the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

*The Plant Journal* (2009) doi: 10.1111/j.1365-313X.2009.04005.x

# A lineage-specific centromere retrotransposon in *Oryza brachyantha*

**Dongying Gao[1], Navdeep Gill[1], Hye-Ran Kim[2,3], Jason G Walling[2], Wenli Zhang[2], Chuanzhu Fan[3], Yeisoo Yu[3], Jianxin Ma[1], Phillip SanMiguel[4], Ning Jiang[5], Zhukuan Cheng[6], Rod A. Wing[3], Jiming Jiang[2] and Scott A. Jackson[1,*]**

[1]*Molecular and Evolutionary Genetics, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA,*
[2]*Department of Horticulture, University of Wisconsin-Madison, 1575 Linden Drive, Madison, WI 53706, USA,*
[3]*Arizona Genome Institute, Department of Plant Sciences, University of Arizona, 1657 E. Helen Street, Tucson, AZ 85721, USA,*
[4]*Genomics Core Facility, Purdue University, 915 W. State Street, West Lafayette, IN 47907, USA,*
[5]*Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA, and*
[6]*State Key Laboratory of Plant Genomics and Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China*

Received 6 June 2009; revised 28 July 2009; accepted 7 August 2009.
*For correspondence (fax 765 496 7255; e-mail sjackson@purdue.edu).

## SUMMARY

**Most eukaryotic centromeres contain large quantities of repetitive DNA, such as satellite repeats and retrotransposons. Unlike most transposons in plant genomes, the centromeric retrotransposon (CR) family is conserved over long evolutionary periods among a majority of the grass species. CR elements are highly concentrated in centromeres, and are likely to play a role in centromere function. In order to study centromere evolution in the *Oryza* (rice) genus, we sequenced the orthologous region to centromere 8 of *Oryza sativa* from a related species, *Oryza brachyantha*. We found that *O. brachyantha* does not have the canonical CRR (CR of rice) found in the centromeres of all other *Oryza* species. Instead, a new Ty3-gypsy (*Metaviridae*) retroelement (FRetro3) was found to colonize the centromeres of this species. This retroelement is found in high copy numbers in the *O. brachyantha* genome, but not in other *Oryza* genomes, and based on the dating of long terminal repeats (LTRs) of FRetro3 it was amplified in the genome in the last few million years. Interestingly, there is a high level of removal of FRetro3 based on solo-LTRs to full-length elements, and this rapid turnover may have played a role in the replacement of the canonical CRR with the new element by active deletion. Comparison with previously described ChIP cloning data revealed that FRetro3 is found in CENH3-associated chromatin sequences. Thus, within a single lineage of the *Oryza* genus, the canonical component of grass centromeres has been replaced with a new retrotransposon that has all the hallmarks of a centromeric retroelement.**

**Keywords: centromere, evolution, LTR retrotransposon, genomics, *Oryza*.**

## INTRODUCTION

Centromeres are essential for chromosome maintenance and transmission through cell division. Despite this absolute necessity, centromeres are highly divergent at the sequence level, both within and between species. For instance, the primary component of most centromeres is a satellite repeat, approximately nucleosomal in iteration (Jiang *et al.*, 2003; Lamb *et al.*, 2004), that can be highly divergent, even within a genus such as *Oryza* (rice) (Lee *et al.*, 2005). These satellite repeats can diverge very rapidly, on the order of a few million years.

In cereal plant species, however, a centromeric retrotransposon (CR) family is conserved among a broad range of species, including rice, maize, sorghum, wheat and sugarcane (Aragón-Alcaide *et al.*, 1996; Jiang *et al.*, 1996). CR is a *Ty3-gypsy*, or refered to as *Metaviridae* type (Hansen and Heslop-Harrison, 2004), retrotransposon that is highly restricted to the centromeric regions in different grass species. CRR (CR of rice) and CRM (CR of maize) elements are intermingled with centromeric satellite repeats, and are associated with CENH3, a centromere-specific histone H3 variant (Cheng *et al.*, 2002; Zhong *et al.*, 2002; Jin *et al.*, 2004; Nagaki *et al.*, 2004). Ideas as to the functional aspects of this conserved retrotransposon and the satellite repeats involve an RNA mechanism that is used to establish an

epigenetic/heterochromatic mark for centromere/kinetochore formation (Gieni *et al.*, 2008; Ishii *et al.*, 2008; Lamb *et al.*, 2008; Zhang *et al.*, 2008).

Retrotransposons are thought to be labile sequences in plant genomes that evolve and diverge quite rapidly, in fact the half life may be <3 Myr in rice (Vitte *et al.*, 2007). Therefore, it is unusual to find a specific retroelement that is conserved in sequence and in chromosomal location, such as this centromeric retrotransposon family. Thus, the invasion and domestication of the retrotransposons in cereal centromeres must predate the divergence of the cereals, and the CRR must predate the radiation of the *Oryza* genus.

Centromeres of many model eukaryotes consist of large quantities of repetitive elements, including centromeric satellite repeats and retrotransposons (Henikoff *et al.*, 2001), which are difficult to clone and sequence completely. For example, centromeres in Arabidopsis consist of 2.8–4 Mb of a 180-bp satellite repeat and a retroelement, *Athila*. Thus far, none of the Arabidopsis centromeres have been fully sequenced (Copenhaver *et al.*, 1999; Fransz *et al.*, 2000; Heslop-Harrison *et al.*, 2003; Hall *et al.*, 2004). Centromere 8 (Cen8) of Nipponbare (*Oryza sativa* L. ssp. *japonica*) was one of the first eukaryotic centromeres to be completely sequenced and analyzed. Cen8 consisted not only of satellite DNA and retrotransposons (including CRR elements), but also of active genes (Nagaki *et al.*, 2004; Wu *et al.*, 2004), which led to the suggestion that it may be a recently formed centromere. Comparative analysis of Cen8 sequences from Kasalath (*O. sativa* L. ssp. indica) and Nipponbare indicates that Cen8 regions of two subspecies of rice have undergone rapid variation, primarily as a result of the insertion and/or deletion of long terminal repeat (LTR)-retrotransposons (Wu *et al.*, 2009). In order to understand the evolution of Cen8 in the genus *Oryza*, we sequenced and analyzed the orthologous Cen8 region from a more distantly related *Oryza* species, *O. brachyantha*, which has a genome designation of FF, based on crossability among *Oryza* species and molecular analyses (Ge *et al.*, 1999). This genome diverged from *O. sativa* approximately 7–9 Mya. Surprisingly, the orthologous sequences from the FF genome lack canonical CRRs; moreover, a new retroelement (named FRetro3) appears to have replaced the CRR, probably acting as part of the functional centromeric retroelement in this species.

## RESULTS

### Identification of five new retrotransposons in *O. brachyantha*

Seed bacterial artificial chromosome (BAC) clones were chosen based on BAC library hybridization using probes derived from genes identified in the Cen8 region of *O. sativa* (Nagaki *et al.*, 2004; Wu *et al.*, 2004), as previously described (Ma *et al.*, 2007), and utilizing the BAC libraries and BAC fingerprint/end sequenced physical maps (Kim *et al.*, 2008). Seven minimum tiling paths of overlapping BACs (AC223438, AC223439, AC223440, 79B10, 03N12, 8B23 and 90E02) were shotgun sequenced, finished and sequence validated using standard procedures, as previously described (IRGSP., 2005). Sequences were then aligned to the *O. sativa* Cen8 sequence to confirm orthology, and for subsequent sequence analysis. We also used Nipponbare coding sequences (CDS) as queries to search against the sequence, and 23 orthologous Cen8 genes were found in the Cen8 region of *O. brachyantha*.

LTR_STRUCT (McCarthy and McDonald, 2003) was used to identify LTR retrotransposons in the orthologous BAC sequences from *O. brachyantha*. Five retrotransposon families were identified that ranged in size from 4934 to 12 450 bp, with LTRs of 418–3268 bp: we named these FRetro1–FRetro5. FRetro1 was classified as *Ty1-copia*, or *Pseudoviridae*, as suggested by Hansen and Heslop-Harrison (2004), the other four retroelements were classified into the *Ty3-gypsy* group based on sequence identity between their internal regions and retrotransposon sequences in GenBank, as well as the order of encoded gene products. The internal region (3385 bp) of FRetro1, which contained a 456-aa open reading frame (ORF), showed high similarity with *Ty1-copia* protein sequences in GenBank; however, the ORF did not carry the conserved domain of the *gag* and polyprotein (*pol*) genes, indicating that FRetro1 is a non-autonomous retrotransposon. The internal region (4037 bp) of FRetro2 encoded two ORFs. The first ORF was 128 aa, and was of unknown function, and the second ORF was 205 aa, and encoded a truncated *gag* gene. The 3804-bp internal sequence of FRetro4 contained a 648-aa ORF that probably encodes a non-functional *gag-pol* polyprotein based on sequence alignments. The internal sequence of FRetro5 is 10 078 bp in size and contains three ORFs: ORF1 and ORF2 encode *gag* and integrase (*INT*), respectively; the function of ORF3 is unclear. However, both *gag* and *INT* are truncated, and no functional reverse transcriptase (RT) was identified. All these data indicated that retrotransposons FRetro1, 2, 4 and 5 are non-autonomous elements (Figure 1). The internal region of FRetro3 (5928 bp) contained two ORFs. ORF1 contains all the necessary genes for movement, so FRetro3 is an autonomous retroelement; more details follow below.

The five retroelements were used as queries in blastn searches against GenBank databases to test whether these retroelements are new and/or specific to the *O. brachyantha* genome. In addition, we compared them with a comprehensive transposable element (TE) database using RECON (Bao and Eddy, 2002; NJ, unpublished data). No hits to Nipponbare or any other species were found using both the LTR and internal region of FRetro2 and FRetro5 LTR as queries. We found few blast hits using the LTRs and internal regions from the other three FRetro elements as queries, however, and all these matches have limited sequence

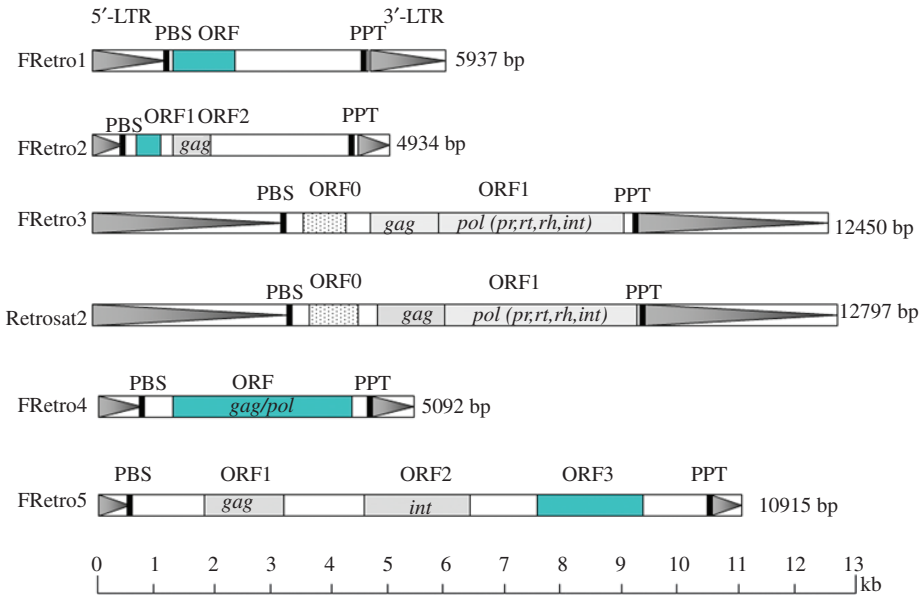


**Figure 1.** Structural comparison of FRetro1–FRetro5 and Retrosat2; 5′-LTR, 5′ long terminal repeat; 3′-LTR, 3′ long terminal repeat; ORF, open reading frame; PBS, primer binding sites; *gag*, group-specific antigen; *pol*, polyprotein; *pr*, protease; *rt*, reverse transcriptase; *rh*, RNAase-H; *int*, integrase; PPT, polypurine tracts. Gray ORFs, conserved domains; blue ORFs, truncated/undetermined similarity; and stippled ORFs, ORF unique to Retrosat2 and FRetro3.

identity with partial regions of the elements (Figure S1). As such, these elements either have no relative or only highly diverged relatives in the *Oryza* genomes, and can therefore be considered as novel retrotransposon families (Wicker *et al.*, 2007).

### FRetro3 is highly abundant in the Cen8 orthologous sequences from *O. brachyantha*

In order to determine the distribution of FRetro1–FRetro5 in the Cen8 orthologous sequences of *O. brachyantha*, all five FRetro elements and an unpublished rice TE database (NJ) were used to screen the sequences using RepeatMasker. Masked sequences were then divided into three groups (intact elements, intact solo LTRs and other elements) based on structural features. Intact elements were the sequences that contain two relatively intact LTRs flanked by target site duplications (TSDs). 'Intact solo-LTRs' includes the elements that contain intact LTR sequences flanked by a TSD. 'Other elements' contain truncated internal transposon sequences or LTR sequences, or both.

Based on these classification criteria, the five FRetro elements comprise approximately 29% of the sequences derived from the BAC clones. FRetro3 is the most abundant element, accounting for 22% of the total sequence (Table 1). Thus far, only a few other plant centromeres have been sequenced, including centromeres 4, 5 and 8 of *O. sativa*

**Table 1** Copy number of five retrotransposons and their percentage in the centromere 8 sequences of *Oryza brachyantha*

| Name of elements | | FRetro1 | FRetro2 | FRetro3 | FRetro4 | FRetro5 | Total |
|---|---|---|---|---|---|---|---|
| Intact elements | Copy number | 2 | 1 | 4 | 1 | 3 | 11 |
| | Coverage (bp) | 10 443 | 4934 | 47 160 | 5092 | 30 669 | 98 298 |
| | Percentage[a] | 1.10 | 0.52 | 5.00 | 0.54 | 3.24 | 10.38 |
| Intact solo LTR | Copy number | 0 | 0 | 22 | 0 | 1 | 23 |
| | Coverage (bp) | 0 | 0 | 71 116 | 0 | 417 | 71 533 |
| | Percentage[a] | 0.00 | 0.00 | 7.51 | 0.00 | 0.04 | 7.55 |
| Other elements | Copy number | 2 | 4 | 52 | 4 | 6 | 68 |
| | Coverage (bp) | 799 | 4063 | 91 979 | 2803 | 5025 | 104 669 |
| | Percentage[a] | 0.08 | 0.43 | 9.71 | 0.30 | 0.53 | 11.05 |
| Total | Copy number | 4 | 5 | 78 | 5 | 10 | 102 |
| | Coverage (bp) | 11 242 | 8997 | 210 255 | 7895 | 36 111 | 274 500 |
| | Percentage[a] | 1.19 | 0.95 | 22.19 | 0.83 | 3.81 | 28.98 |

[a]Percentage of total centromeric sequence (947 202 bp).

 *The Plant Journal*, (2009), doi: 10.1111/j.1365-313X.2009.04005.x

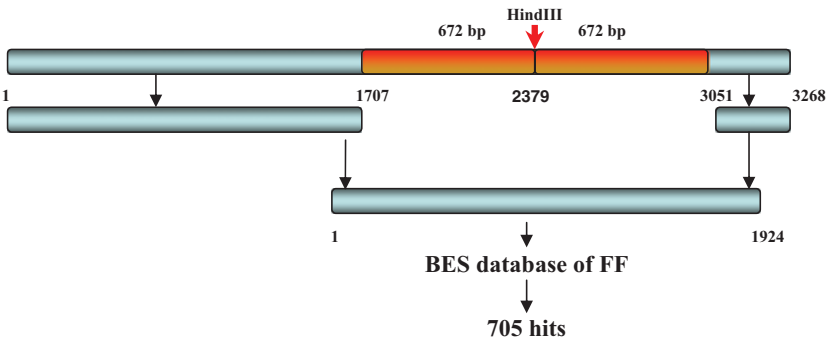


**Figure 2.** Estimation of copy number of FRetro3 in the *Oryza brachyantha* genome.
The red arrow indicates the *Hind*III restriction site, and the orange region shows the flanking 672-bp sequences of *Hind*III; the 1924-bp cut-out part was used to search against the BAC end sequences (BES) database of *O. brachyantha*.

ssp. *japonica* (Nagaki *et al.*, 2004; Wu *et al.*, 2004; Zhang *et al.*, 2004; IRGSP, 2005). None of these centromeres have a retrotransposon as abundant as FRetro3 is in these centromeric sequences of *O. brachyantha*.

In order to investigate the distribution of FRetro3 in the entire FF genome, the LTR sequence of FRetro3 was used as a query to search against a BAC end sequences (BES) database of *O. brachyantha* (http://www.omap.org), similar to the approach used by Jiang *et al.* (2002). The LTR sequence of FRetro3 is 3268 bp, much larger than the BAC end sequences (approximately 600–700 bp), and contains a *Hind*III recognition site, the enzyme used to construct the BAC library of *O. brachyantha*. Thus, we suspected that we may overestimate the copy number using the whole LTR sequence of FRetro3 as a query, as the appearance of the LTR would not be random in the BESs. Therefore, we removed 672 bp on each side of the *Hind*III recognition site from the LTR sequence, the two remaining sequences were joined, and were then used as a query to search the BES of *O. brachyantha* (Figure 2). To improve the accuracy of the data, we used a cut-off *e*-value of $<10^{-15}$. The BESs averaged 672 bp in length, and the copy number of FRetro3 was estimated to be 2816 [(number of hits × FF genome size/nt in the BES database)/2 - (705*362 Mb/45.3 Mb)/2 = 2816]. The results were divided by two, as a typical, intact retroelement carries two LTRs. This is a very conservative estimate, as many elements are truncated, and do not contain both LTRs, even in the Cen8 region (Table 1).

### Structural analysis of FRetro3

After manual analysis of the BAC sequences, we found three other intact retroelements that ranged in size from 10 622 to 12 301 bp. Their LTRs shared 91–93% sequence identity with the LTRs of FRetro3, and their translated internal sequences shared 73–77% amino acid homology with the sequence of FRetro3: therefore, these intact retroelements belonged to the FRetro3 family. We named these FRetro3-1, FRetro3-2 and FRetro3-3. We also found a total of 23 solo LTRs: 22 from the FRetro3 family and one from the FRretro5 family (Table 1). It was interesting that most of the solo LTRs (22/23) were from a single family, FRetro3. Each of the 22 solo LTRs was flanked by identical TSDs, with only two solo LTRs sharing the same TSD. No intact solo LTRs were found for the FRetro1, FRetro2 and FRetro4 families.

Unequal homologous recombination is responsible for the formation of solo LTRs. Intra-element unequal recombination can produce solo LTRs with the same TSD; however, inter-element unequal recombination usually leads to solo LTRs with different TSDs (Devos *et al.*, 2002). All 22 of the FRetro3 solo LTRs are flanked by the same TSDs, indicating that intra-element unequal recombination was more common than inter-element unequal recombination in the Cen8 of the FF genome. The ratio of solo LTRs to intact elements for the FRetro3 family in FF Cen8 is 5.5:1.

In order to provide insight into the history of the FRetro3 family, 22 solo LTRs and eight LTRs from the four intact elements (FRetro3, FRetro3-1, FRetro3-2 and FRetro3-3) were used to construct a phylogenetic tree. These LTRs were grouped into two distinct subfamilies of FRetro3, with LTRs from the four intact elements in subfamily A (Figure 3). The LTRs in subfamily B were larger than subfamily A (3349 bp versus 3128 bp, on average). Further analysis of the aligned LTRs revealed that some regions are more variable than others. For example, a 41-bp T-rich region (from 293 to 333 bp) exhibited a high frequency of deletion and transition mutation (from T to C), so that no two LTRs were identical to each other in this region (Figure S2a). Other variable regions included two GC-rich domains (from 585 to 634 bp and from 647 to 670 bp), where a GCC motif was frequently present (Figure S2b). It is not clear what role these variable regions may have had in the size variation observed, or even in the propensity of this LTR to form solo LTRs.

Genomic contraction can result from the formation of solo LTRs and the removal of the internal part of the retrotransposons via unequal homologous recombination (Shirasu *et al.*, 2000; Devos *et al.*, 2002). Abundant solo LTRs of

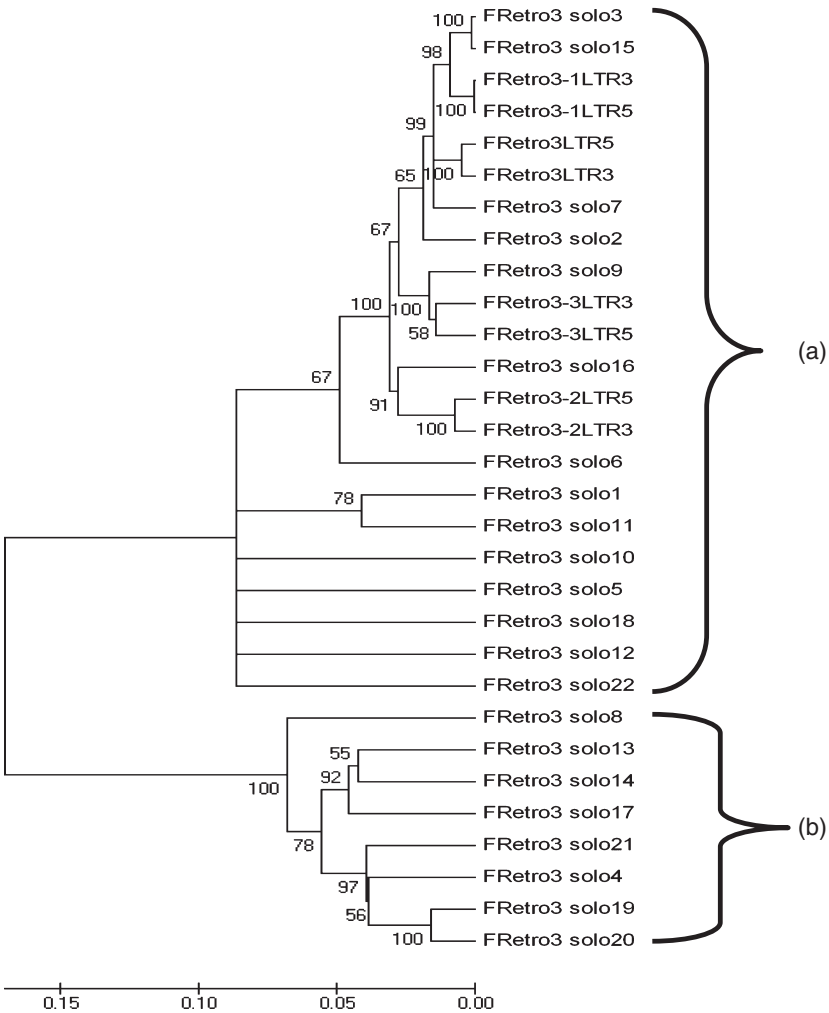


**Figure 3.** Phylogenetic tree based on 30 long terminal repeat (LTR) sequences.

FRetro3 in the BAC sequences suggest that many deletion events occurred during the formation of this centromere. We calculated the quantity of DNA lost as a result of the formation of solo LTRs and/or deletions from these five retroelements (Table 2). The results indicate that about 76% of all five TEs had been deleted, and about 78% of FRetro3 (760 845 bp) has been deleted, which is 86.5% (760 845 bp/879 628 bp $\times$ 100) of the total deleted TE DNA in this region.

The five FRetro elements were often nested. Three regions showed significant nesting, where FRetro1 and FRetro5 were interrupted by insertions of FRetro3. These three regions also had a higher than average number of solo LTRs (on average, one FRetro3 solo LTR per 48 kb). In the first region, two intact solo LTRs and three truncated solo LTRs of FRetro3 were inserted into a single copy of FRetro5 (Figure 4a), resulting in an average of one FRetro3 solo LTR per 12 kb. In the second region, FRetro5 was interrupted by three intact solo LTRs, two truncated solo LTRs and one truncated FRetro3 element (Figure 4b), resulting in one FRetro3 solo LTR per 8.6 kb. In the third region, an FRetro1 element was

**Table 2** Genome contraction resulting from the formation of solo long terminal repeats (LTRs) in the centromere 8 of *Oryza brachyantha*

| Transposon (TE) | FRetro1 | FRetro2 | FRetro3 | FRetro4 | FRetro5 | Total |
|---|---|---|---|---|---|---|
| Copy number of TE | 4 | 5 | 78 | 5 | 10 | 102 |
| Current TE coverage (bp) | 11 242 | 8997 | 210 255 | 7895 | 36 111 | 274 500 |
| Estimated original TE coverage (bp) | 23 748 | 24 670 | 971 100 | 25 460 | 109 150 | 1 154 128 |
| TE DNA deleted (bp) | 12 506 | 15 673 | 760 845 | 17 565 | 73 039 | 879 628 |
| Percentage of TE DNA that was deleted (%) | 52.66 | 63.53 | 78.35 | 68.99 | 66.91 | 76.22 |

 

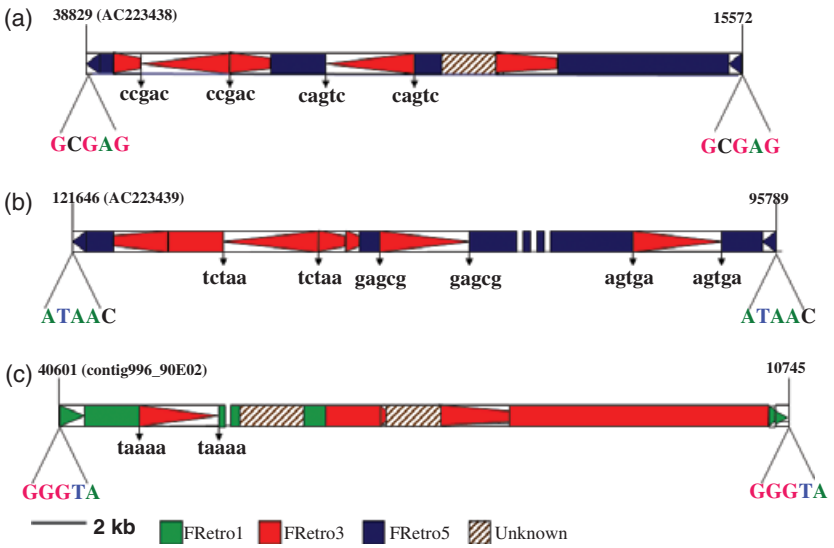


**Figure 4.** Sequences of FRetro3 inserted into FRetro5 (a & b) and FRetro1 (c). Upper-case letters are target site duplications (TSDs) of FRetro5 or FRetro1, small letters are TSDs of solo long terminal repeats (LTRs) of FRetro3.

interrupted by one intact solo LTR, two truncated solo LTRs and two truncated FRetro3 elements (Figure 4c), resulting in one FRetro3 solo LTR per 30 kb.

### Integration time of FRetro3 and its structural similarity to Retrosat2

When a retroelement integrates into the genome, the two LTR sequences are assumed to be identical. Thus, we can estimate the insertion time of LTR-retrotransposons based on the sequence divergence between LTRs. Because the LTR sequences evolve more rapidly than genes, we used an average substitution rate ($r$) of $1.3 \times 10^{-8}$ substitutions per synonymous site per year to estimate insertion times, as described by Ma and Bennetzen (2004). LTR sequences of the four intact copies of the FRetro3 family were used to calculate the times of integration. Estimates for these four elements ranged from 0.04 to 0.96 Myr (Table 3). Times of integration of the other four FRetro elements were estimated to have occurred within the past 3 Myr (Table 3). As most of the FRetro3s are present as either solo LTRs or fragments, it is likely that the majority of the FRetro3 family was amplified 1 Mya.

**Table 3** Insertion time of five retrotransposons in the centromere 8 of *Oryza brachyantha*

| Name | Location | Size (bp) | *K* | Insertion time (Myr) |
|---|---|---|---|---|
| FRetro1 | 780 557–786 493 | 5937 | 0.032 | 1.23 |
| FRetro2 | 834 150–839 083 | 4934 | 0.044 | 1.69 |
| FRetro3 | 732 135–744 584 | 12 450 | 0.009 | 0.35 |
| FRetro3-1 | 203 510–214 131 | 10 622 | 0.001 | 0.04 |
| FRetro3-2 | 230 335–242 116 | 11 787 | 0.02 | 0.77 |
| FRetro3-3 | 277 618–289 918 | 12 301 | 0.025 | 0.96 |
| FRetro4 | 29 485–34 576 | 5092 | 0.017 | 0.65 |
| FRetro5 | 443 249–454 163 | 10 915 | 0.072 | 2.77 |

Two phylogenetic trees were built based on whole-element sequences and on conserved reverse transcriptase (RT) domains. First, complete sequences from 41 different *gypsy*-like retrotransposons were used to construct a phylogenetic tree. Elements known to be located in centromeric and/or pericentromeric regions were included, such as *RIRE7* (Kumekawa *et al.*, 2001), CRR1-CH1-2, CRR2-CH1-1, noaCRR1-CH1-1 and noaCRR2-CH1-1 (Nagaki *et al.*, 2005), cereba from barley (Presting *et al.*, 1998), CRM from maize (Zhong *et al.*, 2002) and *Jinling* from tomato (Wang *et al.*, 2006). *RIRE7*, CRR1-CH1-2, CRR2-CH1-1, noaCRR1-CH1-1, noaCRR2-CH1-1, cereba and CRM were grouped into the same clade; however, FRetro3 was in another clade, along with Retrosat2, *RIRE3*, *RIRE8* and tomato retroelement *Jinling* (Figure 5a).

The RT conserved domains from 25 different *gypsy*-like elements were used to make a second phylogenetic tree. Elements that lacked RT domains or contained truncated RT domains were not inlcuded in this analysis. These 25 elements grouped into two major families: the RIRE2 family and the Reina family (Figure 5b). Although FRetro3 and the CRRs belong to the Reina family, they were grouped into different clades (Figure 5b). Both trees provided similar results: FRetro3 is more similar to *RIRE3*, *RIRE8* and *Jinling* than to the CR family, such as the CRRs (including RIRE7), CRM and cereba (the CR of barley). These data also indicated that FRetro3 is related to Retrosat2.

Retrosat2 is a *Ty3-Gypsy* type retrotransposon originally identified from the *O. sativa* ssp. *indica* genome as a nested element (accession no. AF111709). Sequence alignments between FRetro3 and Retrosat2 indicated that sequence identity is <60%; however, FRetro3 and Retrosat2 share

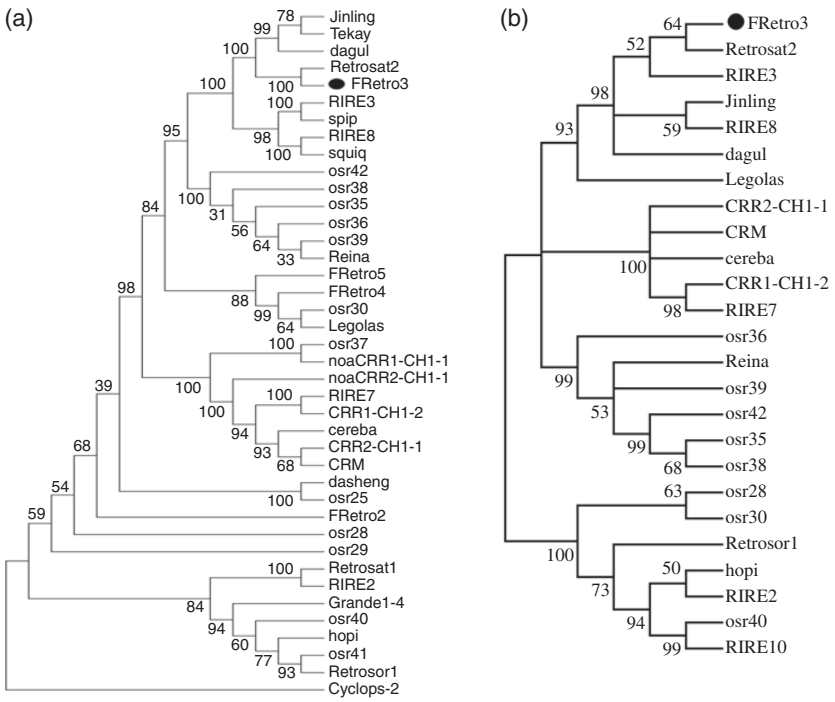


**Figure 5.** (a) Phylogenetic tree based on whole transposable element (TE) sequences.
(b) Phylogenetic tree based on conserved reverse transcriptase (RT) domains of retrotransposons.

many similarities. They are similar in size: FRetro3 is 12 450 bp and Retrosat2 is 12 797 bp. Both retroelements are autonomous elements, and their internal regions have the capacity to encode *gag*- and *pol*-like proteins. Interestingly, internal sequences of both elements encode an extra ORF (ORF0), located between the primer binding sites (PBSs) and the ORF encoding multiple proteins (Figure 1). FRetro3 and Retrosat2 also share the same PBS (primer binding sites) sequence and nearly identical PPT (polypurine tracts) sequences. They also share a similar PBS with *Jinling*, *CRRs*, *RIRE7* and *CRM*; however, PPT sequences of FRetro3 and Retrosat2 are different than those of *Jinling*, *CRRs*, *RIRE7* and *CRM*.

### CRR-related sequences in the FF genome

Lee *et al.* (2005) cloned DNA fragments from CENH3-associated chromatin, and surprisingly did not recover any CRR-related sequences from *O. brachyantha*. Moreover, Southern blot hybridization using a CRR probe showed only weak hybridization to genomic DNA from *O. brachyantha* (Lee *et al.*, 2005). We searched the sequences from the seven Cen8 BAC clones from *O. brachyantha* with CRR sequences (Nagaki *et al.*, 2005), and did not find any intact CRR elements. We found only 12 short sequences, ranging in size from 117 to 933 bp, with low sequence similarity to the CRR elements, suggesting that they may be relics of ancient CRR elements. We also blasted the BES database of *O. brachyantha* using the CRRs as a query and found a few hits, all of which aligned to short regions of the CRRs. For example, the best match was only 289 bp in length with 89% sequence similarity when using LTR of CRR2_CH1-1 (900 bp) as a query. LTR fragments of CRR1, CRR2, noaCRR1 and noaCRR2 subfamilies were PCR amplified and used as a probe to a Southern blot of genomic DNA from 14 rice species, and a weak signal was detected in *O. brachyantha* (Figure 6a). All these results suggest that CRR was present in the ancestor of *O. brachyantha*; however, it is in the process of extinction.

### FRetro3 is a new centromeric retrotransposon specific to *O. brachyantha*

A Southern blot of *Oryza* species was conducted using a 436-bp sequence of FRetro3 LTR region as a probe to assess the distribution of FRetro3 in the genus. A strong signal was seen for the *O. brachyantha* genome (Figure 6b), but little or no hybridization was seen in the other species, indicating that FRetro3 is more highly abundant in the *O. brachyantha* genome, or was either absent or very divergent in other species.

In order to determine the chromosomal distribution of FRetro3 in the FF genome, we conducted fluorescence *in situ* hybridization (FISH) analysis using FRetro3 and the centromeric satellite repeat CentO-F as probes. On meiotic pachytene chromosomes, the majority of the FISH signals derived from the FRetro3 probe flanked the CentO-F satellite repeat arrays (Figure 7a–d). FRetro3 signals that overlapped the CentO-F repeats were also observed, but these signals were generally weaker than those in the pericentromeric regions. Fiber-FISH analysis using the same two probes revealed that most of the long clustering FRetro3 signals were indepen-

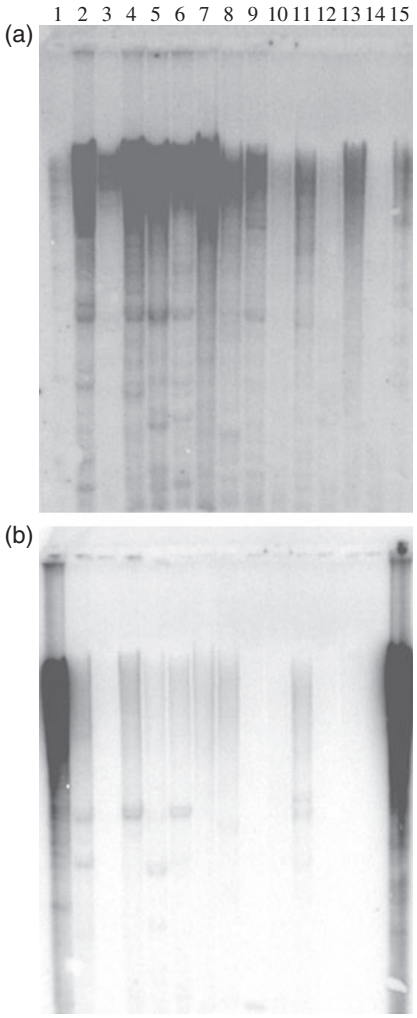


**Figure 6.** Southern blot of 14 rice species using the centromeric retrotransposon of rice (CRR) long terminal repeats (LTRs) (a) and the FRetro3 LTR sequence (b) as probes, respectively: (1) *Oryza brachyantha*, (2) *Oryza sativa* (Nipponbare), (3) *Oryza glaberrima*, (4) *Oryza nivara*, (5) *Oryza longistaminata*, (6) *Oryza rufipogon*, (7) *Oryza minuta*, (8) *Oryza officinalis*, (9) *Oryza punctata*, (10) *Oryza alta*, (11) *Oryza australiensis*, (12) *Oryza granulata*, (13) *Oryza ridleyi*, (14) *Oryza coarctata* and (15) *Oryza brachyantha*.

dent from the CentO-F signals. However, insertion of the FRetro3 within CentO-F arrays was also observed (Figure 7e).

## DISCUSSION

Transposable elements, specifically the retrotransposon class, are labile components of plant genomes that can proliferate rapidly (SanMiguel *et al.*, 1996; Bennetzen and Kellogg, 1997). These elements can decompose either by deletion (recombinational or non-recombinational) or by sequence degeneration (Devos *et al.*, 2002; Vitte and Panaud, 2003; Ma *et al.*, 2004; Vitte *et al.*, 2007). An exception to this general rule is the centromeric retrotransposon family found in cereal genomes, such as the CRRs of rice (Cheng *et al.*, 2002; Nagaki *et al.*, 2005) and the CRMs of maize (Zhong *et al.*, 2002; Nagaki *et al.*, 2003). Both LTRs and coding sequences derived from the CR elements are highly conserved across the Gramineae (Miller *et al.*, 1998; Presting *et al.*, 1998). Thus, it was surprising to observe among the 12 genome types of the genus *Oryza* that only *O. brachyantha* does not contain the CRR elements at its centromeres.

Comprehensive analyses of the FRetro3 element in the present study show that this retroelement has replaced CRR as the dominant centromeric retroelement in *O. brachyantha.* First, based on sequence analysis and FISH, it is located predominantly in the centromeres. Second, the copy number in the orthologous region of *O. brachyantha* is much higher than any of the other retroelements (present in single copies only). Third, the organization of FRetro3 is similar to the organization of CRRs. Insertions of the FRetro3 elements into arrays of centromeric satellite repeat CentO-Fs were observed in fiber-FISH analysis, although the FRetro3/CentO-F intermingling does not appear to be as extensive as CRR/CentO. Fourth, in centromeric heterochromatin, the histone 3 (H3) subunit is replaced with a centromere-specific H3 (Henikoff *et al.*, 2001). We found that eight out of the 96 sequences derived from the CENH3-associated chromatin reported by Lee *et al.* (2005) belong to the FRetro3 family. These results showed that at least some FRetro3 elements are located in the functional domains of *O. brachyantha* centromeres.

Retroelements of centromeres are often inserted into satellite arrays (Lamb *et al.*, 2008). However, at lower frequencies, other copies may also be found interspersed throughout the genome. Satellite arrays, unlike the centromeric retroelements, are highly variable and evolve quickly, even within a genus (Lee *et al.*, 2005). It is not clear what evolutionary constraints act upon the centromeric retroelements to maintain their sequence and chromosomal position, but in the FF genome of rice, these constraints have been broken, and the CRR element has been replaced by another *Ty3-gypsy* retroelement that is specific to the FF genome. Two questions arise: (i) where did this element come from and (ii) how/why did it selectively replace the CRR element, only in this genome?

The *O. brachyantha* genome last shared an ancestor with *O. sativa* about 7–9 Mya (Ge *et al.*, 1999); therefore, we propose that FRetro3 is <7–9 Myr old. This is supported by the observation that all full-length copies of FRetro3 were <1 Myr old. Maize centromeric retrotransposon CRM and barley centromere element *cereba* share sequence similarity with the rice centromere elements CRRs (Miller *et al.*, 1998; Sharma and Presting, 2008). The PBS and PPT domains are highly conserved between CRR and CRM, indicating a common link between rice and maize centromeric retrotransposons. FRetro3 belongs to a different clade than the CRRs, *ceraba* and CRM; moreover,

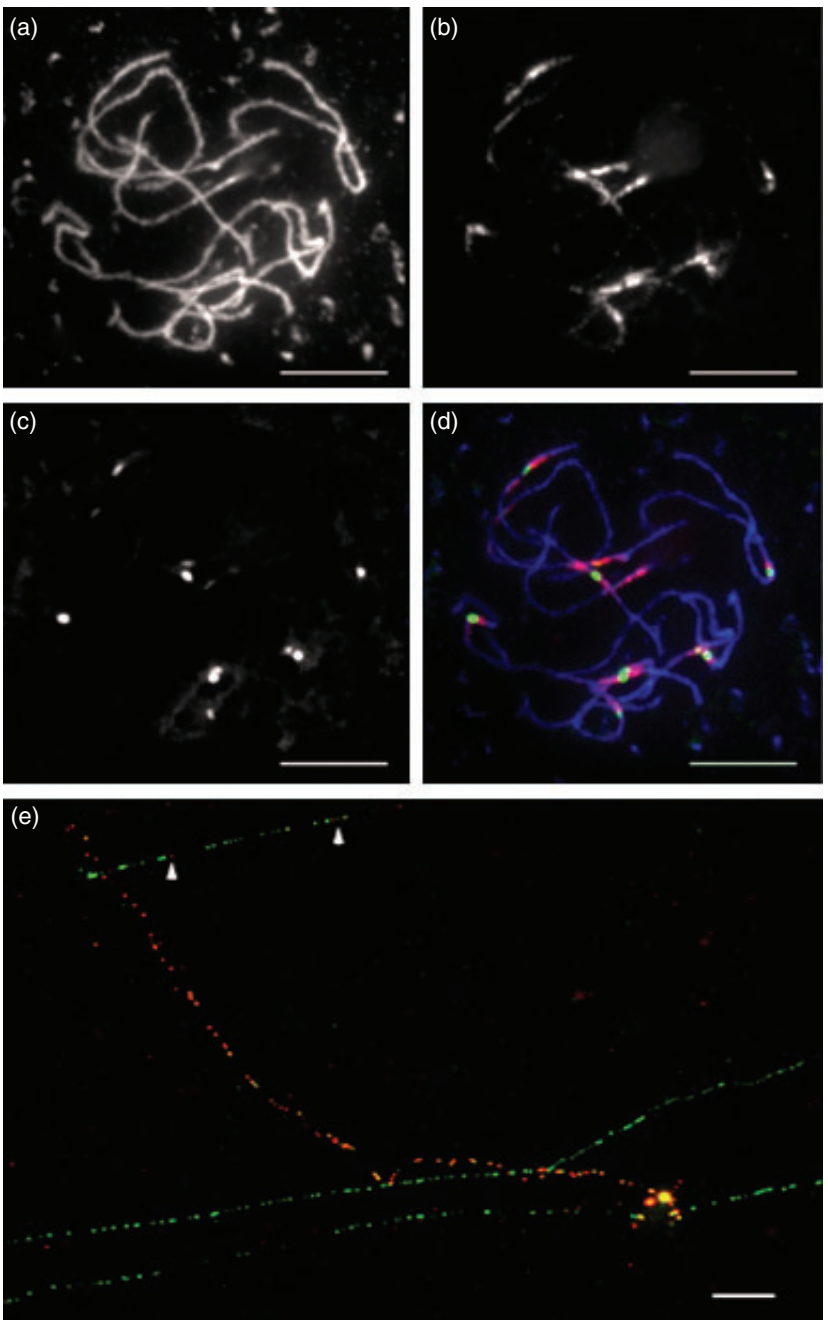


**Figure 7.** (a) Fluorescence *in situ* hybridization (FISH) of FRetro3 to pachytene chromosomes of *Oryza brachyantha*: (a) DAPI counterstain (blue), (b) FRetro3 (red), (c) CentO-F (green) and (d) merged image. (e) Fiber-FISH of FRetro3 (red signals) and CentO-F (green signal) to extended DNA fibers from *O. brachyantha*. Arrows indicated red signal (FRetro3) in arrays of CentO-F. Scale bars: 10 μm.

the CentO-F satellite repeats of the FF genome show no sequence similarity to the centromere repeats of other rice species (Lee *et al.*, 2005). All these results suggest that the centromeres of the *O. brachyantha* genome have recruited a different set of sequences than *O. sativa* (AA).

Many retrotransposons in the rice genus have been discovered using *in silico* or experimental approaches. Some retrotransposons have been reported with no species-specific distribution. For example, *RIRE1* was originally identified in *Oryza australiensis* (EE), but is also present in other rice species (Noma *et al.*, 1997; Roulin *et al.*, 2008). Three other retrotransposons, *RIRE2*, *Wallabi* and *Gran3*, characterized in *O. sativa* (AA), *O. australiensis* (EE) and *Oryza granulata* (GG), respectively, were also detected across the rice genus (Ammiraju *et al.*, 2007). These results indicated that these elements, such as *RIRE1*, might be ancient and already present in the ancestor of the rice genus before these *Oryza* species diverged from a common ancestor. In contrast to these elements, FRetro3 is

 

found only in the *O. brachyantha* genome, which implies that FRetro3 is a younger family than *RIRE1* and the other three TEs.

In order to provide more insight into the evolutionary history of FRetro3, a detailed TE annotation of chromosome 8 in Nipponbare was undertaken. FRetro3 was completely absent in chromosome 8. However, 102 Retrosat2 elements were identified, including 16 full elements and 46 solo LTRs, of which one intact element and nine intact solo LTRs were found in the Cen8 region. None of the centromeric Retrosat2s have a TSD in common with the FRetro3s from FF Cen8. It is interesting to note that Retrosat2 is distributed along the entire chromosome 8 (Figure S3), but is not concentrated at the centromeric region, as is FRetro3. Insertion times of Retrosat2s on chromosome 8 vary from 0 to 2.28 Myr (Table S1).

Given the overall sequence and structural similarity between Retrosat2 and FRetro3, it is possible that they derived from a common ancestor at a certain evolutionary point, although it is not clear whether they share an immediate ancestor. The absence of FRetro3-like LTRs in other species of *Oryza* could have resulted from either the fast divergence of LTR sequences or the lineage that led to FRetro3 being lost in these species. Finally, we cannot rule out the possibility that FRetro3 was introduced to *O. brachyantha* via horizontal transfer.

It remains to be seen if the FRetro3 elements function similarly to CRRs. When and why this genome type recruited a new retrotransposon to its centromeres, and 'eliminated' the family conserved across the cereals, remain questions to be answered. They could probably be answered, in part, by functional assays to show where the active kinetochore is established in the Cen8 of *O. brachyantha*, by the replacement of the canonical H3 subunit by CENH3 (Jiang *et al.*, 2003). The timing of the replacement of the CRR element by the FRetro3 can be estimated in part by the timing of insertions of the FRetro3s that occurred in the last 1 Myr. Finally, the mechanism by which the CRR elements were eliminated is not clear, but we do find low levels of homology with the CRRs in the orthologous Cen8 sequences from *O. brachyantha*, although very fragmented. The elimination or removal may have been a passive process, although we suspect, given the timing, that it was most likely an active process. One hypothesis might be that the FRetro3 family invaded the centromeres of *O. brachyantha* followed by the elimination of the CRRs. In the FF centromere there is an active turnover of retroelements to form solo LTRs in the centromere: as shown by the high levels of solo LTRs compared with full-length elements. If the CRRs lost their ability to transpose, they may have been lost through active deletions to form solo LTRs and other fragments, and so the FRetro3s accumulated there instead.

## EXPERIMENTAL PROCEDURES

### Plant materials

The cultivated rice (*O. sativa*, AA) variety Nipponbare and another 13 wild-rice species: *Oryza glaberrima* (AA), *Oryza nivara* (AA), *Oryza longistaminata* (AA), *Oryza rufipogon* (AA), *Oryza punctata* (BB), *Oryza minuta* (BBCC), *Oryza officinalis* (CC), *Oryza alta* (CCDD), *O. australiensis* (EE), *O. brachyantha* (FF), *O. granulata* (GG), *Oryza ridleyi* (HHJJ) and *Oryza coarctata* (HHKK) were planted in a glasshouse at Purdue University. DNA was extracted from young leaves of all 14 rice species using the cetyltrimethyl ammonium bromide (CTAB) method.

### Analysis of the TEs of the Cen8 sequence of *O. brachyantha*

In order to identify transposable elements in the centromere sequence, all identified retrotransposons in the *O. brachyantha* genome and the rice transposon library (NJ, unpublished data) were combined and used as a TE library database to screen the centromere sequence with RepeatMasker (http://www.repeatmasker.org). The program was run using the 'nolow' option to avoid masking the low-complexity DNA or simple repeats, besides other default parameters. In addition to the above parameters, we also set a cut-off score of >300, and a hit sequence length of >50 bp. Any hits that did not fit these criteria were removed when our analysis identified a sequence as a TE or TE fragment. All the desired hits were then inspected manually to determine the exact boundaries of each element and their TSD. Although a global TE annotation of the centromere sequence was carried out, this study focused on the analysis of retrotransposons that we originally characterized in the *O. brachyantha* genome. Other TEs data will be reported later.

### Fluorescence *in situ* hybridization and fiber-fluorescence *in situ* hybridization

FISH and fiber-FISH experiments were performed using CentO-F (CentO-F 37-2; Lee *et al.*, 2005) and FRetro3 (clone HIv2BC10) as probes to either meiotic chromosomes (Cheng *et al.*, 2001) or extended DNA fibers (Jackson *et al.*, 1998), following previously published protocols.

Briefly, DNA extracts from both clones were nick translated with either biotin dUTP or digoxigenin dUTP (Roche, http://www.roche.com). Pachytene chromosomes were isolated on slides from fixed *O. brachyantha* anther tissue, denatured and co-hybridized with the two differently labeled probes. DNA fibers for fiber-FISH were isolated from *O. brachyantha* nuclei, extended on poly-L-lysine slides and co-hybridized as above. The probes used for pachytene FISH, CentO-F (biotin) and FRetro3 (digoxigenin), were visualized using a single layer of Alexafluor 488 streptavidin (Invitrogen, http://www.invitrogen.com) and mouse anti-digoxigenin (Roche), conjugated with rhodamine, respectively. Chromosomes were counterstained using 4′,6-diamidino-2-phenylindole (DAPI). Probe detection on extended fibers required multiple layers of antibodies to enhance detection, as described in Walling *et al.* (2005).

Slides were analyzed and digital images captured using an Olympus BX60 epifluorescence microscope (Olympus, http://www.olympus.com) coupled to a Hamamatsu CCD (Hamamatsu, http://www.hamamatsu.com) camera, controlled with Metamorph imaging software (http://www.moleculardevices.com/pages/software/metamorph.html). Final adjustments and publication images were made using Adobe Photoshop 7.0. (Adobe, San Jose, CA).

### Estimation of the insertion time of LTR-retrotransposons

5′ and 3′ terminal repeat sequences of all retrotransposons were first aligned using blastn2 (http://blast.ncbi.nlm.nih.gov/bl2seq/wblast2.cgi) comparisons, in order to determine and confirm the exact LTR boundaries of each element. Subsequently, two LTR sequences of the elements were aligned, and the *K* value (average number of substitutions per aligned site) was estimated with the Kimura-2 parameter using MEGA 4 (Tamura *et al.*, 2007). An average substitution rate (*r*) of $1.3 \times 10^{-8}$ substitutions per synonymous site per year was used to calibrate insertion times, as described by Ma and Bennetzen (2004). The insertion times (*T*) were calculated using the formula: $T = K/2r$.

### Southern blot

Genomic DNAs of all 14 rice species were digested by *Eco*RI (Invitrogen) at 37°C for 10 h. The digested DNAs were separated by electrophoresis on a 1.0% (w/v) agarose gel at 55 V for 11 h, and were blotted onto Hybond $N^+$ membrane (Amersham Biosciences, now part of GE Lifesciences, http://www.gelifesciences.com). A 436-bp sequence of the FRetro3 LTR region was used as a probe to detect the presence of FRetro3 in different rice species. The PCR fragment was labeled with [$^{32}$P]dCTP using the rediprime II random prime labeling system (Amersham Biosciences, now part of GE Lifesciences) according to the manufacturer's instructions. Genomic DNA of *O. brachyantha* and Nipponbare DNA were used as templates to amplify FRetro3 and CRR fragments, respectively. The primers used were as follows: FRetro3 (forward, 5′-AGTCTCCGTTTAGGTCCATT-3′; reverse, 5′-TCCCATGAGCTATTTGTTCT-3′); CRR1 (forward, 5′-GCAAGGACCAATGACTAGAG-3′; reverse, 5′-CAAGCAAGAACAAGTTGACA-3′); CRR2 (forward, 5′-TGTACAGCATGATGGTCCTA-3′; reverse, 5′-AATCGAAGAACAAGCAAGAA-3′); noaCRR1 (forward, 5′-TACACTGCTGACTTCAAACG-3′; reverse, 5′-CTTAGCGATCGATACACCTC-3′); noaCRR2 (forward, 5′-ATGATGAGGAAATCACTTCG-3′; reverse, 5′-AATGCAAACGAGAGAACACT-3′). Blots were hybridized at 58.5°C for overnight, and were washed in 1.5 × SSC solution for 30 min, and then in 1 × SSC for 30 min. The membrane was exposed on a Fuji-image plate, and the hybridization signals were captured using a Fujifilm FLA-5100 multifunctional scanner (Fujifilm, http://www.fujifilm.com).

### Construction of phylogenetic trees

In total, 41 *gypsy*-like plant retrotransposon sequences were used to make phylogenetic trees, including: four novel retrotransposons of *O. brachyantha*, identified in this study; 28 rice retrotransposons; three maize retrotransposons – *Tekay* (accession no. AF050455), *Reina* (accession no. U69258) and CRM (accession no. AY129008); the teosinte retrotransposon *Grande1-4* (accession no. X97604); *Retrosor1* in sorghum genome (accession no. AF098806); *cereba* in barley (AY040832); *Cyclops-2* in pea, *Jinling* in tomato (accession no. DQ445619) and *Legolas* in Arabidopsis (accession no. AC006570).

The internal region of each retrotransposon was annotated for ORFs and translated into amino acid sequences using FGENESH (http://linux1.softberry.com/berry.phtml) and GENEMARK (http://exon.gatech.edu/GeneMark). Multiple sequence alignment of all these amino acid sequences of retrotransposons were performed with the conserved regions of RT domains, which have been described previously (Xiong and Eickbush, 1990; Kumekawa *et al.*, 1999). In addition, the amino acid sequences were also used as queries to search against the Gypsy Database (GyDB) (Lloréns *et al.*, 2008), to detect RT conserved sequences in the GyDB. The full element sequences and conserved RT sequences were used to generate multiple alignments using CLUSTALW (http://www.ebi.ac.uk/clustalw) with default options. Phylogenetic trees were generated using the neighbor-joining method in MEGA. The analysis was based on 1000 bootstrap replicates, using the nucleotide maximum composite likelihood model.

## ACKNOWLEDGEMENTS

This study was supported by grants from The National Science Foundation DBI 0603927 (JJ, SAJ and RAW) and 0424833 (SAJ).

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Figure S1.** Graphic summary of sequences producing significant alignments using long terminal repeats (LTRs) and internal regions of the five FF Cen8 retroelements as queries.

**Figure S2.** Two variable regions of the FRetro3 long terminal repeat (LTR) sequence.

**Figure S3.** Distribution of Retrosat2 on chromosome 8 of *Oryza sativa* cv. Nipponbare.

**Table S1.** Insertion times of Retrosat2 on chromosome 8 of Nipponbare.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## REFERENCES

**Ammiraju, J.S.S., Zuccolo, A., Yu, Y. *et al.*** (2007) Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus Oryza. *Plant J.* **52**, 342–351.

**Aragón-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S. and Moore, G.** (1996) A cereal centromeric sequence. *Chromosoma*, **105**, 261–268.

**Bao, Z. and Eddy, S.R.** (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **12**, 1269–1276.

**Bennetzen, J.L. and Kellogg, E.A.** (1997) Do plants have a one-way ticket to genomic obesity? *Plant Cell*, **9**, 1509–1514.

**Cheng, Z., Buell, C.R., Wing, R.A., Gu, M. and Jiang, J.** (2001) Toward a cytological characterization of the rice genome. *Genome Res.* **11**, 2133–2141.

**Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R. and Jiang, J.** (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell*, **14**, 1691–1704.

**Copenhaver, G.P., Nickel, K., Kuromori, T. *et al.*** (1999) Genetic definition and sequence analysis of Arabidopsis centromeres. *Science*, **286**, 2468–2474.

**Devos, K.M., Brown, J.K. and Bennetzen, J.L.** (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in Arabidopsis. *Genome Res.* **12**, 1075–1079.

**Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. and Jones, G.H.** (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell*, **100**, 367–376.

**Ge, S., Sang, T., Lu, B.R. and Hong, D.Y.** (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl. Acad. Sci. USA*, **96**, 14400–14405.

**Gieni, R.S., Chan, G.K. and Hendzel, M.J.** (2008) Epigenetics regulate centromere formation and kinetochore function. *J. Cell. Biochem.* **104**, 2027–2039.

**Hall, A.E., Keith, K.C., Hall, S.E., Copenhaver, G.P. and Preuss, D.** (2004) The rapidly evolving field of plant centromeres. *Curr. Opin. Plant Biol.* **7**, 108–114.

**Hansen, C. and Heslop-Harrison, J.S.** (2004) Sequence and phylogenies of plant pararetroviruses, viruses, and transposable elements. *Adv. Bot. Res.* **41**, 165–193.

**Henikoff, S., Ahmad, K. and Malik, H.S.** (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098–1102.

**Heslop-Harrison, J.S., Brandes, A. and Schwarzacher, T.** (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of Arabidopsis species. *Chromosome Res.* **11**, 241–253.

 

**International Rice Genome Sequencing Project** (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

**Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y. and Takahashi, K.** (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science*, **321**, 1088–1091.

**Jackson, S.A., Wang, M.L., Goodman, H.M. and Jiang, J.** (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome*, **41**, 566–572.

**Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.S., Wing, R.A., Gill, B.S. and Ward, D.C.** (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc. Natl. Acad. Sci. USA*, **93**, 14210–14213.

**Jiang, N., Bao, Z., Temnykh, S., Cheng, Z., Jiang, J., Wing, R.A., McCouch, S.R. and Wessler, S.R.** (2002) Dasheng: a recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. *Genetics*, **161**, 1293–1305.

**Jiang, J., Birchler, J.A., Parrott, W.A. and Daw, R.K.** (2003) A molecular view of plant centromeres. *Trends Plant Sci.* **8**, 570–575.

**Jin, W., Melo, J.R., Nagaki, K., Talbert, P.B., Henikoff, S.Dawe. and R.K., Jiang.J.** (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. *Plant Cell*, **16**, 571–581.

**Kim, H., Hurwitz, B., Yu, Y. *et al.*** (2008) Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus Oryza. *Genome Biol.* **9**, R45.

**Kumekawa, N., Ohtsubo, E. and Ohtsubo, H.** (1999) Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom. *Genes Genet. Syst.* **74**, 299–307.

**Kumekawa, N., Ohmido, N., Fukui, K., Ohtsubo, E. and Ohtsubo, H.** (2001) A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. *Mol. Genet. Genomics*, **265**, 480–488.

**Lamb, J.C., Theuri, J. and Birchler, J.A.** (2004) What's in a centromere? *Genome Biol.* **5**, 239.

**Lamb, J.C., Yu, W., Han, F. and Birchler, J.A.** (2008) Plant centromeres. *Genome Dyn.* **4**, 95–107.

**Lee, H.R., Zhang, W., Langdon, T., Jin, W., Yan, H., Cheng, Z. and Jiang, J.** (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. *Proc. Natl. Acad. Sci. USA*, **102**, 11793–11798.

**Lloréns, C.F.R., Bezemer, D. and Moya, A.** (2008) The Gypsy Database (GyDB) of mobile genetic elements. *Nucleic Acids Res.* **36**, D38–D46.

**Ma, J. and Bennetzen, J.L.** (2004) Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl. Acad. Sci. USA*, **101**, 12404–12410.

**Ma, J., Devos, K.M. and Bennetzen, J.L.** (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* **14**, 860–869.

**Ma, J., Wing, R., Bennetzen, J.L. and Jackson, S.A.** (2007) Evolutionary history and positional shift of a rice centromere. *Genetics*, **177**, 1217–1220.

**McCarthy, E.M. and McDonald, J.F.** (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. *Bioinformatics*, **19**, 362–367.

**Miller, J.T., Dong, F., Jackson, S.A., Song, J. and Jiang, J.** (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics*, **150**, 1615–1623.

**Nagaki, K., Song, J., Stupar, R.M. *et al.*** (2003) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics*, **163**, 759–770.

**Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R. and Jiang, J.** (2004) Sequencing of a rice centromere uncovers active genes. *Nat. Genet.* **36**, 138–145.

**Nagaki, K., Neumann, P., Zhang, D., Ouyang, S., Buell, C.R., Cheng, Z. and Jiang, J.** (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. *Mol. Biol. Evol.* **22**, 845–855.

**Noma, K., Nakajima, R., Ohtsubo, H. and Ohtsubo, E.** (1997) RIRE1, a retrotransposon from wild rice Oryza australiensis. *Genes Genet. Syst.* **72**, 131–140.

**Presting, G.G., Malysheva, L., Fuchs, J. and Schubert, I.** (1998) TY3/GYPSY retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J.* **16**, 721–728.

**Roulin, A., Piegu, B., Wing, R.A. and Panaud, O.** (2008) Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE 1 within the genus Oryza. *Plant J.* **53**, 950–959.

**SanMiguel, P., Tikhonov, A., Jin, Y.-K. *et al.*** (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, **274**, 765–768.

**Sharma, A. and Presting, G.G.** (2008) Centromeric retrotransposon lineages predate the maize/rice divergence and differ in abundance and activity. *Mol. Genet. Genomics*, **279**, 133–147.

**Shirasu, K., Schulman, A.H., Lahaye, T. and Schulze-Lefert, P.** (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res.* **10**, 908–915.

**Tamura, K., Dudley, J., Nei, M. and Kumar, S.** (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

**Vitte, C. and Panaud, O.** (2003) Formation of Solo-LTRs Through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice Oryza sativa L. *Mol. Biol. Evol.* **20**, 528–540.

**Vitte, C., Panaud, O. and Quesneville, H.** (2007) LTR retrotransposons in rice (Oryza sativa, L.): recent burst amplifications followed by rapid DNA loss. *BMC Genomics*, **8**, 218–232.

**Walling, J.G., Pires, J.C. and Jackson, S.A.** (2005) Preparation of samples for comparative studies of plant chromosomes using in situ hybridization methods. *Methods Enzymol.* **385**, 443–460.

**Wang, Y., Tang, X., Cheng, Z., Mueller, L., Giovannoni, J. and Tanksley, S.D.** (2006) Euchromatin and pericentromeric heterochromatin: comparative composition in the Tomato genome. *Genetics*, **172**, 2529–2540.

**Wicker, T., Sabot, F., Hua-Van, A. *et al.*** (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982.

**Wu, J., Yamagata, H., Hayashi-Tsugane, M. *et al.*** (2004) Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell*, **16**, 967–976.

**Wu, J., Fujisawa, M., Tian, Z. *et al.*** (2009) Comparative analysis of complete orthologous centromeres from two subspecies of rice reveals rapid variation of centromere organization and structure. *Plant J.* doi:10.1111/ j.1365-313X.2009.04002.x.

**Xiong, Y. and Eickbush, T.H.** (1990) Origin and evolution of retroelements based upon their reverse-transcriptase sequences. *EMBO J.* **9**, 3353–3362.

**Zhang, Y., Huang, Y., Zhang, L. *et al.*** (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res.* **32**, 2023–2030.

**Zhang, W.L., Lee, H.R., Koo, D.H. and Jiang, J.** (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell*, **20**, 25–34.

**Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J. and Dawe, R.K.** (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell*, **14**, 2825–2836.