LETTERS

Unexpected consequences of a sudden and massive transposon amplification on rice gene expression

Ken Naito[1,2], Feng Zhang[1]†, Takuji Tsukiyama[2], Hiroki Saito[2], C. Nathan Hancock[1], Aaron O. Richardson[1], Yutaka Okumoto[2], Takatoshi Tanisaka[2] & Susan R. Wessler[1]

High-copy-number transposable elements comprise the majority of eukaryotic genomes where they are major contributors to gene and genome evolution[1]. However, it remains unclear how a host genome can survive a rapid burst of hundreds or thousands of insertions because such bursts are exceedingly rare in nature and therefore difficult to observe in real time[2]. In a previous study we reported that in a few rice strains the DNA transposon *mPing* was increasing its copy number by ~40 per plant per generation[3]. Here we exploit the completely sequenced rice genome to determine 1,664 insertion sites using high-throughput sequencing of 24 individual rice plants and assess the impact of insertion on the expression of 710 genes by comparative microarray analysis. We find that the vast majority of transposable element insertions either upregulate or have no detectable effect on gene transcription. This modest impact reflects a surprising avoidance of exon insertions by *mPing* and a preference for insertion into 5′ flanking sequences of genes. Furthermore, we document the generation of new regulatory networks by a subset of *mPing* insertions that render adjacent genes stress inducible. As such, this study provides evidence for models first proposed previously[4–6] for the involvement of transposable elements and other repetitive sequences in genome restructuring and gene regulation.

In a previous study we discovered that rice strain EG4 has over 1,000 *mPing* elements whereas the sequenced rice genome (Nipponbare) has only 50[3,7,8]. In addition, with almost ~40 new inserts per plant, even small populations of EG4 contain thousands of new insertions. Because the transposable element burst is occurring in a species with a complete genome sequence, virtually all insertion sites can be determined. To this end, we used 454 sequencing[9] to identify *mPing* insertion sites in a small population (24 plants) of selfed progeny from a single parent. Sequences flanking *mPing* insertions were amplified by vectorette PCR[10] using primers with barcodes to identify the plant of origin of new insertions (Supplementary Fig. 1). In all, 928 insertions were detected in more than two EG4 siblings (shared insertions) and 736 insertions were individual-specific (unshared; from 20–65 per plant). All 1,664 insertions were characterized using the rice annotation project database[11] (http://rapdb.dna.affrc.go.jp/) (Supplementary Tables 1 and 2). A control dataset of 1,664 randomly selected genomic sites was generated 1,000 times for comparison (see Methods).

The vast majority of insertions, whether shared or unshared, were into single copy sequences of the genome (1,521 of 1,664 = 91.4% compared with control, 898 of 1,664 = 53.9%). Superimposition of all insertion sites on the physical map of the 12 rice chromosomes shows that both shared and unshared sites are enriched in euchromatic regions of high gene density and largely absent from heterochromatic regions (Fig. 1a and Supplementary Fig. 2). A preference for insertion into genic regions was noted previously for other DNA transposons[12,13]. However, the availability of over 1,500 insertions that are near or in genes, of which almost half are *de novo*, afforded a unique opportunity to investigate whether a successful element like *mPing* has evolved target preferences that serve to mitigate the impact of a massive and rapid increase in copy number.

A closer look at the insertion sites in rice genes indicates that both shared and unshared insertions of *mPing* are significantly underrepresented in coding exons (9 of 928 = 1.0% of shared and 6 of 736 = 0.8% of unshared compared with 103.4 of 1,664 = 6.3% of the control; $P \ll 0.01$) (Fig. 1b and Supplementary Tables 3 and 4). A similar distribution for most (old) transposable elements in eukaryotic genes has been attributed to selection acting to filter detrimental exon insertions[3], but lack of unshared inserts in exons suggests that *mPing* avoids exons. One could argue however that even *de novo* exon insertions may have been filtered by selection (for example, gametophyte- or dominant-lethality). This argument, however, is not supported by previous studies where 30–35% of *de novo* insertions of *Tos17*[14,15] and *Ac/Ds*[16] in rice were into exons (Supplementary Table 5).

In contrast with insertions into coding regions, both shared and unshared insertions are significantly overrepresented in the vicinity of genes compared to the control data set (Fig. 1c, d and Supplementary Tables 3 and 4). The most dramatic deviation is within −1 kilobase (kb) of the transcription start site (shared, 192 of 928 = 20.7%, and unshared, 114 of 736 = 15.5%, compared with control, 92.5 of 1,664 = 5.6%).

No single mechanism can account for the biased distribution of *mPing* inserts. For example, *mPing* may avoid exons because its 9 base pair (bp) target site preference is (A+T)-rich (GC content = 31%, Supplementary Fig. 3) whereas rice exons are, on average, (G+C)-rich (~55%)[17]. However, *mPing* does not avoid the (G+C)-rich 5′ untranslated region and is enriched just upstream of the transcription start site. An understanding of the mechanisms underlying these preferences is beyond the scope of this study as they may be influenced by other factors such as chromatin structure[18,19], which, so far, has not been thoroughly characterized in rice.

Taken together the analysis of 1,664 *mPing* insertion sites indicates that the spectrum of shared and unshared insertions is virtually indistinguishable. Furthermore, the position of insertions suggests that the rapid amplification of successful transposable elements like *mPing* may have a more modest impact on the host than previously thought because of highly evolved targeting mechanisms that minimize deleterious effects on host gene expression.

To test this hypothesis, we used comparative microarray analysis to assess the impact of *mPing* amplification on the transcription of rice genes. We focused on the expression of 710 genes harbouring *mPing* inserts (Fig. 2a; within 5 kb upstream and downstream) that are present in all EG4 plants (shared inserts) (928 total shared

[1]Department of Plant Biology, University of Georgia, Athens, Georgia 30602, USA. [2]Division of Agronomy and Horticulture Science, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Sakyo-ku, Kyoto 606-8502, Japan. †Present address: Department of Genetics, Cell Biology and Development, University of Minnesota, Minnesota 55455, USA.

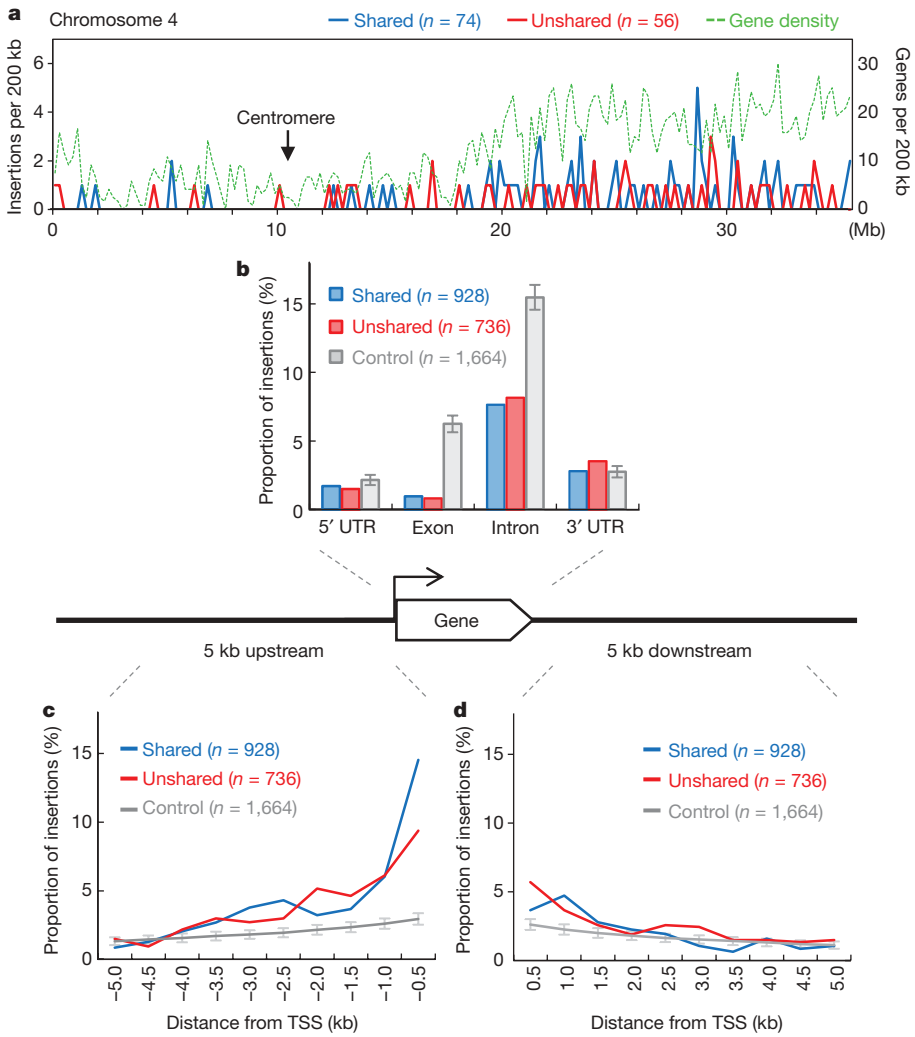

Figure 1 | Distribution of *mPing* insertions in strain EG4. **a**, Distribution along chromosome 4. The *x* axis indicates the distance along the chromosome and the *y* axes indicate the number of *mPing* insertions (left) and genes (right) per 200 kb. **b**–**d**, Proportion of all *mPing* insertions within the coding region of annotated genes (**b**) and upstream (**c**) or downstream of annotated genes (**d**). TSS, transcription start site; TTS, transcription termination site; UTR, untranslated region. Mean ± s.d., $n = 1{,}000$ (for 'control').

inserts – 218 intergenic shared inserts = 710; Supplementary Table 3). Unshared insertions were excluded from this analysis because most do not exist in the EG4 plants used for RNA isolation. To perform a comparative analysis of the 710 EG4 alleles (with *mPing*) to the Nipponbare alleles (without *mPing*, called NB alleles hereafter) RNA from seedlings of both strains grown under normal conditions was hybridized to 44K microarray chips containing 31,439 rice genes (see Methods)[20]. Aside from the difference in *mPing* copy number, NB and EG4 are genetically very similar as EG4 is one of the strains of the cultivar Gimbozu, which is a great-grandparent of NB (see Supplementary Fig. 4 where the extreme *mPing* polymorphism is compared to the monomorphic pattern of the *Dasheng* retrotransposon[21]). As another measure of genetic similarity, only 7.0% of simple sequence repeat sites are polymorphic between EG4 and NB[22].

Comparison of the expression of 31,439 EG4 and NB alleles showed that 82% are indistinguishable under normal growth conditions (Supplementary Table 6, 25,809 of 31,439). Of the remaining 18%, approximately half of the EG4 alleles are upregulated relative to the NB alleles whereas half are downregulated. The distribution of this 'background' of expression differences between EG4 and NB is reported as the control (see Fig. 2b, c).

Comparison of the distribution of gene expression differences between the 710 EG4 (*mPing*-containing) and NB (no *mPing*) alleles and controls reveals significant differences (Fig. 2b, c and Supplementary Tables 6 and 7). First, among the EG4 alleles that show a difference in expression (156 of 710) there are significantly more upregulated EG4 alleles (111 of 156), largely due to alleles harbouring *mPing* within 1–5 kb upstream of the transcription start site ($P < 0.01$). Although there were proportionally more upregulated EG4 alleles with *mPing* in promoters (~−1 kb), introns and downstream of the transcription termination site (~+5 kb), these were not significantly different from the control. Second, as expected, a significantly higher proportion of the 51 EG4 alleles with exon insertions were downregulated compared with the control ($P < 0.01$). Finally, microarray data were independently verified for 17 alleles using quantitative real-time PCR (qPCR) (Supplementary Figs 5 and 6).

The microarray data support the hypothesis that under normal growth conditions *mPing* insertions have a modest impact on host transcription, as the vast majority of EG4 alleles either have no impact or preferentially enhance transcription.

The enhancement of transcription by *mPing* motivated us to survey *mPing* sequence for regulatory motifs. Scanning *mPing* sequences with the plant *cis*-element database[23] (http://www.dna.affrc.go.jp/PLACE/index.html) identified 96 putative regulatory motifs; about one-third of which were stress responsive (Supplementary Table 8).

The involvement of a variety of stresses has been correlated with the activation of transposable elements ever since their discovery in a maize strain undergoing the breakage-fusion-bridge cycle[24]. The *mPing* element is a deletion derivative of *Ping*, an autonomous

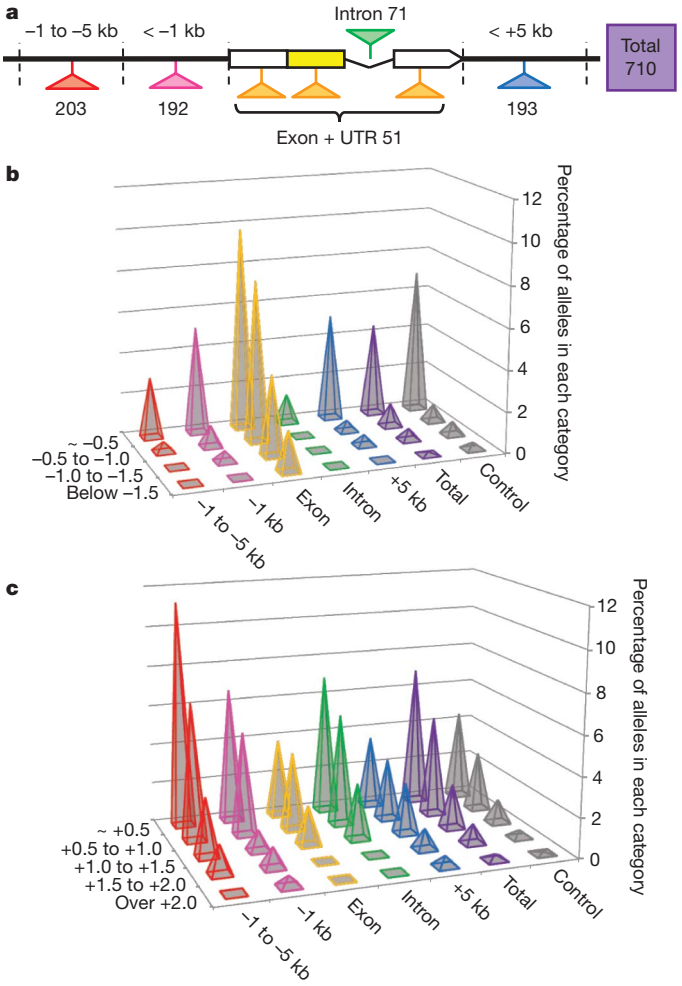

Figure 2 | Impact of *mPing*-containing alleles on host transcription. **a**, Schematic of *mPing*-harbouring alleles from the shared insertions. Triangles represent *mPing* insertions that are colour-coded and include the number of insertions in the region shown. Yellow and white boxes represent coding and non-coding exons, respectively. **b**, **c**, Distribution of the degree of expression differences of *mPing*-harbouring alleles (in strain EG4) against Nipponbare alleles from microarray analyses among downregulated alleles (**b**) and upregulated alleles (**c**). The *x* axis is the degree of expression differences ($\log_{10}$ value of fold change). The *y* axis is the proportion of all *mPing* alleles in the corresponding region.

element that codes for two proteins that are both required for transposition of *Ping* and *mPing* (Fig. 3a)[25]. As such, *mPing* contains all of the sequences upstream of the basal promoter of ORF1. Whether these sequences are sufficient to confer stress-inducibility was tested by generating transgenic *Arabidopsis* plants containing this region fused to the GUS coding region. As shown in Fig. 3b, seedlings exposed to cold showed more intense GUS staining, which correlated with higher levels of GUS RNA in cold-stressed seedlings (Fig. 3c).

To investigate whether the presence of *mPing* conferred stress inducibility on nearby rice genes we began by performing qPCR on the products of 10 genes whose transcription was unaffected (in the microarray analysis) by the presence of *mPing* within 55 bp of the transcription start site. In addition, we exploited the availability of two landraces (A123, A157) that are very closely related to EG4 and have over 1,000 *mPing* inserts, but the spectrum of *mPing* insertions is dramatically different (Supplementary Fig. 4)[3]. As such, these strains allow us to compare, for example, the expression of a NB-type allele (no *mPing*) to an EG4-type allele (with *mPing*) where both are in strains with ~1,000 *mPing* insertions.

For each gene, we tested RNA isolated from the seedlings of four strains (NB, EG4, A123, A157) under conditions of cold, salt and dehydration stress (see Methods). For each allele there are landraces

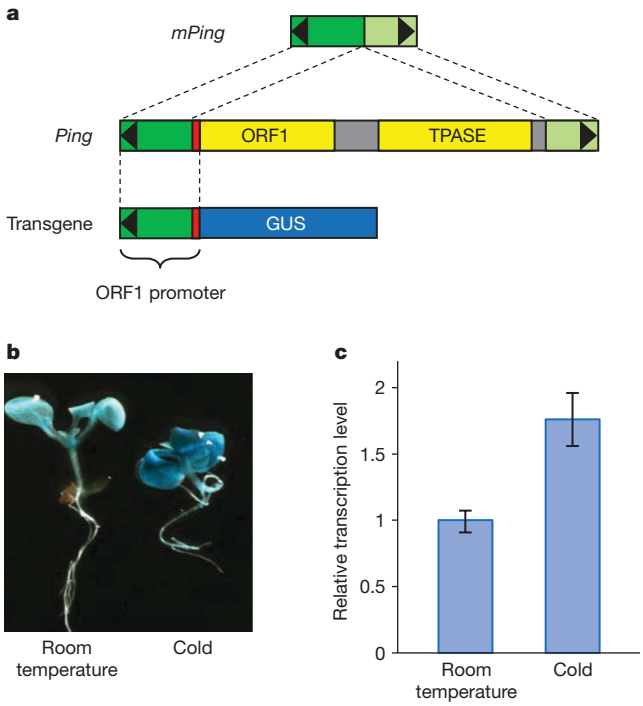

Figure 3 | Cold inducibility of a transgene in *Arabidopsis* containing the *Ping* ORF1 promoter fused to GUS. **a**, Derivation of *mPing* and a transgene promoter from the *Ping* element. Black triangles represent the terminal inverted repeat (TIR) and the red region represents the basal promoter of *Ping* ORF1. **b**, GUS staining of *Arabidopsis* transformants at room temperature and after cold treatment. **c**, Relative transcription levels of the GUS transgene in *Arabidopsis* seedlings. Mean ± s.d., $n = 3$.

that have an NB-type allele. Of the ten loci tested, seven EG4-type alleles are inducible by cold and salt, but not by dehydration (Fig. 4 and Supplementary Fig. 7). Furthermore, sequencing of the NB and EG4 alleles of the three genes shown in Fig. 4 demonstrated that they differed only by the *mPing* insertion, except for a single nucleotide difference in one of the alleles (at −364 nucleotides in Os02g0582900).

To address the question of whether *mPing* is providing the promoter for nearby genes or whether it contains sequences that activate transcription from a distance, we analysed the RNA levels (at room temperature and after cold exposure) of genes where *mPing* was in introns (Supplementary Figs 8 and 9), ~2.5 kb upstream of the transcription start site (Supplementary Fig. 10) or downstream of the transcription termination site (Supplementary Fig. 11). We find that sequences within *mPing* act as an enhancer as the majority of EG4 alleles are cold-inducible regardless of the location of the insertion, while five of five control alleles (no *mPing* in any strain) showed virtually no difference in stress response (Supplementary Fig. 12). These data suggest that sequences within *mPing* provide new binding sites for transcription factors or other regulatory proteins.

In this study we have caught a transposon in the act of rapid amplification. Although conventional wisdom is that most transposable element insertions are neither beneficial nor neutral to the host, the results of our study demonstrate that populations can survive rapid and massive increases in transposable element copy number, even of transposable elements that prefer to insert into genic regions, because (successful) transposable elements have evolved target preferences that are largely neutral. Furthermore, we demonstrate that a large subset of the new alleles may actually benefit the host by creating potentially useful allelic variants and novel, stress-inducible regulatory networks. At least two stresses have been identified in this study (cold and salt), and one (drought) has been ruled out. It is possible that other subsets of *mPing* alleles respond to additional growth conditions. Taken together, *mPing* amplification can potentially create populations of rice (larger than the 24 rice plants in this study) with

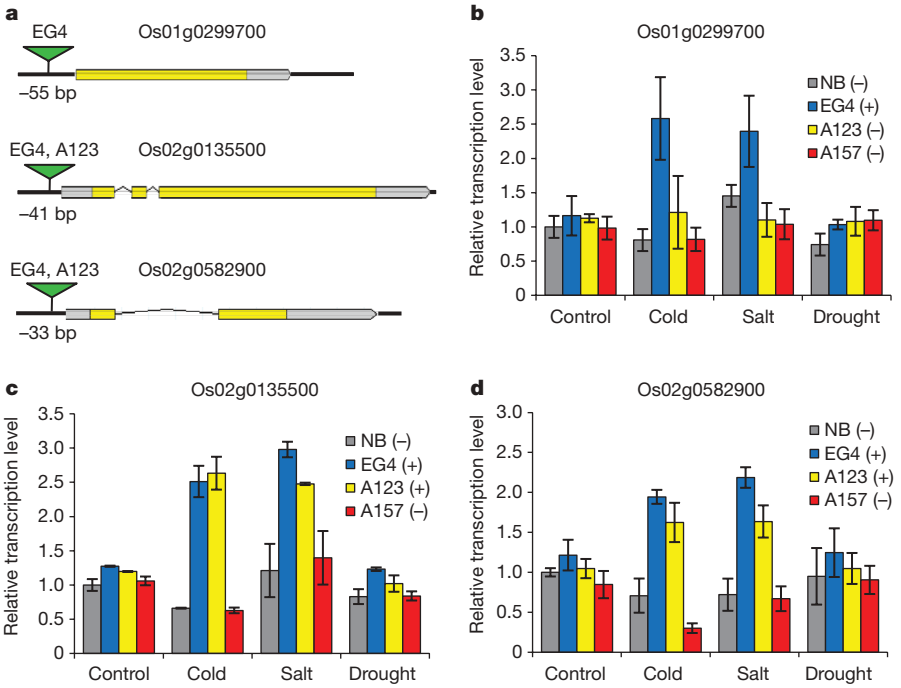

Figure 4 | Stress-induction of transcription of alleles harbouring *mPing*. **a**, Schematic of the gene structure of alleles with *mPing* in EG4 (top) and in strain A123. **b–d**, Relative transcription levels of selected alleles under stress conditions. The *y* axis is the amount of transcript in each sample relative to the RNA in NB, which is given a value of 1. RNA from two landraces (A123 and A157) where *mPing* has amplified to over 1,000 copies was also quantified. A '+' indicates that this strain shares the *mPing* allele with EG4 and a '–' indicates that the strain has the NB allele (no *mPing*). Mean ± s.d., $n = 3$.

tens, perhaps hundreds of thousands of new alleles. Of note is that many of these alleles will be dominant and, as such, one allele, or a subset of alleles, could quickly alter one or more quantitative trait. This is consistent with McClintock's view of transposable elements as agents of evolutionary change[4,26]. For rice and other selfing plants, transposable element bursts may be one of the critical solutions to rapidly generate genetic diversity in the face of an ever-changing environment. Furthermore, because evidence for the rapid and massive amplification of miniature inverted-repeat transposable elements has been found in virtually all sequenced eukaryotic genomes[27] (and even some prokaryotes[28]), features of transposable element amplification documented for *mPing* are likely to be widespread in nature.

METHODS SUMMARY

Genomic DNA extracted from 24 EG4 plants was digested with BfaI and ligated to vectorette adapters[10]. Sequences flanking *mPing* insertions were amplified by nested PCR with *mPing*-specific primers and adapter primers with distinct four-base barcodes, pooled and subjected to pyrosequencing (454 Life Sciences). All sequences were assigned into distinct groups based on the DNA barcode. Shared and unshared *mPing* insertions were identified between barcoded groups by comparing non-redundant insertion sites. For the rice 44K microarray (Agilent Technologies) analysis, RNA was extracted from 10 seedlings each (seven-day-old) of EG4 and Nipponbare. EG4 versus NB alleles selected as up- or downregulated were determined according to the manufacturer's protocol. The *Ping* ORF1 promoter was cloned into the pDONR vector (Invitrogen), transferred to pMDC162 (ref. 29) by recombination, and transformed into *Arabidopsis thaliana* ecotype Columbia by the floral dip method[30]. Whether EG4 landraces (A123 and A157) contain EG4 alleles (with *mPing*) or NB alleles (no *mPing*) was determined by PCR using primers flanking *mPing* inserts. For analysis of stress inducibility, 10-day-old plants were exposed to 4 °C for 2 h (cold), 150 mM NaCl for 24 h (salt), or deprived of water (dehydration) for 5 days. Microarray data can be found at GEO accession number GSE15021.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 July; accepted 2 September 2009.

1. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9,** 397–405 (2008).
2. Feschotte, C., Jiang, N. & Wessler, S. R. Plant transposable elements: where genetics meet genomics. *Nature Rev. Genet.* **3,** 329–341 (2002).
3. Naito, K. *et al.* Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl Acad. Sci. USA* **103,** 17620–17625 (2006).
4. McClintock, B. The significance of responses of the genome to challenge. *Science* **226,** 792–801 (1984).
5. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* **165,** 349–357 (1969).
6. Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* **46,** 111–138 (1971).
7. Jiang, N. *et al.* Active DNA transposon family in rice. *Nature* **421,** 163–167 (2003).
8. Kikuchi, K., Terauchi, K., Wada, M. & Hirano, H. Y. The plant MITE *mPing* is mobilized in anther culture. *Nature* **421,** 167–170 (2003).
9. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437,** 376–380 (2005).
10. Arnold, C. & Hodgson, I. J. Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* **1,** 39–42 (1991).
11. Itoh, T. *et al.* Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.* **17,** 175–183 (2007).
12. Cresse, A. D., Hulbert, S. H., Brown, W. E., Lucas, J. R. & Bennetzen, J. L. *Mu1*-related transposable elements of maize preferentially insert into low copy number DNA. *Genetics* **140,** 315–324 (1995).
13. Liao, G. C., Rehm, E. J. & Rubin, G. M. Insertion site preferences of the *P* transposable element in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **97,** 3347–3351 (2000).
14. Miyao, A. *et al.* Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell* **15,** 1771–1780 (2003).
15. Piffanelli, P. *et al.* Large-scale characterization of *Tos17* insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* **65,** 587–601 (2007).
16. Kolesnik, T. *et al.* Establishing an efficient *Ac/Ds* tagging system in rice: Large-scale analysis of *Ds* flanking sequences. *Plant J.* **37,** 301–314 (2004).
17. Yu, J. *et al.* A draft sequence of the rice genome (*Oryza sativa L.* ssp. *indica*). *Science* **296,** 79–92 (2002).
18. Mavrich, T. N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453,** 358–362 (2008).
19. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458,** 362–366 (2009).
20. Shimono, M. *et al.* WRKY45 plays a crucial role in benzothiadiazole-inducible blast resistance. *Plant Cell* **19,** 2064–2076 (2007).
21. Jiang, N. *et al. Dasheng*: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics* **161,** 1293–1305 (2002).
22. Monden, Y. *et al.* High potential of a transposon *mPing* as a marker system in *japonica* × *japonica* cross in rice. *DNA Res.* **16,** 131–140 (2009).

23. Higo, K., Ugawa, Y., Iwamoto, M. & Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acids Res.* **27,** 297–300 (1999).
24. McClintock, B. The relation of homozygous deficiencies to mutations and allelic series in maize. *Genetics* **29,** 478–502 (1944).
25. Yang, G., Zhang, F., Hancock, C. N. & Wessler, S. R. Transposition of a rice miniature inverted-repeat transposable element *mPing* in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **104,** 10962–10967 (2007).
26. Capy, P., Gasperi, G., Biemont, C. & Bazin, C. Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85,** 101–106 (2000).
27. Feschotte, C., Zhang, X. & Wessler, S. R. in *Mobile DNA II* Ch. 50 (eds Craig, N. L., Craigie, R., Gellert, M. & Lambowitz, A. M.) 1147–1158 (ASM Press, 2002).
28. Delihas, N. Small mobile sequences in bacteria display diverse structure/function motifs. *Mol. Microbiol.* **67,** 475–481 (2007).
29. Curtis, M. D. & Grossniklaus, U. A gateway cloning vector set for high-throughput functional analysis of genes *in planta*. *Plant Physiol.* **133,** 462–469 (2003).
30. Clough, S. J. & Bent, A. F. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16,** 735–743 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Rice Genome Resource Center for the use of the rice microarray system and Y. Nagamura and R. Motoyama for technical support; the GenBank project of the National Institute of Agrobiological Science in Japan for providing seeds of Gimbozu landraces (A123 and A157); and X. Zhang and C. Feschotte for critical discussions and reading of the manuscript. S.R.W. is funded by a NSF Plant Genome grant and the University of Georgia Research Foundation and T.Tanisaka by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions K.N. and F.Z. performed 454 sequencing and analysed the data. A.O.R. provided statistical analyses. T.Tsukiyama and Y.O. performed microarray, and K.N. and H.S. analysed the data. C.N.H. performed *Arabidopsis* transformation. K.N. performed stress treatment and real-time PCR. K.N., F.Z., T.Tanisaka and S.R.W. contributed the experimental design and wrote the paper.

Author Information Sequence data has been submitted to GEO under accession number GSE15021. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.R.W. (sue@plantbio.uga.edu) or T.Tanisaka (tanisaka@kais.kyoto-u.ac.jp).

METHODS

Plants. Rice plants of Nipponbare, EG4 and its landraces A123 and A157 were grown in plastic pots in an environment-controlled greenhouse with a photoperiod of 14 h light (25–28 °C). DNA and RNA were extracted 7 days after germination. Before RNA extraction, plants were exposed to cold temperature (4 °C for 2 h), high-salinity (150 mM NaCl for 24h), or dehydration (deprived of water supply for 5 days) stress.

Sequencing and determination of *mPing* insertion sites. To determine *mPing* insertion sites in 24 individual sibling EG4 plants and their parent, flanking DNA fragments of *mPing* insertions were amplified using vectorette PCR[10]. Each DNA sample (~500 ng) was first digested overnight with BfaI at 37 °C, then ligated to linkers and amplified by nested PCR. To sequence all of the samples in a single sequencing run, the primers used in the second round of PCR contained unique 4-bp barcodes at their 5′ end. PCR products were column-purified with a PCR purification kit (Qiagen), pooled and subjected to pyrosequencing (454 Life Science, the University of Oklahoma).

All reads generated from pyrosequencing were processed as follows by a Python script. All sequences were first filtered to remove the sequences without discernable 5′ *mPing* sequences and then assigned into distinct groups based on their DNA barcode. Within each group sequences flanking *mPing* were extracted and those larger than 25 bp were searched against the rice genome database (RAP-DB) using BLAST with default parameters. Information on the best hits from the BLAST search was parsed to determine the location of *mPing* insertion sites in the rice genome. Redundant *mPing* insertions, defined as those that hit the same position in the genome, were removed from each group. Shared and unshared (unique) *mPing* insertions were identified between barcode groups by comparing non-redundant insertion sites.

For the control data set, 1,664 genome coordinates across the 12 chromosomes were randomly generated, and information on the surrounding sequence was parsed from the rep.gff file from RAP Annotation (release 2) (http://rapdb.dna.affrc.go.jp/rapdownload/rap2/gff_RAP2.tar.gz). The random insertion process was repeated 1,000 times. Replicates were used to estimate the expected number of insertions (and standard deviations) in each category under a random insertion model. The observed data of *mPing* insertions were compared to each replicate by a χ^2 test for the deviation from a random distribution.

RNA extraction and synthesis of Cy3- and Cy5-labelled cRNA. Total RNA was extracted from 7-day-old shoots of rice plants using an RNeasy Plant Mini Kit (Qiagen). Cy3- and Cy5-labelled cRNA was prepared from 400 ng of total RNA with Low RNA Input Linear Amplification Kit (Agilent Technologies) and Cy3- and Cy5-CTP (Perkin Elmer). Labelled cRNA was purified with RNeasy mini spin columns (Qiagen).

Microarray conditions and data analysis. A 44K Rice Oligo Microarray kit (Agilent Technologies) was used for microarray analysis. One microgram of Cy3-labelled cRNA was mixed with the same amount of Cy5-labelled cRNA and used for subsequent hybridization. Hybridization was carried out for 17 h with rotation at 60 °C. After washing, slides were scanned using a GenePix 4000A scanner (Axon Instruments) with 550 V and 680 V of PMT voltage for Cy3 and Cy5 detection, respectively, and quantified by Microarray Suite 2.0 (IPLab Spectrum Software, Scanalytics). Subsequent analysis was performed using GeneSpring 7 software (Agilent Technologies).

Genes that were up- or downregulated were selected as differentially transcribed between EG4 and NB. Signal intensity, amplitude of expression fluctuation and standard error of the mean F (F = the ratio of normalized data between NB and EG4) were also considered. Genes meeting the criteria were selected as follows: the average signal intensity of the control RNA in the experiment was within the range 53,103 to 13,107; the F of duplicate samples were all significantly higher or lower than 1 ($P < 0.01$).

Quantitative real-time PCR. Total RNA was prepared (for both rice and *Arabidopsis* seedlings) using an RNeasy Plant Mini Kit (Qiagen). Accumulation levels of the target transcripts were analysed by real-time PCR with a LightCycler 480 real-time PCR system (Roche Applied Science) by monitoring amplification with iQ SYBR Green Supermix (Bio-Rad) as described in the manufacturer's protocol.

Transformation of *Arabidopsis*. The *Ping* ORF1 promoter was cloned into the pDONR vector (Invitrogen) after PCR amplification. After sequencing, the promoter was transferred to pMDC 162 (ref. 30) by recombination, transformed into *Agrobacterium tumefaciens* (GV3103), and transformed into *Arabidopsis thaliana* ecotype Columbia by the floral dip method[29]. Seedlings were germinated on Murashige and Skoog solid medium (0.2% phytogel, 1% sucrose), containing 50 μg ml^{-1} kanamycin and 50 μg ml^{-1} hygromycin. Staining with the β-Glucuronidase Reporter Gene Staining Kit (Sigma-Aldrich) was performed before and after cold treatment (4 °C for 2 h).