The Plant Journal (2007) **52**, 342–351 doi: 10.1111/j.1365-313X.2007.03242.x

Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*

Jetty S. S. Ammiraju[1,†], Andrea Zuccolo[1,†], Yeisoo Yu[1], Xiang Song[1], Benoit Piegu[2], Frederic Chevalier[2], Jason G. Walling[3,‡], Jianxin Ma[3], Jayson Talag[1], Darshan S. Brar[4], Phillip J. SanMiguel[5], Ning Jiang[6], Scott A. Jackson[3], Olivier Panaud[2] and Rod A. Wing[1,*]

[1]*Arizona Genomics Institute, Department of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA,*
[2]*Laboratoire Génome et Développement des Plantes, Université de Perpignan, Perpignan, France,*
[3]*Department of Agronomy, Purdue University, West Lafayette, IN 47907-2054, USA,*
[4]*Department of Plant Breeding and Genetics, International Rice Research Institute, Los Baños, The Philippines,*
[5]*Genomics Core Facility, Purdue University, West Lafayette, IN 47907-2054, USA*
[6]*Plant and Soil Sciences, Michigan State University, East Lansing, MI 48823, USA*

Received 30 March 2007; revised 11 June 2007; accepted 21 June 2007.
*For correspondence (fax +1 520 621 1259; e-mail rwing@ag.arizona.edu).
†These authors contributed equally to this work.
‡Present address: Department of Horticulture, University of Wisconsin, Madison, WI, USA.

Summary

Long terminal repeat (LTR) retrotransposons constitute a significant portion of most eukaryote genomes and can dramatically change genome size and organization. Although LTR retrotransposon content variation is well documented, the dynamics of genomic flux caused by their activity are poorly understood on an evolutionary time scale. This is primarily because of the lack of an experimental system composed of closely related species whose divergence times are within the limits of the ability to detect ancestrally related retrotransposons. The genus *Oryza*, with 24 species, ten genome types, different ploidy levels and over threefold genome size variation, constitutes an ideal experimental system to explore genus-level transposon dynamics. Here we present data on the discovery and characterization of an LTR retrotransposon family named *RWG* in the genus *Oryza*. Comparative analysis of transposon content (approximately 20 to 27 000 copies) and transpositional history of this family across the genus revealed a broad spectrum of independent and lineage-specific changes that have implications for the evolution of genome size and organization. In particular, we provide evidence that the basal GG genome of *Oryza* (*O. granulata*) has expanded by nearly 25% by a burst of the *RWG* lineage *Gran3* subsequent to speciation. Finally we describe the recent evolutionary origin of *Dasheng*, a large retrotransposon derivative of the *RWG* family, specifically found in the A, B and C genome lineages of *Oryza*.

Keywords: *Oryza,* long terminal repeat (LTR) retrotransposon, *RWG* family, genome size, large retrotransposon derivative.

Introduction

Plant genomes are extraordinarily dynamic. One prominent force behind this is the flux of transposable elements, especially long terminal repeat (LTR) retrotransposons, that can result in dramatic increases or decreases in genome size (Bennetzen *et al.*, 2005; Devos *et al.*, 2002; Feschotte *et al.*, 2002; Hawkins *et al.*, 2006; Ma and Bennetzen, 2004; Ma *et al.*, 2004; Meyers *et al.*, 2001; Petrov and Wendel, 2006; Piegu *et al.*, 2006; SanMiguel *et al.*, 1996, 1998; Schulman and Kalendar, 2005; Shirasu *et al.*, 2000; Uozu *et al.*, 1997; Vicient *et al.*, 1999; Vitte and Panaud, 2005; Wendel and Wessler, 2000; Wendel *et al.*, 2002; Zhang and Wessler, 2004). However, little is known about how LTR retrotransposons evolve and reshape genomes within a family structure. For example, previous

comparative studies of orthologous regions between maize, sorghum and wheat were unable to identify conserved LTR retrotransposons that pre-date their divergence (Ramakrishna *et al.*, 2002; SanMiguel *et al.*, 2002; Tikhonov *et al.*, 1999). This is mainly because of long divergence times between these species relative to the half-life of LTR retrotransposons, which is estimated to be <6 million years (MY) (Ma and Bennetzen, 2004), thus confounding the study of patterns of ancestral LTR retrotransposon evolution and their impact on genome biology.

The *Oryza* genus represents a powerful model system to test and validate the mechanisms and timing of genome size evolution within a genus (Piegu *et al.*, 2006; Wing *et al.*, 2005), having ten distinct genome types (six diploid and four polyploid), a threefold genome size variation (Ammiraju *et al.*, 2006), a robust phylogenetic tree (Ge *et al.*, 1999), a comprehensive collection of BAC libraries (Ammiraju *et al.*, 2006), BAC end sequences (BES) and physical maps (http://www.omap.org; Kim *et al.*, unpublished results), and a reference genome sequence for the cultivated species *O. sativa* (International Rice Genome Sequencing Project, 2005; Yu *et al.*, 2005). Previously, using this system, we demonstrated that there is a positive correlation between LTR retrotransposon content and genome size in *Oryza* (Ammiraju *et al.*, 2006), and that the *O. australiensis* EE genome doubled in size by three LTR retrotransposon bursts (*RIRE1*, *Wallabi* and *Kangaroo*) approximately 3 million years ago (MYA) (Piegu *et al.*, 2006).

Here we present the discovery and characterization of a second case of genomic obesity in *Oryza* caused by a *Ty-3-gypsy* retrotransposon family named *Gran3* that populates nearly one quarter of the *O. granulata* [GG] genome. Most compellingly, the 22% increase in genome size of this most ancestral lineage of the genus (Ge *et al.*, 1999) occurred subsequent to its speciation. These results are significant in light of previous predictions of unidirectional genome size reduction in the genus *Oryza*, resulting from super-imposition of genome sizes onto known phylogenies (Kellogg, 1998). *Gran3* is closely related to two other LTR retrotransposon families previously characterized in *O. sativa* [AA] and *O. australiensis* [EE], i.e. *RIRE2* and *Wallabi*, respectively. Comparative analysis of the transpositional history of these three elements (*Gran3*, *RIRE2* and *Wallabi*) across the genus *Oryza* allowed, for the first time, the study of evolutionary dynamics of the same LTR retrotransposon family across a plant genus covering a time span of 25 MY (Piegu *et al.*, 2006). This analysis revealed a broad spectrum of independent and lineage-specific changes that have implications on the evolution of genome size in *Oryza* as a whole. Finally we analyzed the evolutionary dynamics of the element *Dasheng*, a large retrotransposon derivative (LARD) (Kalendar *et al.*, 2004) of the retrotransposon *RIRE2* family, thus providing data and insights on the evolution of a non-autonomous LTR retrotransposon.

Results and discussion

Isolation and characterization of the Gran3 *LTR retrotransposon family from the* O. granulata *[GG] genome*

As part of a large-scale comparative sequencing project within the genus *Oryza* (http://www.omap.org), a BAC clone (149.7 kb) containing the *Adh1–Adh2* region from the *O. granulata* [GG] genome was isolated, sequenced and annotated. *O. granulata* is the most basal lineage of the genus *Oryza* (Ge *et al.*, 1999), and possess the second largest genome (882 Mbp) among the diploid species (Ammiraju *et al.*, 2006). Four *Ty-3-gypsy* retrotransposons and a relic in the form of a solitary LTR were identified from this single BAC and found to belong to the same family, henceforth named *Gran3*. They displayed typical LTR retrotransposon consensus 5′-TG…CA-3′ sequences at the start and end of the LTRs, and intact target site duplications. Two of the elements, *Gran1* and *Gran2*, contained deletions in their internal coding domains relative to the structurally complete element *Gran3* (Supplementary Figure S1). These two elements are therefore considered defective, and were either trans-mobilized by an autonomous *Gran3* element or have degraded or been rearranged after transposition. Another member of the same family, *Gran4*, was considerably rearranged (Figure 1a). *Gran1* and *Gran2* are 5003 and 4620 bp long, respectively, whereas *Gran3* element is 12 098 bp after removal of two unrelated solo LTRs located within this element (Figure 1a). In addition to sharing very similar and short LTR lengths of 425–431 bp, *Gran1–3* possessed highly conserved primer binding sites (PBS) and polypurine tract (PPT) sites adjacent to their 5′ and 3′ LTRs (Figure 1b).

Nearly one quarter of the O. granulata *genome is composed of the* Gran3 *family*

The presence of four members on a single BAC strongly suggested that *Gran3* family is highly abundant in the *O. granulata* genome. To evaluate this hypothesis, the copy number of *Gran3* family in the *O. granulata* genome was estimated using three approaches. We first calculated the copy number mathematically [LTR copy number = matches in BAC ends $\times$ 882 Mb (*O. granulata* genome size)/ the size of the BAC end database (93 Mb)] (Jiang *et al.* (2002a) and obtained an estimate of 59 556 LTRs. Secondly, probes specific to the LTR domain and an internal domain common to both complete and defective *Gran3* elements (Figure 1b) were hybridized to dot blots containing stochiometrically estimated concentrations of the *O. granulata* genome. This analysis identified nearly 25 700 *Gran3* copies with both LTRs (based on the internal domain) and 8600 $\pm$ 2000 apparent single LTRs (Table 1). Lastly, we validated the copy number estimates by screening the *O. granulata* BAC library (Ammiraju *et al.*, 2006), and showed that about 60–80% of

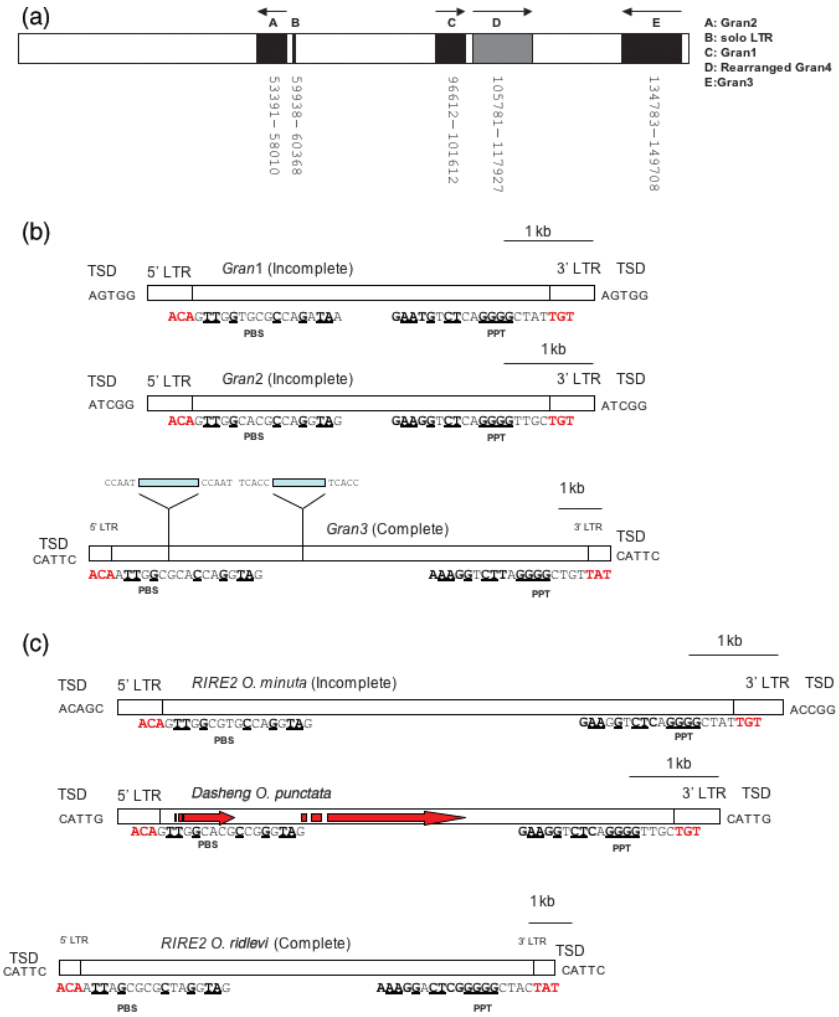

Figure 1. Organization and structural features of *Gran3* retrotransposon family in *O. granulata* BAC.
(a) Distribution of *Gran3* elements in the *Oryza granulata* BAC. Arrows indicate the 5′ → 3′ orientation. The numbers indicate the position of the elements in the BAC OG-Aba077F15 (149.7 kb long).
(b) Detailed structures of the *Gran3* elements isolated. TSD, target site duplication; PBS, primer binding site; PPT, polypurine tract. Proposed PBS and PPT sequences are shown and their shared nucleotide positions are underlined; sequences belonging to LTRs are shown in red. The light-green boxes in *Gran3* represent two solo LTRs insertions flanked by TSD.
(c) Isolated full-length *RWG* family members from various *Oryza* lineages. The arrows indicate tandem repeats.

Table 1 Copy numbers and relative contribution of the *Gran3* family in the *Oryza granulata* genome

Gran3 elements	Number of copies	Size (bp)	Size in genome (Mb)	Genome fraction (%)
Intact elements	8900 ± 1000	12 000	107 ± 12	12.1 ± 1.4
Defective forms	16 800 ± 3000	5000	84 ± 15	9.5 ± 1.7
Apparent solo LTRs	8600 ± 2000	430	3.6 ± 0.9	0.4 ± 0.1
Total			194.6 ± 24.9	22.0 ± 3.2

the tenfold redundancy BAC library contained sequences homologous to the *Gran3* family (data not shown).

These results indicated that the *Gran3* LTR retrotransposon family is highly abundant in the *O. granulata* genome, but could not distinguish the relative proportion of intact versus defective elements. Evidence was gathered from two different analyses to understand such dynamics. Firstly, an additional probe specific to the reverse transcriptase (RT) domain of *Gran3* element that distinguished the defective forms from complete forms was designed (Figure 1b) and hybridized to dot blots. Copy number estimates indicated the presence of approximately 8900 ± 1000 complete *Gran3* elements. Subtraction of this number from the entire number of *Gran3* family members indicated that nearly 16 800 ± 3000 defective forms with internal deletions exist in the GG genome, accounting for up to 22% of the genome size (Table 1). Secondly, an *in silico* survey of the BES database of *O. granulata* was also conducted using the RT domain of *Gran3*, and the results were compared to those obtained using LTR sequences. The results indicated a striking imbalance of RT to LTR domains, with a ratio 0.08, which is far from that expected (0.5) if all the *Gran3* elements are complete (Supplementary Table S1). These results point to an overwhelming presence of incomplete/non-autonomous elements, and are consistent with the results obtained from hybridization approaches and the annotated *O. granulata* BAC. Similar results have been found for other LTR retrotransposon families, such as *Cinful-1* and *Zeon* in maize (Meyers *et al.*, 2001). Further, a considerable portion of the total number of LTR sequences appears to be solo LTRs

(25%). However, the contribution of these solo LTRs is only 0.4 ± 0.1% of the genome (Table 1).

Genomic organization of the Gran3 *family in the* O. granulata *genome*

The copy number estimates above indicate that the *Gran3* retrotransposon family is highly abundant in the GG genome. To better understand the genomic organization of the *Gran3* family, we hybridized a fluorescently labeled *Gran1* probe (see Experimental procedures) to *O. granulata* metaphase chromosomes. As shown in Supplementary Figure S4, a significant clustering of *Gran3* in the heterochromatic peri-centromeric regions is particularly striking for most chromosomes. Additionally, we also mapped *Gran3* family members using BES to a phase I BAC-based physical map of the *O. granulata* genome. The results are presented in Supplementary Table S2 and Supplementary Figure S5, and are congruent with the results of fluorescent *in situ* hybridization. The abundance of *Gran3* appeared to be non-uniform, with the highest accumulation frequency (portion/presence; LTRs/megabase) found on chromosome 8 and the lowest on chromosome 10 in the peri-centromeric regions. Ratios of accumulation frequencies in the peri-centromeric regions versus non-pericentromeric regions were derived for each chromosome (Supplementary Table S2). A ratio above 1 is expected in the case of peri-centromeric clustering. The results indicated that there was significant peri-centromeric clustering on all chromosomes except chromosome 10. Interestingly, the frequency of *Gran3* accumulation (LTR/Mb) is approximately five times higher in peri-centromeric regions than in chromosome arms.

Our data indicate that the *Gran3* family is a major component of peri-centromeric regions in the *O. granulata* genome. This type of organization is similar to that previously found for the LTR retrotransposons *Dasheng* and *RIRE2* in *O. sativa* (Jiang *et al.*, 2002a,b).

Genomic paleontology of the Gran3 *family in* O. granulata

A major retrotranspositional burst underlies Gran3 *abundance in* O. granulata*.* We estimated the timing of insertional activity of *Gran3* using an approach described by Piegu *et al.* (2006). Although, such global simulations may tend to over-estimate the timing, they should nevertheless provide insights into the approximate time frame for such activity. In brief, we first determined the population structure of the *Gran3* retrotransposon family within the *O. granulata* genome using a phylogenetic approach. The resulting neighbor-joining tree, generated using all 150 paralogous RT sequences (Supplementary Table S1) retrieved from the BES database of *O. granulata* (methods) has three major monophyletic clades, namely M1 (which includes 92 paralogous RT), M2 (which includes 22 paralagous RT) and M3 (which includes 13 paralagous RT), plus a fourth group (which includes 23 paralogous RT) that is not clearly monophyletic and is very basal to the first three groups (Figure 2).

Secondly, pairwise distances (Kimura, 1980) were calculated by comparing each paralog to all other paralogs belonging to the same monophyletic group. Distances were then translated into insertion dates as described by San-Miguel *et al.* (1996) but using a mutation rate of 1.95×10^{-8} (Vitte *et al.*, 2004) to be consistent with our previous analysis (Piegu *et al.*, 2006). This enabled the identification of three retrotranspositional waves, mostly overlapping in terms of time (Figure 3). From a quantitative perspective, a significant part of the amplification seems to have occurred at around 3–6.2 MYA. Further, a 4.4 MY insertion time was estimated for complete and intact *Gran3* element, which is consistent with the dating of the proliferation waves.

The Gran3 LTR retrotransposon burst is more recent than host species radiation. Two lines of evidence suggest that the *Gran3* burst in the GG genome occurred subsequent to the species radiation. Firstly, the estimated time of divergence between *O. granulata* and other *Oryza* species is approximately 9–25 MY (Guo and Ge, 2005; Piegu *et al.*, 2006). On the other hand, the proliferation waves were dated at <6.2 MY (Figure 3), suggesting that the genome size increase in *O. granulata* is recent. Secondly, all *Gran3* RT paralogs from *O. granulata* form a distinct cluster relative to the other *Oryza* species at high bootstrap values (Figure 4). A strikingly similar scenario was observed in *O. australiensis* [EE], echoing our previous observations (Piegu *et al.*, 2006). Combined, these results provide an intriguing picture of recent and independent increases in genome sizes for the GG and EE genome types of *Oryza* because of the same ancestral retrotransposon family. As in the case of *O. australiensis*, the increase in the genome size was subsequent to its speciation. Further, phenetic analysis suggests that the episodic burst dated at around the same evolutionary time in *O. australiensis* (Piegu *et al.*, 2006) was independent of the bursts in *O. granulata*.

Indirect evidence for the ongoing activity of *Gran3* was detected in the *O. granulata* genome. *Gran3* LTR sequences identified from the BES database were compared to each other in all pairwise combinations and then classified into groups according to the extent of sequence divergence (Supplementary Figure S3). Identical LTRs that could potentially result from cloning bias, i.e derived from the same restriction site cloned during the BAC library construction, were removed by careful manual comparison. A very conservative estimate indicated that more than 9% of the LTRs (142 out of 1559 LTRs analyzed) have at least one identical counterpart within the same group. As LTRs of an intact element are identical at the time of insertion, these

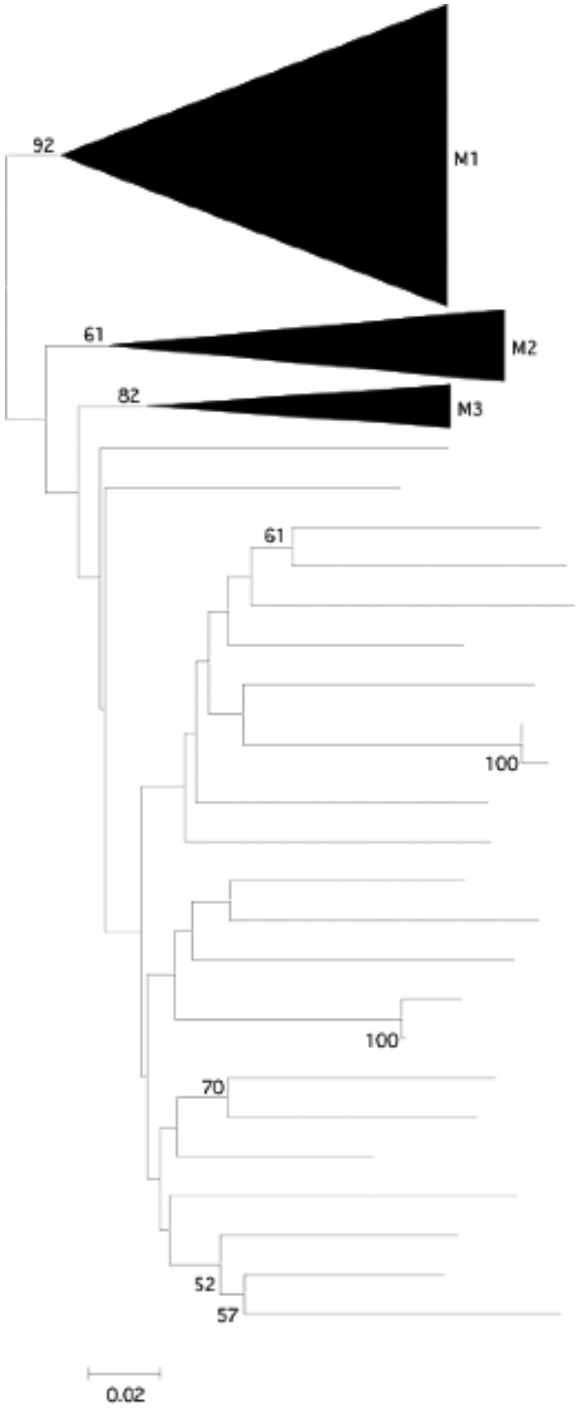

Figure 2. Neighbor-joining phylogenetic tree of the *Gran3* retrotransposon in *Oryza granulata*.
Reverse transcriptase coding domain nucleotide sequences were used. Bootstrap values are calculated for 1000 replicates, and are shown only if greater than 50. Three monophyletic clades are indicated (M1, M2 and M3).

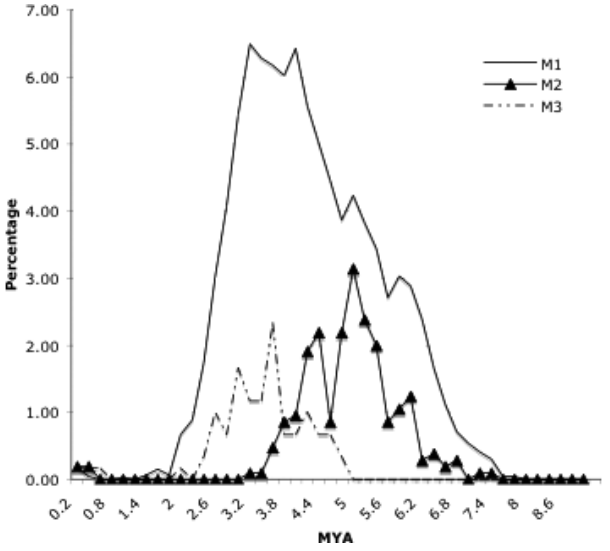

Figure 3. Timing of the *Gran3* retrotranspositional burst in *Oryza granulata.*
Groups of paralogs were isolated on the basis of the monophyletic clades indicated in Figure 2. Each paralog was compared to all others included in the same clade; Kimura distances were then converted to million of years ago using a substitution rate of 1.95×10^{-8}. The *y* axis shows the percentage of the total pairwise comparisons.

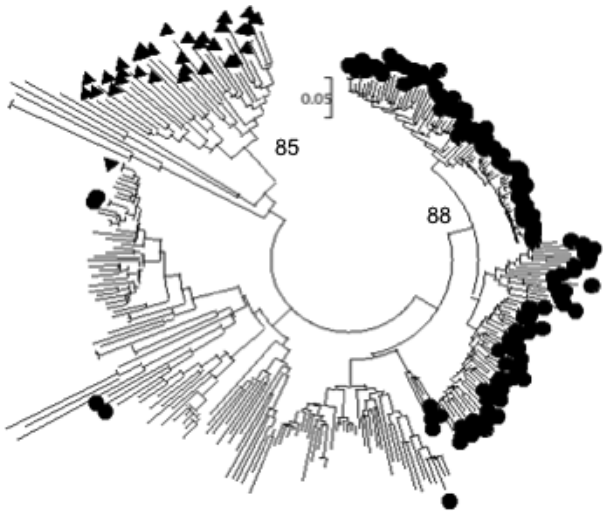

Figure 4. Neighbor-joining phylogenetic tree of the *Gran3* retrotransposon from 12 *Oryza* species.
Gran3 reverse transcriptase (RT) coding domain nucleotidic sequences were used. Of all the available *Gran3* RT (Supplementary Table S1), 20% were randomly chosen for the analysis for each *Oryza* species. Bootstrap values were calculated for 1000 replicates. Circles indicate *O. australiensis* sequences and triangles indicate *O. granulata* sequences.

observations may imply ongoing *Gran3* activity or such activity in the recent past. However, the overall majority of LTRs showed only partial similarity with others, reflecting the heterogeneity of the LTR set as expected in the case of an ancient retrotranspositional process. Solo LTRs with terminal site duplications (TSDs) (Supplementary Figure S3) were present in most of these classes, suggesting that the process leading to their formation has been active since the

very beginning of the *Gran3* colonization of the *O. granulata* genome.

RIRE2, Wallabi *and* Gran3 *are all members of the same LTR retrotransposon family*

To determine whether the *Gran3* family isolated from *O. granulata* was related to other LTR retrotransposons in *Oryza*, similarity searches were conducted. This analysis revealed that *Gran3* family members were similar to two previously characterized LTR retrotransposons, *RIRE2* from *O. sativa* (Ohtsubo *et al.*, 1999) and *Wallabi* from *O. australiensis* [EE] (Piegu *et al.*, 2006). *Gran3* was 55% similar over the entire length of *RIRE2*, 55.9% in the LTRs and 71% in the RT domain at the nucleotide level. In addition, they shared remarkable conservation in sequence length (11 365 bp for *RIRE2* versus 12 098 bp for *Gran3*), and a PBS sequence that is homologous to $tRNA^{Arg}$, thereby suggesting that *Gran3* is a distant member of the *RIRE2* clade of LTR retrotransposons. For *Wallabi*, *Gran3* was 53.8% similar over the entire length, 57.4% in the LTRs and 72.1% in the RT regions.

Phylogenetic analysis of RT sequences of *Gran3* (*O. granulata*) and *Wallabi* (*O. australiensis*) with those belonging to the most abundant *Ty-3-gypsy* elements from *O. sativa* resulted in a distinct grouping of these two elements with that of *RIRE2* at high bootstrap value (Supplementary Figure S2). Based on sequence and structural conservation as well as phylogenetic grouping, we propose that *RIRE2*, *Wallabi* and *Gran3* belong to the same LTR retrotransposon family, henceforth named *RWG* (*RIRE2*, *Wallabi*, *Gran3*).

Several lines of evidence also suggested that the *RWG* LTR retrotransposon family was ancient in the genus *Oryza*. Using Southern blots, we previously demonstrated the presence and differential abundance of the *Wallabi* element (now a member of the *RWG* family) across all 10 genome types of *Oryza* (Piegu *et al.*, 2006). These results were further confirmed by the presence of RT homologs (see Experimental procedures) of this family in BES derived from all 10 known *Oryza* genome types (http://www.omap.org) and by isolation of complete *RWG* family elements from several sequenced BACs of *O. punctata* [BB], *O. minuta* [BBCC] and *O. ridleyi* [HHJJ] (Figure 1c; GenBank accession numbers EF193783–EF193785). In addition, a detailed phylogenetic analysis of RT domains derived from the *Oryza* BES databases revealed a coherence between element phylogeny and organismal phylogeny, suggesting primary colonization by the *RWG* family in the ancestor of *Oryza* (Figures 4 and 5) and subsequent diversification during the course of evolution. The closest homologs for the *RWG* family outside *Oryza* are the *Athila* and *Grande-1* elements in Arabidopsis and *Zea mays*, respectively (Ohtsubo *et al.*, 1999).

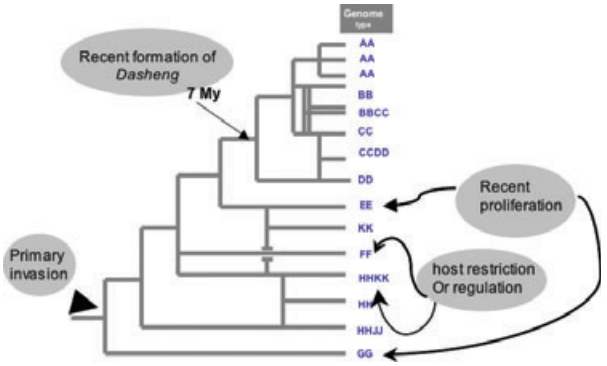

Figure 5. Evolutionary scenario of the *RWG* family and its impact on the evolution of genome size. (Phylogenetic tree is from Ge *et al.*, 1999.)

Discontinuous phylogenetic distribution of the RWG *LTR retrotransposon family in the genus* Oryza

Dot-blot analysis was conducted using probes specific to LTR and internal domains to estimate the relative abundance of the *RWG* family in *Oryza* and its contribution to genome size variation. Different steps were taken to minimize biases inherent in the use of dot-blot hybridization approaches to estimate copy numbers. Firstly, we designed specific primers from the *RWG* family sequences (LTR and RT) mined from BES of each *Oryza* genome type (see Experimental procedures), and polymerase chain reaction (PCR)-amplified probes were used to hybridize dot blots of the same species to minimize biases from sequence divergence. Secondly, hybridization conditions were calibrated by first estimating *RWG* copy number from *O. sativa* genomic DNA dot blots and then comparing the results to those obtained from *in silico* analysis of the completed reference sequence (International Rice Genome Sequencing Project, 2005). Thirdly, the results obtained from hybridization analysis were largely supported by copy number estimates from a set of an unbiased sampling of random sheared shotgun sequences generated for each *Oryza* species (Supplementary Table S3) and the number of significant hits from BES analysis (Supplementary Table S1). The results are presented in Table 2. Copy number variation of an unprecedented magnitude was found across *Oryza*, from approximately 20 in *O. brachyantha* [FF] to 27 000 in *O. australiensis* [EE] (Piegu *et al.*, 2006). Two diploid species, *O. brachyantha* [FF] and *O. glaberrima* [AA], and the tetraploid *O. coarctata* [HHKK] contained the lowest number of *Gran3* members. The copy number in the case of *O. coarctata* should be considered a rough estimate as the genome size of this species is unknown (Ammiraju *et al.*, 2006). Two diploid genome types (GG and EE) contained the largest number of *RWG* family members among the ten known genome types (Table 2). Other genome types contained moderately increased numbers of *RWG* family elements in comparison to the International Rice Genome Sequencing

Table 2 Copy number and genomic contribution of the *RWG* family across *Oryza*

Species	Genome type	Genome size (Mb)	Copy number		Percentage genome share			
			LTR	Internal[a]	By intact elements	By apparent solo LTR	Total	LTR/in
O. nivara	AA	448	3240 ± 645	512 ± 130	1.37	0.2	1.6	4–10.2
O. rufipogon	AA	439	1620 ± 321	209 ± 37	0.57	0.1	0.7	5.3–11.3
O. glaberrima	AA	357	207 ± 35	30 ± 7	0.10	0.17	0.3	4.6–10.8
O. punctata	BB	425	2291 ± 469	188 ± 32	0.53	0.2	0.7	8.3–17.7
O. officinalis	CC	651	5946 ± 573	762 ± 113	1.40	0.3	1.7	6.1–10
O. minuta	BBCC	1124	2850 ± 533	407 ± 113	0.43	0.1	0.5	4.5–11.5
O. alta	CCDD	1008	1327 ± 200	759 ± 222	0.90	0.0	0.9	1.1–2.8
O. australiensis[b]	EE	965	66 000 ± 1000	27 000 ± 3000	24.9	0.6	25.5	2.2–2.8
O. brachyantha	FF	362	< 20	< 20	0.03*	0.0	0.03*	NA
O. granulata[b]	GG	882	60 000 ± 3000	25 700 ± 2000	21.6	0.4	22.0	2.1–2.7
O. ridleyi	HHJJ	1283	1033 ± 257	469 ± 155	0.44	0.004	0.45	1.2–4.1
O. coarctata	HHKK	1283	< 20	< 20	0.01*	0.0	0.01*	NA

[a]Internal values represent the copy numbers obtained from hybridization of the reverse transcriptase domain of *RWG* elements isolated from BAC end sequences of each *Oryza* species, except in the case of *O. granulata* and *O. australiensis.*
[b]For *O. granulata*, internal values represent copy numbers obtained using probes that are common for both complete and defective *Gran3* elements (Table 1). *O. australinesis* values were obtained from Piegu *et al.* (2006).
NA, not available; asterisks indicate that the value is a rough estimate

Project reference sequence, suggesting differential lineage-specific proliferation and/or elimination or regulation. We estimated the ratio of the number of LTRs versus the internal region (LTR/in) for the *RWG* family for each *Oryza* genome type. A ratio of 2 is expected in the case of an intact element, and deviations from this could result from deletion or recombination. The LTR/in ratios for these species suggest a differential excess of LTRs in all genome types except *O. coarctata* and *O. brachyantha* (Table 2). The contributions of the *RWG* family varied widely across the genus, ranging from 0.02% in *O. coarctata* to 25% in *O. australiensis*.

Recent evolution of a non-autonomous element of the RWG *family*

Previous studies (Jiang *et al.*, 2002a,b) showed that the *O. sativa* [AA] genome harbors a high copy number of a large non-autonomous retrotransposon derivative (LARD) (Kalendar *et al.*, 2004) called *Dasheng*. Its autonomous counterpart is the LTR retrotransposon *RIRE2*, based on sequence similarity of the LTRs of both elements, target site preference and genomic distribution. Two alternative hypotheses have been proposed to account for the origin of LARDs (Jiang *et al.*, 2002a). One is that they are the product of transduction of a genomic sequence from the host genome, flanked by two solo LTRs. The other is that LARDs originate in the virus-like particle by co-encapsulation of an mRNA of the autonomous element with an mRNA of any host gene, followed by strand exchange between the two during the reverse transcription step. So far, there is no experimental evidence to support either of these hypotheses. In this study, we have provided evidence that the *RWG* family is ancient in the genus *Oryza*. We thus looked for the presence of *Dasheng*-related sequences in the genus to tentatively trace its origin by probing Southern blots containing genomic DNA from representatives of all ten *Oryza* genome types using *Dasheng*-specific probes. The results clearly indicate the presence of strong hybridization signals in *Oryza* lineages belonging to the AA, BB and CC genomes, and their absence from other distantly related genomes (Supplementary Figure S6). The Southern blot results were further confirmed by PCR using a primer in the 5′ LTR of *Dasheng* and another in its internal region, for which PCR amplification bands were only detected in the AA, BB and CC genome species (data not shown). These results indicate that retrotransposition of *Dasheng* has occurred recently and specifically in the A/B/C genome lineages, i.e. around 7 MYA (Figure 5) (Panaud *et al.*, unpublished results).

Dynamics of the RWG *family and its impact on the evolution of genome size in* Oryza

Here we have presented a comparative evolutionary dynamics study of an ancestral LTR retrotransposon family (*RWG*) in terms of its recent history, timing of transposition and differential fates of proliferation among lineages of the genus *Oryza*. Recent activity of *Dasheng* and *RIRE2* has already been demonstrated in domesticated rice (*O. sativa* [AA]) by estimating their age as well by identifying transcripts in cDNA libraries (Jiang *et al.*, 2002a,b). Nearly 25% of all full-length *Dasheng* and *RIRE2* elements isolated from *O. sativa* were amplified within the last 150 000 years (Jiang *et al.*, 2002a,b). A global estimate

of their insertion time derived by dating phenetically structured groups of *Dasheng*- and *RIRE2*-like LTR sequences indicated *Dasheng* to be one of the youngest elements (7.2 MY) and *RIRE2* to be relatively ancestral (17.2 MY) (Jiang *et al.*, 2002b). However, the substitution rate used for molecular clock estimation is at least three times slower than that used in our analysis. Extrapolation of the previous results of Jiang *et al.* (2002b) to the mutation rate we used suggests that most *RIRE2*- and *Dasheng*-like elements were inserted within the past 5–6 MY in the *O. sativa* genome. These results indicate that, as in the case of the distant ancestors *O. australiensis* and *O. granulata*, the *O. sativa* genome also experienced recent amplification of the *RWG* family (Figure 5). The magnitude of such amplification, however, is several thousand-fold different among these lineages, leading to massive increases in genome size for both the EE and GG genomes. The contribution of this single LTR retrotransposon family in these two genomes parallels the total contribution of all classes of retrotransposons in the *O. sativa* genome (20%) (International Rice Genome Sequencing Project, 2005). These results suggest that a large proportion of the more than twofold genome size difference in the domesticated species versus the most ancestral lineage is mainly accounted for by differential amplification of the *RWG* family.

We also demonstrated that the *RWG* family propagates to different levels, resulting in a number of lineages with complex phylogenetic patterns. The findings raise more questions regarding the biological causes for massive *RWG* family proliferation in some lineages (i.e. *O. granulata* and *O. australiensis*) and almost sudden death or attenuation in others (i.e. *O. brachyantha* and *O. coarctata*) (Figure 5). Such differential fates of the *RWG* family could mechanistically result from lineage-specific differences in the regulation and/or suppression of transposition. Alternatively, they could also result from differences in the rates and efficiencies of the removal processes. One indirect line of evidence, a lack of correlation between transposition and transcription of this family in *O. sativa*, *O. australiensis* and *O. granulata*, suggests that differential amplification of the *RWG* family is mainly regulatory, at transcriptional or post-transcriptional level.

Beyond this mechanistic information, surprisingly little is known about the global evolutionary forces underlying genome size variation. Several models for the evolution of genome size have been proposed, ranging from the role of adaptive changes to a particular ecological niche (Kalendar *et al.*, 2000) to selection at the organismal level (Kidwell and Lisch, 2000). With large-scale sequences from several *Oryza* species expected in the near future, the stage is set for exploring such exciting avenues. The results presented here bring clarity to the bi-directional evolution of genome size and show that the *RWG* family has played a major role in restructuring and evolution of genome size in *Oryza*.

Experimental procedures

Data mining and phylogenetic analysis

Similarity searches were carried out using the BLASTN algorithm (Altschul *et al.*, 1997); only hits with an E value ≤ 1e-10 and spanning at least 80 bp of the *Gran3* elements were taken into account. All alignments were performed using the program CLUSTAL W (Thompson *et al.*, 1994). Sequences homologous to the RT domain of *Gran3* elements were identified in the BES databases through BLASTN searches (Altschul *et al.*, 1997) run under relaxed settings (-q -4 –r 5) in order to isolate the most diverged elements of the family; only sequences spanning at least 90% of the length of the RT domain were considered. The specificity and the origin of these sequences were checked by comparison with RTs from the most abundant LTR retrotransposons of *O. sativa*. Only those having the best scores of similarity with *Gran3* elements were retained and used to build neighbor-joining trees using the program MEGA version 3 (Kumar *et al.*, 2004).

A similar approach was used to identify *Gran3* LTRs from *O. granulata* BES. Alignments of LTR pairs were performed using the program STRETCHER (EMBOSS package; Rice *et al.*, 2000). Kimura distances (Kimura, 1980) were calculated for each pair using the DISTMAT program (EMBOSS package).

Hybridization analysis

Total genomic DNA from young leaf tissue of the *Oryza* species *O. brachyantha* ([FF]; accession 101232), *O. alta* ([CCDD]; accession 105143), *O. officinalis* ([CC]; accession 100896), *O. ridleyi* ([HHJJ]; accession 100821), *O. punctata* ([BB]; accession 105690), *O. coarctata* ([HHKK]; accession 104502), *O. minuta* ([BBCC]; accession 101141), *O. granulata* ([GG]; accession 102118), *O. glaberrima* variety CG14 ([AA]; accession 96717), *O. rufipogon* ([AA]; accession 105491), *O. nivara* ([AA]; accession 100897) and *O. australiensis* ([EE]; accession 100882) and *O. sativa* ssp. *japonica* (cv. Nipponbare; [AA]) was isolated as described by Doyle and Doyle (1987). Large-scale genomic resources in the form of BAC libraries, BES and physical maps (Ammiraju *et al.*, 2006; Wing *et al.*, 2005; Kim *et al.*, unpublished results) are available for these *Oryza* accessions as part of the OMAP consortium (http://www.omap.org).

LTR and the internal domain (1817–2128 bp) of *Gran2* (Genbank accession number EF187427) and the RT domain (3993–4430 bp) of *Gran3* (Genbank accession number EF187429) were used as BLAST probes to retrieve paralogs from BES databases of all *Oryza* species. Retrieved paralogs were aligned, and specific PCR primers were designed to amplify these domains for each species to reduce the copy number biases inherent with heterologous probes (Supplementary Table S4). All PCR primers were designed using PRIMER3 software (Rozen and Skaletsky, 2000). PCR reactions were carried out in a 25 μl reaction volume, using 10 ng of genomic DNA, 0.2 mM of each dNTP, and 10 pmol of each primer. The following thermal cycling conditions were used; 94°C for 4 min, then 35 cycles of 94°C for 30 sec, 65°C for 30 sec and 72°C for 90 sec, followed by 72°C for 5 min.

Polymerase chain reaction products were purified from the agarose gels. Genomic DNA from *Oryza* species as well as the isolated PCR products was quantified using a Nano Drop® ND-1000 spectrophotometer (Nano Drop; http://www.nanodrop.com). Dot-blot preparation, hybridization, washing conditions and copy number estimations were essentially as described by Piegu *et al.* (2006). At least three replications were carried out for each species. High-density colony filters of the *O. granulata* library were screened

using LTR and internal domain probes. Hybridizations and imaging were carried out as described by Ammiraju *et al.* (2006).

Two probes corresponding to the internal region (IR) of *Dasheng* were generated. Primer pair 1 (5′-TATTATGCCACCGTTGTTGG-3′/5′-AACAACGACGGCATAATAGG-3′) was designed to span the tandemely repeated sequence within the IR. Primer pair 2 (5′-TGGGAAACTTCGTCATCATAGTC-3′/5-GTGCTTAAGTCCCTAGAAGACAC-3′) was designed to amplify a portion of the IR that did not contain any repeat motif. PCR amplification with primer pair 1 resulted in a ladder-like pattern on agarose gel. The PCR products were cloned, and the inserts obtained from the recombinant clones were of various sizes, as expected. The largest clone was sequenced to confirm the presence of five tandem repeats, and was then used in Southern hybridization experiments. PCR amplification with primer pair 2 yielded a single band, which was cloned, sequenced and then used as a hybridization probe.

Fluorescent in situ *hybridization analysis*

The *Gran1* element was sub-cloned from the *O. granulata* BAC clone (OG_ABa077 F15) using a PCR strategy. PCR amplification was performed in a final volume of 50 μl containing 0.4 μM of primers (CTGCGACAGGCAGGGTAA and TGGGTTCTCTAGACCTACATTATCGT), 2.5 U of Klen Taq DNA polymerase (Sigma, http://www.sigmaaldrich.com/), 5 μl of 10 × PCR buffer (Sigma), 0.12 5 mM of each dNTP, 1.5 mM $MgSO_4$, 5 μl of PCRx enhancer solution (Invitrogen, http://www.invitrogen.com/), 3 μl of DMSO and 50 ng of BAC DNA. PCR amplification was performed in a Tetrad thermocycler (Bio-Rad, http://www.bio-rad.com/) with 35 cycles of 96°C for 15 sec and 60°C for 15 min, and checked by 1% agarose gel electrophoresis with 1 × TAE buffer. The gel-purified insert was cloned into pGEM-T Easy vector (Promega, http://www.promega.com/) according to the manufacturer's recommendation, and the recombinant clones were validated by *Not*I restriction enzyme digestion and sequencing using T7 (TA ATACGACTCACTATAGGG) and Sp6 (ATTTAGGTGACACTAT AGAA) primers. Chromosomes of *O. granulata* were prepared according to the method described by Walling *et al.* (2005). Probe labeling, fluorescent *in situ* hybridization and image analysis were carried out as described by Walling *et al.* (2006).

Physical mapping

Finger printed contigs (FPC) for the BAC clones that span *Gran3* LTRs in their ends were identified in the phase I physical map of *O. granulata* (http://www.omap.org). The distance between two ends of each BAC clone was measured in CB units. The CB distances were then translated into physical distances based on the total length of the physical contig and also from anchoring information of these contigs to the *O. sativa* reference genome sequence using SYMAP software (Soderlund *et al.*, 2006). The relative positions of *Gran3* LTRs in the physical map were inferred from these data.

Acknowledgements

This work was supported by grants from the National Science Foundation Plant Genome Program DBI-0321678 to R.A.W., S.A.J. and P.J.S.M., and the Bud Antle Endowed Chair to R.A.W. We thank members of the Arizona Genomics Institute's Resource Center, Sequencing, Finishing and Bioinformatics groups for generating and processing the high-quality sequence data used for this analysis. We are grateful to Dr Paul Sanchez for providing genomic DNA of *O. granulata* for initial dot-blot analysis. We also thank other members of the OMAP research team and advisory committee for fruitful discussions.

Supplementary Material

The following supplementary material is available for this article online:

Figure S1. Dot plot comparisons of complete (horizontal) and incomplete (vertical) forms of *Gran3* elements.

Figure S2. Neighbor joining tree generated from reverse transcriptase domains of most abundant *Ty-3-gypsy* elements from *Oryza sativa* with those from *Gran3* (*O. granulata*) and *Wallabi* (*O. australiensis*).

Figure S3. Distribution of *Gran3* LTR sequence in *O. granulata* according to their reciprocal similarity.

Figure S4. Genomic distribution of *Gran3* family in the *O. granulata* genome as revealed by Fluorescent in situ hybridization (FISH).

Figure S5. Distribution of *Gran3* LTRs in the physical map of *O. granulata*.

Figure S6. Southern hybridization of *Dasheng* on *Oryza* species.

Table S1. Dynamics of *RWG* family across the genus *Oryza*.

Table S2. Distribution of *Gran3* LTRs in the comparative physical map of *O. granulata* genome.

Table S3. Copy number estimates of *RWG* element LTRs in different *Oryza* species obtained using random sheared libraries.

Table S4. List of specific primers used for amplification of specific domains from *RWG* family for each *Oryza* species.

This material is available as part of the online article from http://www.blackwell-synergy.com.

References

Altschul, S.F., Madden, T.L., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Ammiraju, J.S., Luo, M., Goicoechea, J.L. *et al.* (2006) The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res.* **16**, 140–147.

Bennetzen, J.L., Ma, J. and Devos, K.M. (2005) Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* **95**, 127–132.

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.

Doyle, J.J. and Doyle, J.L. (1987) Genomic Plant DNA preparation from fresh tissue – CTAB method. *Phytochem. Bull.* **19**, 11–15.

Feschotte, C., Jiang, N. and Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nature Rev. Genet.* **3**, 329–341.

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.

Guo, Y.L. and Ge, S. (2005) Molecular phylogeny of Oryzeae (Poacea) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am. J. Bot.* **92**, 1548–1558.

Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A. and Wendel, J.F. (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* **16**, 1252–1261.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **7052**, 793–800.

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

Jiang, N., Jordan, I.K. and Wessler, S.R. (2002b) *Dasheng* and *RIRE2*. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. *Plant Physiol.* **130**, 1697–1705.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E. and Schulman, A.H. (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl Acad. Sci USA*, **97**, 6603–6607.

Kalendar, R., Vicient, C.M., Peleg, O., Anamthawat-Jonsson, K., Bolshoy, A. and Schulman, A.H. (2004) Large retrotransposon derivatives: abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics*, **166**, 1437–1450.

Kellogg, E.A. (1998) Relationships of cereal crops and other grasses. *Proc. Natl Acad. Sci. USA*, **95**, 2005–2010.

Kidwell, M.G. and Lisch, D.R. (2000) Transposable elements and host genome evolution. *Trends Ecol. Evol.* **15**, 95–99.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

Kumar, S., Tamura, K. and Nei, M. (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163.

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.

McCarthy, E.M., Liu, J., Lizhi, G. and McDonald, J.F. (2002) Long terminal repeat retrotransposons of *Oryza sativa*. *Genome Biol.* **3**(10): research 0053.1-0053.11.

Meyers, B.C., Tingey, S.V. and Morgante, M. (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res.* **11**, 1660–1676.

Ohtsubo, H., Kumekawa, N. and Ohtsubo, E. (1999) *RIRE2*, a novel gypsy-type retrotransposon from rice. *Genes Genet. Syst.* **74**, 83–91.

Petrov, D.A. and Wendel, J.F. (2006) Evolution of eukaryotic genome structure. In *Evolutionary Genetics: Concepts and Case Studies* (Fox, C.W. and Wolf, J.B., eds). Oxford: Oxford University Press, pp. 144–156.

Piegu, B., Guyot, R., Picault, N. *et al.* (2006) Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* **16**, 1262–1269.

Ramakrishna, W., Dubcovsky, J., Park, Y.J., Busso, C., Emberton, J., SanMiguel, P. and Bennetzen, J.L. (2002) Different types and rates of genome evolution detected by comparative sequence analysis of orthologous segments from four cereal genomes. *Genetics*, **162**, 1389–1400.

Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet.* **16**, 276–277.

Rozen, S. and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (Krawetz, S. and Misener, S., eds). Totowa, NJ: Humana Press, pp. 365–386.

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

SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y. and Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize. *Nature Genet.* **20**, 43–45.

SanMiguel, P.J., Ramakrishna, W., Bennetzen, J.L., Busso, C.S. and Dubcovsky, J. (2002) Transposable elements, genes and recombination in a 215-kb contig from wheat chromosome 5A(m). *Funct. Integr. Genomics*, **2**, 70–80.

Schulman, A.H. and Kalendar, R. (2005) A movable feast: diverse retrotransposons and their contribution to barley genome dynamics. *Cytogenet. Genome Res.* **110**, 598–605.

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**, 893–894.

Soderlund, C., Nelson, W., Shoemaker, A. and Paterson, A.H. (2006) SyMAP: a system for discovering and viewing syntenic regions of FPC maps. *Genome Res.* **16**, 1159–1168.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Tikhonov, A.P., SanMiguel, P.J., Nakajima, Y., Gorenstein, N.M., Bennetzen, J.L. and Avramova, Z. (1999) Colinearity and its exceptions in orthologous adh regions of maize and sorghum. *Proc. Natl Acad. Sci. USA*, **96**, 7409–7414.

Uozu, S., Ikehashi, H., Ohmido, N., Ohtsubo, H., Ohtsubo, E. and Fukui, K. (1997) Repetitive sequences: cause for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol. Biol.* **35**, 791–799.

Vicient, C.M, Suoniemi, A., Anamthawat-Jonsson, K., Tanskanen, J., Beharav, A., Nevo, E. and Schulman, A.H. (1999) Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell*, **11**, 1769–1784.

Vitte, C. and Panaud, O. (2005) LTR-retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet. Genome Res.* **110**, 91–107.

Vitte, C., Ishii, T., Lamy, F., Brar, D.S. and Panaud, O. (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa*, L.). *Mol. Genet. Genomics*, **272**, 504–511.

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.* **395**, 443–460.

Walling, J.G., Shoemaker, R., Young, N., Mudge, J. and Jackson, S. (2006) Chromosome-level homeology in paleopolyploid soybean (*Glycine max*) revealed through integration of genetic and chromosome maps. *Genetics*, **172**, 1893–1900.

Wendel, J.F. and Wessler, S.R. (2000) Retrotransposon-mediated genome evolution on a local ecological scale. *Proc. Natl Acad. Sci USA*, **97**, 6250–6252.

Wendel, J.F., Cronn, R.C., Johnston, J.S. and Price, H.J. (2002) Feast and famine in plant genomes. *Genetica*, **115**, 37–47.

Wing, R.A., Ammiraju, J.S.S., Luo, M. *et al.* (2005) The *Oryza* map alignment project: the golden path to unlocking the genetic potential of wild rice species. *Plant Mol. Biol.* **59**, 53–62.

Yu, J., Wang, J., Lin, W. *et al.* (2005) The genomes of *Oryza sativa*: a history of duplications. *PloS Biol.* **3**(2), e38.

Zhang, X. and Wessler, S.R. (2004) Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc. Natl Acad. Sci. USA*, **101**, 5589–5594.

