rufipogon accessions (Supplementary Table 2) indicating that mPingA-D arose in O. rufipogon prior to domestication.

Like mPing, none of the 3000 genomes analyzed in this study have more Ping elements (7–10) than the four accessions undergoing mPing bursts7. Ping elements were detected in only 199 of 3000 accessions (6.6%) (Fig. 2 and Table 1) with most of the 199 (74.8%) having only a single copy and two accessions having 4 Pings (Fig. 2b). In contrast, Ping elements were detected in 21 of 48 (43.7%) of the O. rufipogon accessions analyzed (Table 1 and Supplementary Figure 2). These data suggest that it is likely that Ping was selected against or lost from most accessions during the hypothesized two or more domestication events from O. rufipogon populations10,14.

Origin of a Ping variant and its possible significance. Analysis of the extensive collection of rice genomes revealed that a SNP distinguishing Ping and mPing (+16 G/A), located adjacent to the 15-bp terminal inverted repeat (TIR) (Fig. 3a), may be implicated in mPing bursts. Pings having these SNPs are distinguished herein as Ping16G (identical shared sequences with mPing) and Ping16A. First, all 21 O. rufipogon accessions with Ping have only Ping16G, which has the same sequence at +16G/A as mPing (Table 1). Thus, Ping16G is the original Ping and all four mPing subtypes (mPingA-D, Supplementary Table 2) arose
Table 1 Distribution of Ping variants and Ping16A_Stow genotypes in domesticated rice and O. rufipogon

| Subgroups       | Number of accessions | Number of accessions with Pinga | Ping variants: Ping16G | Ping variants: Ping16A | Ping16A_Stow: Stowaway only | Ping16A_Stow: Stowaway with Ping |
|-----------------|----------------------|---------------------------------|------------------------|------------------------|----------------------------|-------------------------------|
| O. sativa       | 3000                 | 199 (6.6%)                      | 31                     | 154                    | 188                        | 11                            |
| -indica         | 1651                 | 20 (1.2%)                       | 8                      | gb                     | 10                         | 0                             |
| -aus/boro       | 189                  | 28 (14.8%)                      | 19                     | 0                      | 0                          | 0                             |
| -temperate japonica | 250              | 61 (24.4%)                      | 1                      | 61                     | 121                        | 8                             |
| -tropical japonica | 335              | 51 (15.2%)                      | 0                      | 51                     | 2                          | 0                             |
| -aromatic       | 65                   | 0 (0%)                          | 0                      | 0                      | 0                          | 0                             |
| -admixed        | 510                  | 39 (7.6%)                       | 3                      | 33b                    | 55                         | 3                             |
| O. rufipogon    | 48                   | 21 (43.7%)                      | 21                     | 0                      | 4                          | 0                             |
| -Or-I           | 13                   | 7 (53.8%)                       | 7                      | 0                      | 0                          | 0                             |
| -Or-II          | 23                   | 10 (43.4%)                      | 10                     | 0                      | 1                          | 0                             |
| -Or-IIIa        | 6                    | 2 (33.3%)                       | 2                      | 0                      | 3                          | 0                             |
| -Or-IIIb        | 6                    | 2 (33.3%)                       | 2                      | 0                      | 0                          | 0                             |

Ping16A_Stow is defined as a locus where Ping has inserted into the Stowaway element on chromosome 1 (2640500-2640502).

*Number of accessions with Ping16G plus Number of accessions with Ping16A is less than or equal to Number of accessions with Ping because Ping genotypes in some accessions cannot be determined from available sequences. An exception is “temperate japonica”, where one accession (IRIS_313-10564) has both Ping16G (Chr8: 2964281-2964283) and Ping16A (Chr6: 2352641-23526981).

Eight indica accessions have Ping16A that are located in regions showing evidence of introgression from japonica (seven accessions share the loci Chr3: 21965880-21965882 and one accession has the Nipponbare Ping locus Chr6: 2352641-23526981). One indica accession has Ping16A in a region with indica background. Analyses were performed with RFMix v2.03.

Thirty-one admixed accessions have Ping16A from japonica. Two admixed accessions have Ping16A that are located in regions with ambiguous origin. Analyses were performed with RFMix v2.03.

Fig. 3 Transposition frequency of mPing variants in the yeast assay. a Sequence alignment of mPing and Ping16A terminal sequence (1-40 bp). The SNP between mPing and Ping16A at position 16 (+16 G/A SNP) is indicated by the red arrow. b Transposition frequency of mPing variants with mutations at the 5’ end in the yeast assay. X axis indicates mPing variants with mutations at 14 positions in the 5’ TIR and two positions outside the TIR. For example, mPing16A represents an mPing variant having a G-to-A mutation at position 16. A variant mPingC4A was not included because the lack of qualified experiments. Y axis shows transposition frequency that was measured as mPing excision events per million cells and normalized to the control mPing. The error bars show standard deviation (s.d.) of 2-9 independent biological replicates. Letters (a and b) above the bars indicate significant differences of transposition frequency between mPing variants and control (adjusted P-value ≤ 0.05). The adjusted P-values are based on a one-way ANOVA (P-value = 2.37e-15, F-value = 12.34, DF = 16) followed by a Tukey’s honest significant difference (Tukey’s HSD) test. Source data for Fig. 3b are provided as a Source Data file.

prior to domestication by internal deletion. Second, of the 199 domesticated rice accessions with Ping, 31 have Ping16G, whereas 154 have Ping16A (Table 1). The presence of the derived Ping16A in both indica and japonica accessions was initially confusing as it suggested the unlikely scenario that this variant arose independently during or after the hypothesized two domestication events that led to these subspecies[10,14]. However, closer examination of local sequence ancestry revealed that, where a determination
could be made, all of the Ping16A loci in indica and admixed accessions originated by introgression from japonica (Table 1). Thus, Ping16A has experienced limited but significant proliferation during or after japonica domestication such that it now accounts for the majority of Ping elements present in domesticated rice accessions (Table 1).

**Reduced mobility of Ping16A in yeast assays.** The TIRs and adjacent sequences of several DNA transposons have been shown to be functionally significant with mutations of these sequences reducing transposition frequency by decreasing the binding of transposase. Because the SNP distinguishing Ping16A from Ping16G is adjacent to the 15-bp 5′ TIR (Fig. 3a), we employed a yeast assay to assess transposition rates of 14 mutations within and 2 mutations adjacent to the 5′ TIR (Fig. 3b). In this assay, Pong transposase and an enhanced Ping ORF1 (the putative binding domain) catalyzes transposition of mPing inserted in an ADE2 reporter gene, thereby allowing growth of yeast cells. The results indicate that both the mutations adjacent to the TIRs (G16A and G17T) and 12 of 14 mutations in the TIR significantly reduced mPing transposition (one-way ANOVA with Tukey’s HSD test, adjusted P-value ≤ 0.05; Fig. 3b), supporting the hypothesis that this SNP (+16 G/A) may have functional significance by reducing Ping16A’s mobility. Although Pong transposase, which was shown previously to catalyze higher transposition frequency than Ping, was used in this experiment to facilitate the yeast transposition assays, its catalytic mechanism is likely indistinguishable from Ping transposase. Furthermore, the reduced transposition of the G16A mutant (mPingG16A) was independently confirmed using Ping transposase (Supplementary Figure 5).

**A Ping locus correlates with higher mPing copy number.** The four accessions previously shown to be undergoing mPing bursts (HEG4, EG4, A119, A123) have many (7–10) Pings, and all share only a single Ping, Ping16A_STow7. This correlation suggests that acquisition of Ping16A_STow may have initiated the burst. Ping16A_STow, located on chromosome 1 (2640500–240502), is comprised of the Ping16A variant inserted in a 769-bp Stowaway element (Fig. 4a). Of interest was whether any of the 3000 accessions had Ping16A_STow and, if so, did they also have more mPings.

Among the 3000 accessions, 11 have Ping16A_STow (188 have only the Stowaway insertion at this locus) (Table 1) and these accessions have significantly more mPings (two-tailed Wilcoxon–Mann–Whitney test, \( P = 2.5e^{-08} \); Fig. 4b, Table 2, and Supplementary Table 3), providing additional correlative evidence for the involvement of Ping16A_STow in mPing bursts. Pong has been stably silenced since domestication. Pong encoded proteins catalyze transposition of mPing in yeast and A. thaliana assays and in rice cell culture. However, because Pong elements are epigenetically silenced in Nipponbare and in accessions undergoing mPing bursts (HEG4, EG4, A119, A123), there is no evidence to date that Pong has an impact on Ping or mPing copy number or distribution.

Data from this study extend previous findings and suggest that Pong was silenced in O. rufipogon and has been stably silenced in domesticated rice. Pong elements are present in the genomes of almost all of the analyzed rice accessions (99.1%, 2972/3000), and Pong copy numbers vary little within or between subgroups (Supplementary Figure 6). On average, rice accessions have four Pong elements (Fig. 1d). All O. rufipogon accessions have Pong elements (Supplementary Figure 2), except four (W1849, W1850, W2022, W2024), which appear to contain only Pong deletion derivatives (see Methods). As in domesticated rice, there is minimal Pong copy number variation among the O. rufipogon accessions examined (Supplementary Figure 2).
Six rice accessions with higher Ping copy numbers (14–25) were analyzed to determine if this resulted from Ping activation. First, because active Ping elements produce proteins that catalyze mPing transposition, we tested if the genomes of these lines contained more mPings. However, all six accessions had the same range of mPing copies as accessions with few Pongs (Supplementary Table 4). Second, because host regulatory mechanisms suppress transposition, other potentially active TEs (elements shown previously to transpose when epigenetic regulation is impaired) may have been activated in these accessions along with Ping. However, the six accessions harbored average copy numbers of nine potentially active TEs (Supplementary Table 4). Taken together these data suggest that these six accessions have accumulated silenced Ping elements since domestication. Finally, additional evidence for the stability of Ping silencing can be inferred from the observation that none of the 2801 accessions lacking Ping have a higher mPing copy number than accessions with Ping.

**Discussion**

Results of the evolutionary inventory of the members of the mPing/Ping/Pong TE family in wild and domesticated rice genomes suggest the following scenario for the origin of the mPing burst. All mPing subtypes in domesticated accessions (mPingA-D) were generated prior to domestication, probably in O. rufipogon, by internal deletion from Ping16G. Furthermore, Ping16G, but not Ping16A, was detected in 21 of 48 O. rufipogon accessions. The fact that only 31 of the 3000 extant domesticated accessions examined have Ping16G suggests that there has been a massive loss of this element in domesticated rice. In contrast, the Ping16A variant was identified in the majority of the domesticated accessions with Ping (154/199). Its absence in O. rufipogon genomes indicate that it was either very rare in wild populations or that it arose in japonica after domestication. Ping16A has experienced limited but significant proliferation in japonica and has even been introgressed into a small number of indica accessions (Table 1). Taken together, these data indicate that Ping16A has become more widely distributed in domesticated accessions, whereas Ping16G is disappearing.

Yeast assays testing the functional impact of several mutations in and adjacent to the Ping TIR demonstrate that the +16G (Ping16G) to +16A (Ping16A) polymorphism significantly reduces transposition frequency. Thus, Ping16A encoded proteins (which are identical to Ping16G encoded proteins) are more likely to catalyze the transposition of mPing (with its +16G) than Ping16A. This situation is reminiscent of other autonomous elements that harbor sequences that reduce transposition frequency[26,27]. It has been hypothesized that autonomous TEs enhance their survival by evolving self-regulating mutations that reduce both host impact and epigenetic detection and silencing[27].

The vast majority of accessions with Ping16A have only one Ping (105/154 accessions) and a moderate number of mPing elements (mean = 28). One of these accessions is the reference accession Nipponbare where the inability to detect transposition of Ping or mPing was initially attributed to Ping silencing[28]. In fact, Ping is not silenced in Nipponbare nor in any other accession analyzed to date[6]. Rather it is transcribed and catalyzes (infrequent) transposition of mPing[6,7]. We speculate that accessions with a single copy of Ping16A may be experiencing a balance, perhaps under stabilizing selection, between host survival and the maintenance of an active TE family in the genome.

The hypothesized balance between Ping16A and mPing elements and the host was perturbed in the subset of temperate japonica accessions experiencing mPing bursts[7] and it was suggested that the shared Ping16A_Stow locus may have been responsible[7]. Based on the evolutionary inventory presented in this study, it follows that Ping16A_Stow was generated in a temperate japonica accession when Ping16A transposed into a Stowaway element on chromosome 1. The Stowaway element (without the Ping insertion) was also present at this locus in O. rufipogon (Table 1). It is unlikely that this Stowaway is active as there are only four family members, each with <96% sequence identity, in the Nipponbare genome. Here we find that Ping16A_Stow is also shared by five of the six accessions with the highest mPing copy numbers among the 3000 accessions analyzed (Table 2). The sixth accession, IRIS_313_15904, has a region of introgressed indica or aromatic alleles at this location, which may have replaced the Ping16A_Stow locus in prior generations. The association of Ping16A_Stow with higher mPing copy numbers is consistent with its suggested role in triggering mPing bursts. However, the mechanism by which Ping16A_Stow may initiate the burst is unknown and warrants further investigation. Prior studies indicated that increased Ping transcripts were correlated with more mPing transpositions in accessions undergoing mPing bursts[7,28]. Our unpublished data suggest that Ping16A_Stow does not produce more transcripts compared with other Ping elements, suggesting that mechanisms other than an increased transcript level from this locus may be responsible.

In conclusion, the results of this study demonstrate that mPing bursts are likely restricted to the past century as none of the thousands of genomes analyzed have as many mPing (hundreds) and Ping (7–10) elements as the four previously characterized accessions. Further, analysis of the 3000 rice genomes and wild progenitors indicates that two variants of the autonomous Ping element, Ping16A and the subsequently evolved Ping16A_Stow locus appear to be critical for mPing hyperactivity. Other studies have shown that domestication can be associated with the loss of epigenetic regulation[29], which may lead to the activation of TEs.

**Table 2** Ping copy numbers and genotypes in rice accessions with high copy numbers of mPing

| Accession | Origin | Subgroups | mPing copy number | Ping copy number | Ping + 16G/A SNP genotypes | Ping16A_Stow |
|-----------|--------|-----------|-------------------|------------------|-------------------------------|--------------|
| HEG4      | Japan  | Temperate japonica | 503          | 7                | Ping16A                       | Yes          |
| EG4       | Japan  | Temperate japonica | 437          | 7                | Ping16A                       | Yes          |
| A123      | Japan  | Temperate japonica | 231          | 10               | Ping16A                       | Yes          |
| A119      | Japan  | Temperate japonica | 333          | 7                | Ping16A                       | Yes          |
| B160      | China  | Temperate japonica | 180          | 3                | Ping16A                       | Yes          |
| IRIS_313_15904 | South Korea | Temperate japonica | 178 | 3 | Ping16A | No |
| B235      | China  | Temperate japonica | 113          | 2                | Ping16A                       | Yes          |
| B005      | Japan  | Temperate japonica | 16          | 1                | Ping16A                       | Yes          |
| B003      | China  | Admixed       | 72           | 2                | Ping16A                       | Yes          |
| B001      | China  | Temperate japonica | 71           | 2                | Ping16A                       | Yes          |

*From Lu et al.[7]
However, our data indicate that Pong element copy number has been stably maintained from the wild ancestor through the germination of the thousands of domesticated accessions, suggesting that epigenetic regulation was unaffected. In contrast, Ping activity has been sustained during domestication, resulting in the spread and amplification of the Ping16A variant and the germination of the Ping16A_Stow locus in rice. Yet, the spread of Ping activity associated with exceptional mPing activity has been very limited in rice, likely due to its high level of self-fertilization.

**Methods**

**Dataset.** Illumina DNA sequencing reads of 3000 rice accessions were obtained from NCBI SRA project PRJEB6180. The metadata incorporating name and origin of the 3000 rice accessions was extracted from previously published Table S1A and S1B

The raw reads of 48 O. rafipogon accessions were obtained from NCBI SRA under project accession numbers listed in Supplementary Data 2. The metadata associated with the subgroup classification of these 48 O. rafipogon accessions was extracted from prior studies

Population structure and ancestral component analysis. The genotyped SNP dataset (release 1.0 3 K RG 4.8 million filtered SNP Dataset) of the 3000 rice genomes was obtained from SNP-Seek Database

A subset of 270,329 SNPs was selected by removing SNPs in approximate linkage equilibrium (round trip: 10^-6 and standard deviation: 0.005). A total of 768 accessions with major ancestral component over 99.99% were used as reference panels for 158 tropical accessions, and 158 tropical accessions were obtained from NCBI SRA project accession numbers listed in Supplementary Data 2. The metadata incorporating name and origin of the 3000 rice accessions was extracted from previously published Tables S1A and S1B

**Analysis of Ping16A_Stow.** The pre-aligned BAM files of 3000 rice genomes (http://s3.amazonaws.com/3kricegenome/Nipponbare/"Accession_Name"_realigned.bam) were analyzed to determine if a Stowaway element was present at the Ping16A_Stow locus. A pseudogene was built using a Ping element and its flanking sequences, which are 1–305 bp of the Stowaway element upstream of Ping and 306–770 bp of the Stowaway element downstream of Ping. The sequencing reads of these 199 rice genomes were aligned to the pseudogene using BWA with default parameters (mem -k 19 -w 100 -T 30) followed by analysis of the BAM files to identify junctions, including both the Stowaway and its flanking sequence. All of these 199 accessions were confirmed to have the Stowaway element at positionChr: 2640500–2640502. A similar approach that identified the Stowaway insertion was used to identifyPing insertions in the Stowaway element at the Ping16A_Stow locus. A pseudogene was built using a Ping element and its flanking sequences, which are 1–305 bp of the Stowaway element upstream of Ping and 306–770 bp of the Stowaway element downstream of Ping. The sequencing reads of these 199 rice genomes were aligned to the pseudogene using BWA with default parameters (mem -k 19 -w 100 -T 30). Analysis of junction reads covering both the Stowaway and its flanking sequence. All of these 199 accessions were confirmed to have the Stowaway element at positionChr: 2640500–2640502.
2582230–25802232 (a mPing locus in Ausboro), were assigned through manual inspection because these loci were either reference Ping (53 accessions with Nipponbare mPing, 36 accessions with Ausboro mPing, and 36 accessions with Kitaake mPing) or orphans identified by RelocaTE2 (11 accessions with Ping16A_Stow and 4 accessions with the Ausboro Ping locus). For O. rufipogon, all reads aligning to mPing, Ping, and Pong were pooled to analyze the base composition at the +16 G/A SNP because mPing, Ping, and Pong insertions could not be efficiently sorted. An O. rufipogon genome was categorized as a genome having Ping162 or Ping16A based on whether they had two or more reads supporting G or A. Accessions that have two or more reads supporting both G and A were further analyzed to clarify whether the Ping16A is present in these genomes. For example, accession W1230 had both G (288 reads) and A (23 reads) at the +16 SNP. These A-supporting reads were extracted from W1230 sequences and aligned to pseudogenomes that have W1230 mPing or Ping inserted in MSU7. All of these A-supporting reads were uniquely aligned to mPing locus Chr3: 25526483-25526485 that contains a 430-bp mPingC element successfully assembled from locus-specific paired-end reads, suggesting these A-supporting reads were from mPing not from Ping.

**Assembly and classification of mPing sequences.** A locus-specific assembly was performed to recover full-length mPing sequences from rice sequences. The sequencing reads matching mPing were obtained using RelocaTE2, assembled using velvet v1.2.9 (MAXKMERLENGTH = 31 -ins_length 500 -exp_cov auto -scaffolding yes)44. The flanking non-mPing sequences were removed from the assembled sequences. Any mPing candidate loci containing sequence gaps were removed from the analysis. The remaining full-length mPing sequences were compared using ClustalW v2.2.26 to build the mPing catalog with python package NetworkX (https://networks.github.io). Each node in the graph is an mPing sequence and each edge is a connection, which requires two mPing sequences are properly aligned (number of gaps or mismatches ≤ 4). The mPing sequences in each subgraph represent a subtype of mPing. Representative sequences were extracted from each mPing subtype and aligned with four canonical defined mPing subtypes (mPingA, mPingB, mPingC, and mPingD) from the prior study using MUSCLE v3.8.425 with default parameters (–maxiters 16). The multiple sequence alignment in MSA format was converted into VCF format using msavcfjar tool (https://github.com/lindenb/jvarkit) to identify polymorphic sites. The assembled mPing sequences were classified into subtypes based on their breakpoints and point mutations compared with the four canonical mPing subtypes.

The reads of O. rufipogon accessions were aligned to four canonical defined mPing subtypes (mPingA, mPingB, mPingC, and mPingD) using BWA with default parameters (mem -k 10 -w 100 -T 30). Alignments with ≤ 2 mismatches were manually inspected using IGV v2.3.0 to determine if the reads cover breakpoint of each mPing subtype in each accession. An accession with two or more reads covering the breakpoint of an mPing subtype was identified as an accession containing this mPing subtype.

**Phylogenetic analysis.** The 270,329 SNPs used for ADMIXTURE analysis were used to genotype HEG4, EG4, AI19, and AI23 using GATK UnifiedGenotyper v3.4-46. The phylogenetic tree of rice accessions was built using a neighbor-joining method implemented in FastTree v2.1.10 (http://ab荒l.sourceforge.net). The sequencing reads for the 48 O. rufipogon accessions were obtained to analyze a SNP dataset. Briefly, paired-end reads were aligned to MSU7 using SpeedSeq v 0.1.0 (align -t 24 -R "@RG\tID=\tSM=\tTL\tLibrary\"). The resulting BAM files were analyzed using GATK UnifiedGenotyper to perform SNP calling. Filtering parameters (QD < 2.0, MQ < 20.0, MQ0 > 20.0, FS < 0.05, HaplotypeScore < 13.0, MQRankSum < -0.1, ReadPosRankSum < -8.0, MQ6 > 4 & (MQ0/(1.0×DP)) > 0.1), QUAL < 30.0, DP < 6, DP > 5000, HSR > 5) were used to retain high-quality SNPs using GATK VariantFilter. Only homozygous SNPs that did not overlap the repetitive sequences were used in the phylogenetic analysis. These high-quality SNPs were extracted and converted into PHYLPFP format multiple sequence alignment for phylogenetic analysis with RAxML v8.8.248 under a GTRGAMMA model (–m GTRGAMMA). Bootstrap was performed using 100 iterations (-f a – b 100). O. glaberrima, Oryza glumaepatula, and O. meridionalis were treated as outgroups. Graphical representations of the phylogenetic trees were generated in R using “APE” libraries.

**Yeast transposition assay.** mPing was amplified with Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific) using the control mPing primers (mPing F and mPing R) or mutation containing primers (i.e., mPing F and mPing16A R; Supplementary Table 5). The primary PCR products were then amplified with ADDE2 TSD F and ADDE2 TSD R primers (Supplementary Table 5) to add ADDE2 homologous sequences. Purified PCR products were co-transformed into Saccharomyces cerevisiae strain IMI1725 with HpaI digested pWLP98a plasmid using the lithium acetate/polyethylene glycol method51. Plasmids were isolated from individual yeast clones using the Zymo Yeast Plasmid Miniprep kit (Zymo Research) and transformed into Escherichia coli for plasmid purification and sequence validation.

Sequence verified plasmids were transformed into S. cerevisiae strain BR101 containing previously described pAG413 GAL ORF1 Shuttle1 NLS and pAG415 GAL Pong Tpase L384A, L386A plasmids25. The transposition rate was measured as described in the prior study14. Briefly, 3 ml cultures were grown in CSM-His-Leu-Ura (dextrose) for 24 h at 30 °C, and 100 µl was plated onto 100 mm CSM-His-Leu-Ura-Ade (galactose) plates. The total number of yeast cells was calculated by plating a 10^5 dilution of the cultures onto YPD plates. The numbers of colonies on the galactose plates were determined after 10 days of incubation at 30 °C. The transposition rate was determined by dividing the galactose colony count by the total number of cells plated.

**Statistical analysis.** Sample sizes, statistical tests, and P-values are indicated in figures or figure legends. Linear regression, two-tailed Pearson’s correlation, two-tailed Wilcoxon–Mann–Whitney, one-way ANOVA and Tukey’s HSD test were performed with “lm”, “cor.test”, “wilcox.test”, “aov”, and “TukeyHSD” functions in R. One-sample t-test was performed with ‘ttest_1sample’ function in Python module ‘scistats’.

**Code availability.** RelocaTE2 and other code used in this study are available at https://github.com/stajichlab/Dynamic_rice_publications or https://doi.org/10.5281/zenodo.1492794.
18. Hancock, C. N., Zhang, F. & Wessler, S. R. Transposition of the Tourist-MITE mPing in yeast: an assay that retains key features of catalysis by the class 2 PHEH/Harbinger superfamily. Mol. DNA 1, 5 (2010).

19. Chen, J., Wrightsman, T. R., Wessler, S. R. & Stajich, J. E. Relocate2: a high resolution transposable element insertion site mapping tool for population resequencing. PeerJ 5, e2942 (2017).

20. Li, Z. et al. The 3,000 rice genomes project. Gigascience 3, 7 (2014).

21. Karki, S. et al. Analysis of distribution and proliferation of telomere repeats in a wild rice (Oryza rufipogon Griff.). Breed. Sci. 59, 297–307 (2009).

22. Hu, H., Mu, J., Zhang, H., Tao, Y. & Han, B. Differentiation of a miniature inverted transposable element (MITE) system in Asian rice cultivars and its inference for a diphylectic origin of two subspecies of Asian cultivated rice. J. Integr. Plant Biol. 48, 260–267 (2006).

23. Zhou, M., Bhasin, A. & Reznikoff, W. S. Molecular genetic analysis of the npp family transposons in rice. Gene 215, 225–232 (1998).

24. Feschotte, C., Osterlund, M. T., Peeler, R. & Wessler, S. R. DNA-binding specificity of rice mariner-like transposases and interactions with Stowaway MITES. Nucleic Acids Res. 33, 2153–2165 (2005).

25. Payero, L., Outten, G., Burckhalter, C. E. & Hancock, C. N. Alteration of the MITE in a wild rice (Oryza rufipogon Griff.). Breed. Sci. 59, 297–307 (2009).

26. Claeys Bouuaert, C., Lipkow, K., Andrews, S. S., Liu, D. & Chalmers, R. The Monograft: a tool for muscle integration. Nucleic Acids Res. 42, 8079–8089 (2014).

27. Yang, G., Nagel, D., Feschotte, C., Hancock, C. N. & Wessler, S. R. Tuned autoregulation of a eukaryotic DNA transposon. Nature 440, 207–211 (2006).

28. Alexandrov, N. et al. SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res. 44, D104–D110 (2016).

29. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a flexible and accurate model for inferring local ancestry in admixed individuals. Genet. Epidemiol. 36, 1312–1325 (2012).

30. Li, L. F., Li, Y. L., Jia, Y., Caicedo, A. L. & Olsen, K. M. Signatures of natural selection on rice domestication. Mol. Ecol. 24, 3734–3747 (2015).

31. Alexandrov, N. et al. SNP-Seek database of SNPs derived from 3000 rice genomes. Nucleic Acids Res. 44, D104–D110 (2016).

32. Purcell, S. et al. PLINK: a whole-genome association analysis toolset. Am. J. Hum. Genet. 81, 559–575 (2007).

33. Maples, B. K., Gravel, S., Penny, D. K. & Andolfatto, P. D. DMix: a computational tool for inferring local ancestry in admixed individuals. Mol. Ecol. 24, 3519–3535 (2015).

34. Okonechnikov, K., Conesa, A. & Garcia-Alcalde, F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294 (2016).

35. Van der Ploeg, S. et al. The genome analysis toolkit (GATK): scalable and accurate genotype calling. Genome Res. 20, 625–631 (2010).

36. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

37. Li, H. et al. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

38. Hu, X. et al. PilRS: Profile-based Illumina pair-end reads simulator. Bioinformatics 28, 1533–1535 (2012).

39. Chiang, C. et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat. Methods 12, 966–968 (2015).

40. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. Bioinformatics 31, 2032–2034 (2015).

41. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics 30, 2503–2505 (2014).

42. Robinson, J. T. et al. Integrative genomic viewers. Nat. Biotechnol. 29, 24–26 (2011).

43. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829 (2008).

44. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829 (2008).

45. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004).

46. McMullen, M. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

47. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650 (2009).

48. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).

49. Paradis, E., Claude, J. & Strimmer, K. In APE: analyses of phylogenetics and evolution in R language. Bioinformatics 24, 289–290 (2008).

50. Gilbert, D. M. et al. Precise repair of mPing excision sites is facilitated by target site duplication derived microhomology. Mol. DNA 6, 15 (2015).

51. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyyethylene glycol method. Methods Enzymol. 350, 87–96 (2002).

Acknowledgements
We thank Drs. Hongru Wang and Shujun Ou for discussions of data analysis and Dr. Shaohua Fan and Julia Adams for valuable comments on the manuscript. This work was supported by National Science Foundation grants (IOS-1027542 to S.R.W. and J.E.S. and MCB-1651666 to C.N.H.). Data analyses were performed on the UC Riverside High-Performance Computational Cluster supported by National Science Foundation grant DBI-1429826 and National Institutes of Health grant S10-OD016290.

Author contributions
J.C., J.E.S. and S.R.W. conceived the study. J.C. and L.I.L. analyzed the sequence data. J.B., S.D. and C.N.H. performed the yeast experiment and analyzed the data. J.C., C.N.H., J.E.S. and S.R.W. wrote the paper.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-08451-3.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Journal peer review information: Nature Communications thanks Ronald Chalmers, Terence Brown and Francois Sabot for their contribution to the peer review of this work. Peer reviewer reports are available

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.