

Genome-Scale Sequence Disruption Following Biolistic Transformation in Rice and Maize^[OPEN]

Jianing Liu,^a Natalie J. Nannas,^b Fang-fang Fu,^c Jinghua Shi,^d Brooke Aspinwall,^a Wayne A. Parrott,^e and R. Kelly Dawe^{a,c,1}

^aDepartment of Genetics, University of Georgia, Athens, Georgia 30602

^bDepartment of Biology, Hamilton College, Clinton, New York 13323

^cDepartment of Plant Biology, University of Georgia, Athens, Georgia 30602

^dBionano Genomics, San Diego, California 92121

^eDepartment of Crop and Soil Sciences, University of Georgia, Athens, Georgia 30602

ORCID IDs: 0000-0003-4398-1542 (J.L.); 0000-0002-2617-8822 (N.N.); 0000-0002-3187-9308 (F.F.); 0000-0001-5989-1029 (J.S.); 0000-0003-3422-0506 (B.A.); 0000-0001-7847-1134 (W.P.); 0000-0003-3407-4553 (R.K.D.)

Biolistic transformation delivers nucleic acids into plant cells by bombarding the cells with microprojectiles, which are micron-scale, typically gold particles. Despite the wide use of this technique, little is known about its effect on the cell's genome. We biolistically transformed linear 48-kb phage lambda and two different circular plasmids into rice (*Oryza sativa*) and maize (*Zea mays*) and analyzed the results by whole genome sequencing and optical mapping. Although some transgenic events showed simple insertions, others showed extreme genome damage in the form of chromosome truncations, large deletions, partial trisomy, and evidence of chromothripsis and breakage-fusion bridge cycling. Several transgenic events contained megabase-scale arrays of introduced DNA mixed with genomic fragments assembled by nonhomologous or microhomology-mediated joining. Damaged regions of the genome, assayed by the presence of small fragments displaced elsewhere, were often repaired without a trace, presumably by homology-dependent repair (HDR). The results suggest a model whereby successful biolistic transformation relies on a combination of end joining to insert foreign DNA and HDR to repair collateral damage caused by the microprojectiles. The differing levels of genome damage observed among transgenic events may reflect the stage of the cell cycle and the availability of templates for HDR.

INTRODUCTION

The creation of genetically modified crop lines through transformation is typically performed using *Agrobacterium tumefaciens*-mediated gene transfer (Gelvin, 2017) or particle bombardment (Klein et al., 1989). Both modes of transformation insert recombinant DNA in a random and uncontrolled manner. *Agrobacterium* is viewed as superior because it often delivers complete gene constructs bounded by known left and right borders (Gelvin, 2017). The integration of *Agrobacterium* transfer DNA (T-DNA) occurs at existing double strand breaks through the activity of native polymerase theta and microhomology-mediated repair (van Kregten et al., 2016). Despite its relative precision, most T-DNA insertions are at least dimers (van Kregten et al., 2016) and many are composed of long arrayed multimers (Cluster et al., 1996; Krizkova and Hroudá, 1998; Jupe et al., 2018). In addition, *Agrobacterium* transformation may result in multiple T-DNA insertions at different locations, large deletions (Takano et al., 1997; Kaya et al., 2000), chromosomal inversions, translocations, and duplications (Takano et al., 1997; Nacry et al., 1998; Clark and

Krysan, 2010; Zhu et al., 2010; Anderson et al., 2016; Jupe et al., 2018).

Biolistic transformation offers the advantage that it can deliver any form of DNA, RNA, or protein (Altpeter et al., 2005; Svitashv et al., 2015; Gil-Humanes et al., 2017; Shi et al., 2017), a property that has been exploited to facilitate gene editing technologies (Belhaj et al., 2015; Altpeter et al., 2016; Begemann et al., 2017; Liang et al., 2017). When conditions for biolistic transformation are carefully calibrated, the results can be comparable to *Agrobacterium*-mediated transformation in terms of transformation efficiency and transgene copy number (Lowe et al., 2009; Jackson et al., 2013). Biolistic transformation is also free of the constraints associated with *Agrobacterium*-host plant interactions. Unaltered bacterial artificial chromosome sequences larger than 100 kb (Ercolano et al., 2004; Phan et al., 2007) and an intact linear 53-kb molecule (Partier et al., 2017) have been integrated into plants by biolistic methods. Similarly, very long PCR products containing >100 kb of a simple repeating structure were cobombarded with a selectable marker plasmid to create maize transgenics with inserts ranging from ~200 to 1000 kb in size (Zhang et al., 2012). However, transgene copy number following biolistic transformation can be very high (depending on the amount of DNA delivered into cells; Altpeter et al., 2005) and very little is known about the process or mechanism of insertion following biolistic transformation. Prior literature based primarily on DNA gel blots indicates that sequence breakage and reassembly is common (Pawlowski and Somers, 1996, 1998; Svitashv et al., 2002; Makarevitch et al., 2003). The only detailed sequence-level

¹ Address correspondence to kdawe@uga.edu.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: R. Kelly Dawe (kdawe@uga.edu).

^[OPEN]Articles can be viewed without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.18.00613

IN A NUTSHELL

Background: *Agrobacterium*-mediated gene transfer and biolistic transformation are the two most widely used methods for plant transformation. *Agrobacterium* transformation is the preferred method because it often delivers intact single-copy transgenes, whereas multicopy insertions and fragmented transgenes are prevalent outcomes with biolistic transformation. Despite its deficiencies, biolistic transformation has some notable advantages. It can be used in a wide range of species and can deliver any type of DNA, RNA or protein. Its flexibility has made it an attractive method for introducing Cas9 protein and repair templates in gene editing applications.

Question: We were initially interested in using biolistic transformation to transform long intact DNA molecules into rice and maize, and had chosen linear 48 kb phage lambda and smaller circular plasmids as test cases for this purpose. We wanted to use whole genome sequencing to see how transformation affected both the introduced DNA and the host genome.

Findings: We found that biolistic transformation can result in simple low copy insertions as well as large transgene arrays and severe genome rearrangements. Megabase-scale arrays of broken transgenes and genomic fragments were documented in multiple lines. There was evidence of severe genome damage, including large deletions, partial trisomy, and genome shattering events resembling chromothripsis and breakage-fusion bridge cycling. Other regions of the genome were broken and fragmented but seamlessly repaired. We conclude that transformation is the result of double strand breaks at the level of the transgene and the genome, and that known DNA repair pathways can explain the variety of outcomes observed.

Next steps: Our results highlight the fact that new methods of plant transformation are needed to make full use of emerging gene editing technologies. If biolistic methods are to be used, we recommend performing whole genome sequencing to identify lines without genomic rearrangements, or to outcross the transformants to normal breeding lines prior to detailed analysis.

analysis of transgenes following biolistic transformation revealed a few large fragments and many small shattered pieces, with 50 of 82 insertions being less than 200 bp in length (Svitashev et al., 2002). These limited sequence data suggest there may be unexpected and severe genomic consequences associated with biolistic transformation.

As a means to better understand the mechanistic underpinnings of biolistic transformation, we transformed linear and circular DNA molecules into rice (*Oryza sativa*) and maize (*Zea mays* sp *mays*) and subjected the lines to whole genome sequencing and analysis. The data revealed a wide spectrum of insertions and outcomes, from simple insertions to extraordinarily long shattered arrays. Multiple forms of genome damage were observed, including chromosome breakage and shattering and extreme copy number variation. We also found evidence of homology-directed repair (HDR) at sites that had been damaged during transformation. The data indicate that transformation involves both damage to the genome and fragmentation of the input DNA, creating tens to thousands of double stranded breaks that are repaired by end-joining and HDR in ways that can either create simple insertions or cause large structural changes in the genome.

RESULTS

General Assessments of the Genomes after Cobombardment with Lambda and Plasmid

We biolistically transformed 48-kb linear lambda phage DNA (Casjens and Hendrix, 2015) and appropriate selectable marker plasmids into rice and maize using a twofold (rice) or fourfold (maize) molar excess of lambda. All sequence analyses were performed on genomic DNA extracted from cultured callus tissue

to obtain an unvarnished view of the transformation process; however, three of the rice lines and all of the maize lines were also regenerated to plants (Supplemental Table 1). After screening the transformed callus by PCR to confirm the presence of lambda, we sequenced 14 rice lines and 10 maize lines at low coverage. The data revealed that over a third of the rice events contained less than one copy of lambda, whereas the remaining two-thirds contained ~1 to 43 copies (Supplemental Table 1, where copy number is a sequence coverage value, and does not imply that any single lambda is intact). The maize transgenic events showed a similar wide range from ~1 to 51 copies (Supplemental Table 1). The selectable marker plasmids were observed at lower abundances reflecting their lower representations during transformation.

To interpret the distribution and structure of the insertions, eight rice lines and four maize lines were sequenced at 20× coverage by 75 bp paired-end Illumina sequencing (Table 1). The data were then interpreted using SVDetect, which employs discordant read pairs to predict breakpoint signatures through clustering (Zeitouni et al., 2010), and Lumpy, which uses discordant read pairs and split reads to determine structural variation (SV) types by integrating the probabilities of breakpoint positions (Layer et al., 2014). The paired end Illumina reads were aligned to the rice or maize reference genomes with the complete lambda and plasmid sequences concatenated as separate chromosomes. Insertions were detected as inter-chromosomal translocations between lambda, plasmid, and genome, whereas rearrangements were identified as intra-chromosomal translocations. Based on simulations using in silico modified forms of rice chromosome 1 with randomly inserted lambda/chromosomal fragments, we estimate that our approach identifies ~84% of the breakpoints involving lambda and ~66.5% of the junctions involving two chromosomes but no lambda (Supplemental Table 2).

Table 1. Copy Number of Introduced Molecules (Lambda and Cobombarded Plasmid) and Number of Breakpoints in Rice/Maize Transgenic Genome

Transgenic Events ^a	Genome Coverage	Copy Number		Number of Breakpoints					
		Lambda	Plasmid	Lambda-Lambda	Lambda-Plasmid	Lambda-Genome	Plasmid-Genome	Intra-Genome	Inter-Genome
Os λ-1	22.45	0.12	3.63	1	11	0	2	2	1
Os λ-2	20.95	0.95	0.80	19	1	3	0	2	0
Os λ-3	20.84	0.04	1.37	0	3	1	2	1	1
Os λ-4	19.63	32.48	3.87	517	21	14	0	1	0
Os λ-5	21.58	37.06	2.22	420	18	18	0	1	0
Os λ-6	21.78	17.35	1.18	257	6	123	0	1	13
Os λ-7	19.64	1.72	1.07	51	4	63	3	12	14
Os λ-8	21.84	7.83	0.98	152	3	99	1	67	40
Zm λ-1	14.07	1.73	1.88	22	22	17	6	1	0
Zm λ-2	18.05	19.11	13.97	31	30	15	5	14	19
Zm λ-3	15.55	10.33	2.00	12	8	8	1	1	8
Zm λ-4	17.93	40.48	20.50	241	143	73	4	10	18

^aTransgenic events of rice (*Oryza sativa*) and Maize (*Zea mays*) are labeled as “Os” and “Zm,” respectively.

The sequence data also allowed us to identify deletions and duplications of genomic DNA by changes in read depth as assayed by CNVnator (Abyzov et al., 2011). Unique breakpoints and regions showing copy number variation were plotted using the Circos chromosome visualization software (Figure 1; Supplemental Figures 1 and 2). We found a wide range of sequence complexity, ranging from simple insertions to long complex arrays and massive genome-scale disruptions.

Simple Low-Copy Insertions

Four rice lines and two maize lines had one or few insertions and otherwise did not show evidence of genome damage (Figure 1; Supplemental Figures 1 and 2). In these events, there were fewer than 40 detected breakpoints between lambda, plasmid, and chromosomes (Table 1), and there were small chromosomal deletions of less than 20 kb around insertion sites (Figure 1; Supplemental Figures 1 and 2). For example, in rice λ-1 there is a 27-kb insertion composed of rearranged lambda (5.8 kb) and plasmid (21.2 kb) fragments in a region of chromosome 8 that has sustained an 18-kb deletion. Similarly, maize λ-1 contains 86.3 kb of combined lambda and plasmid DNA in chromosome 9 with no deletion at the point of insertion, and rice λ-4 (discussed in detail below) contains a long array of lambda and plasmid fragments in chromosome 2 and a small 9 bp deletion at the site of insertion. In these and other cases of simple insertions, there was no other evidence of chromosome truncation or duplication as judged by read depth.

Creation of Long Arrays

Several transformants had large amounts of lambda DNA. Rice λ-4 is the simplest of these, with lambda junctions involving three genomic locations (chromosome 2, 9, and 12) and no other evidence of genome damage (Figures 1A and 1E). As assayed by sequence coverage and SV estimates, this event contains the equivalent of 37 copies of lambda broken into a minimum of 552

pieces. Local sequence assembly indicated that the apparent insertions in chromosome 9 and 12 are small sections of chromosomal DNA flanked on both sides by lambda fragments. Two fragments of chromosome 9 are 102 and 464 bp in length, and one fragment of chromosome 12 is 108 bp in length (Figure 2; Supplemental Figure 3). In contrast, on chromosome 2, the assemblies revealed two simple lambda-genome junctions. These data suggest that rice λ-4 has a large insertion on chromosome 2 and that small sections of chromosome 9 and 12 are intermingled within it. Analysis of 23 self-cross progeny from rice λ-4 supported this view showing that the fragments of chromosome 9 and 12 and the junctions on chromosome 2 are genetically linked (Supplemental Figure 3B).

To confirm our interpretation of the rice λ-4 event, we analyzed the original transgenic plant by Bionano optical mapping, where long DNA molecules were fluorescently labeled at the restriction site BspQI, imaged, and assembled into megabase-scale restriction map contigs (Udall and Dawe, 2018). The data revealed no insertions on chromosome 9 or 12, but an unequivocal large insertion on chromosome 2 at the location predicted. There are two assemblies over this region, one for the wild type chromosome 2 and one showing an insertion of at least 1.6 Mb containing novel sequence. The 48-kb lambda molecule contains six BspQI sites in a distinctive pattern. However, Bionano alignment software failed to detect any similarity between lambda and the BspQI recognition pattern within the array on chromosome 2, as expected if lambda molecules were broken and rearranged. To more accurately assess the internal structure of the array, we sequenced a T1 plant that was homozygous for the insertion on chromosome 2 at 25X coverage using PacBio technology. A total of 1810 (45,280/25) lambda fragments ranging in size from 31 to 11,387 bp were identified. Over 96% of the lambda fragments were less than 2 kb with a mean fragment size of 410 bp (Figure 2C).

Evaluation of breakpoint junctions provided information on the mechanisms of repair that operate to create long arrays (Supplemental Figure 4A). The two major forms of non-homologous repair are nonhomologous end joining (NHEJ), which

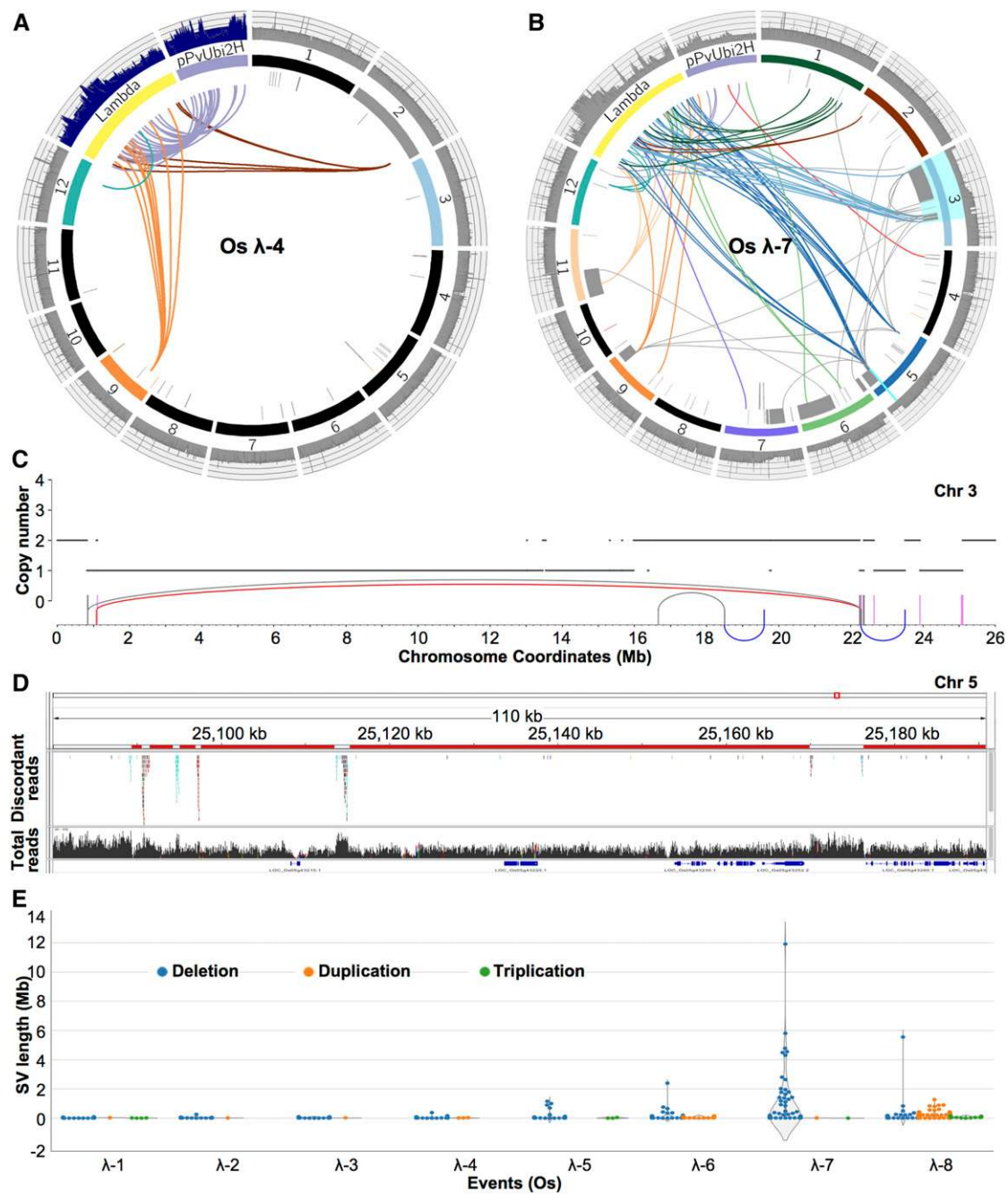


Figure 1. Spectrum of Genomic Outcomes Following Transformation with Lambda and Plasmid In Rice.

All Circos plots are annotated as follows. The twelve rice chromosomes are shown along with λ and plasmid pPvUbi2H magnified at $1,000\times$ and $5,000\times$. The outer track shows sequence coverage over each molecule or chromosome as histograms. The inner ring demonstrates DNA copy number profiles derived from read depth, with gray shown as 1 copy, orange as 3 copies, dark red as 4 copies, and black as more than 4 copies. The inner arcs designate inter- and intra-chromosomal rearrangements. Breakpoints within the genome are colored gray, whereas the breakpoints between λ or plasmid and the genome are colored to match the respective chromosomes.

(A) Rice event λ -4, which contains a long transgene array in chromosome 2. The coverage values in histogram tracks of λ and plasmid are divided by 15 and 1.5, respectively.

(B) Rice event λ -7, illustrating a complex event with severe genome damage.

(C) A 26 Mb region on chromosome 3 (highlighted in cyan in Figure 1B) at high resolution. The horizontal lines show copy number states and vertical bars represent inter-chromosomal breakpoints (gray) and breakpoints involving λ (plum), the arcing links show local rearrangements of the deletion-type (gray),

is typified by blunt end junctions and short insertions (Pannunzio et al., 2017), and microhomology-mediated end joining (MMEJ), which is characterized by junctional microhomology of at least 5 bp (McVey and Lee, 2008). Computational analyses of the junctions in rice and maize transgenic events revealed blunt-end connections (25%), short insertions varying in size from 1 to 80 bp (21%) and junctions displaying microhomology in the range of 1 to 4 nucleotides (50%) and 5 to 25 nucleotides (4%), suggestive of both NHEJ and MMEJ (Supplemental Figure 4B). It is also possible that some of the longer insertions were an outcome of synthesis-dependent strand annealing (SDSA), an alternative form of homology directed repair pathway (HDR, see below). The four relative orientations of lambda fragments (tail-head, tail-tail, head-tail, and head-head), were nearly uniformly distributed (Supplemental Figure 4C) as expected for a random rejoining process. We also investigated whether the natural overlapping single stranded ends of lambda (the 12-bp cos sites; Casjens and Hendrix, 2015) may have played a role in multimerization. The data showed that five rice lines and three maize lines contained a single annealed cos site; a low frequency that supports the view that homology-based annealing and ligation have minor roles in the assembly of broken fragments.

Evidence of Homology-Directed Repair

A second major form of repair is homology-directed repair (HDR), where double stranded breaks are seamlessly corrected using undamaged homologous molecules such as sister chromatids as templates. If a segment of the genome is broken away and not repaired, we expect to find a deletion at the original coordinates, whereas if the damaged region is repaired by HDR, we expect to find no evidence of damage at the original coordinates. Incorporation of a displaced fragment at a new location followed by repair of the original site will result in a total of three copies of the region affected.

The analysis of rice λ -4 revealed that small sections of chromosome 9 and 12 were included in a long array of lambda fragments but that there were no changes from wild type where the original damage occurred (as assayed by optical mapping; Figure 2A). We also analyzed the coordinates surrounding the affected sites on chromosomes 9 and 12 (plus or minus 1 Mb) for a clustering of discordant reads or significant changes in read depth and found no evidence of sequence disruption. Further, PCR analysis of the T0 line revealed no evidence of small deletions at these coordinates. These data are consistent with a model where chromosome 2 and 9 were damaged, broken fragments were included in the assembly of the long chromosome 2 array, and the damaged chromatids were repaired by HDR.

To determine whether HDR had occurred in any of the other lines assayed, we identified 78 additional displaced genomic fragments in four rice events and four maize events. We then systematically checked for increases in read depth and clusters of discordant reads that map to the native locations of these displaced fragments. The data provide evidence of HDR in three rice events and three maize events (Supplemental Tables 3 and 4). For example, a 110-bp displaced fragment from chromosome 1 and a 69-bp fragment from chromosome 9 in rice λ -5, both flanked by lambda pieces, exhibited increased sequence coverage by 50% and no apparent deletions at the original coordinates (Figure 3). Although most of the displaced genomic fragments in lambda arrays were on the order of a few hundred bases, we also found evidence of breakage and repair among the chromosomes on a larger scale (Supplemental Tables 3 and 4). For example in rice λ -8, a 21-kb and a 34-kb region from chromosome 2 were broken away, connected by a small fragment from lambda and reinserted in the genome, followed by repair at the original locations. This resulted in duplication regions clearly visible by read depth (Figure 3B). The limits on the size of a deletion that can be repaired by HDR are not known, but in animals HDR can be used to incorporate new (knock-in) constructs as large as 34 kb (He et al., 2016).

It is formally possible that some of the displaced fragments are an outcome of SDSA (Gorbunova and Levy, 1997). SDSA occurs when one strand from a double stranded break invades an intact DNA molecule and begins to initiate DNA synthesis, but is then released and processed by end joining (Verma and Greenberg, 2016). Under this model the DNA scored as displaced would actually have been copied from an undamaged location. However, SDSA events tend to be short (<50 bases; Kleinboelting et al., 2015) and this mechanism probably cannot explain the longer displaced regions we have observed (13 are larger than 1 kb, Supplemental Tables 3 and 4). The fact that the majority of displaced fragments are associated with deletions at the original location also tends to favor the HDR model over the SDSA model.

Deletions and Evidence of Breakage-Fusion-Bridge Cycling

Copy number profiling provided evidence for many deletions ranging in size from 3.5 kb to 11.9 Mb in rice and 115 kb to 62 Mb in maize (Supplemental Data Set). Deletions and duplications/triplications greater than 1 Mb were found in four rice events and three maize events (Figure 1E; Supplemental Figure 2B; Supplemental Data Set). Deletions were particularly common around transgene insertions and at the ends of chromosomes, and the majority were associated with the presence of lambda or plasmid DNA, indicating that the breaks occurred as a

Figure 1. (continued).

duplication-type (red), and intra-chromosomal translocation-type (blue). For a visual depiction of how local rearrangements are defined using paired end reads, see Supplemental Figure 5.

(D) A region from 25.1 Mb to 25.2 Mb on chromosome 5 (highlighted in cyan in Figure 1B) as visualized with IGV. Deleted regions are shown in red and retained regions in white (top), as indicated by the alignment of discordant reads (middle) and read depth (bottom).

(E) Swarm and violin-plots showing the distribution of the size and number of deletions, duplications, and triplications in all rice events transformed with λ . Each dot in the swarm plots represents a different SV. Violin plots represent the statistical distribution, where the width shows the probability of given SV lengths.

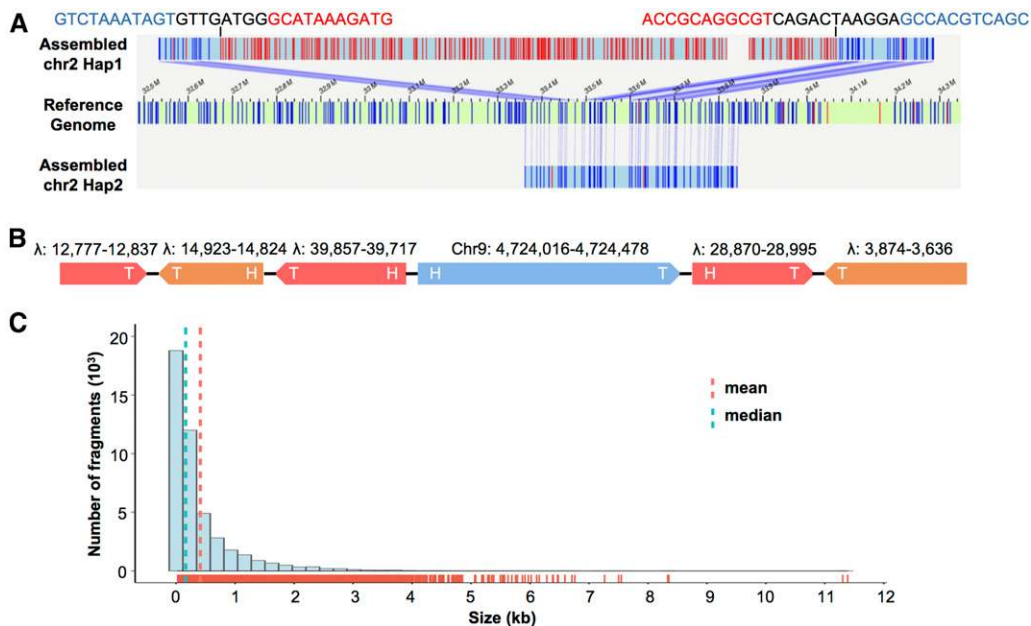


Figure 2. Characteristics of the Long Transgene Array in Rice Event λ -4.

(A) Bionano assembly depicting the 1.6-Mb insertion in chromosome 2. The middle panel represents the reference genome, and the top and bottom panels depict the assembled transgenic and wild type chromosomes in this heterozygous line. The blue bars indicate matching restriction sites between the reference and assembled contigs, and red bars denote restriction sites within the insertion. The nucleotide sequences above the top panel show the breakpoint sequences, with chromosome sequences highlighted in blue, λ sequences highlighted in red, and new sequences in black.

(B) A 1.1-kb region assembled from Illumina data showing five λ pieces and a single fragment of chromosome 9 in rice event λ -4. The direction of the arrows indicates the 3' ends (Tails) of λ and chromosomal genomic fragments. Four different relative orientations between intra- and inter-chromosomal pieces can be found in this sequence: Tail (3')-Head (5'), Tail-Tail, Head-Tail, Head-Head.

(C) Size distribution of λ fragments in the array as determined by PacBio sequencing.

consequence of the transformation process. Deletions that appeared to have no connection with lambda or plasmid may either reflect our imperfect (84%) ability to detect such junctions, or identify regions that were damaged and repaired without the involvement of introduced DNA. No deletions were observed in the single non-transformed rice callus line used as a control.

Chromosome breakage is expected to yield a double stranded break that is repaired by ligation to an introduced DNA molecule or to another broken chromosome. The fusion of two different chromosomes can cause the formation of a dicentric chromosome that is unstable during mitosis. When the centromeres on a dicentric chromosome move in opposite directions during anaphase, the pulling forces cause a re-breaking of the chromosome that initiates a breakage-fusion-bridge (BFB) cycle that may repeat for many cell divisions (McClintock, 1942; Zakov et al., 2013; Storchová and Kloosterman, 2016). The BFB cycle can lead to local duplications and higher order expansions (Campbell et al., 2010; Mardin et al., 2015). Chromosomes 4 and 7 in rice λ -8 show complex rearrangements and evidence of trisomy (Figure 4) that is consistent with errors at the level of chromosome segregation. Copy number gains were observed on chromosome 6 in maize λ -4, where the amplified regions are adjacent to a terminal deletion (Figure 4D). At least two inversions of 3.9 Mb and 2.8 Mb were found in the amplified area. Read depth increases adjacent to a terminal deletion

were also found on chromosome 9 in maize λ -3, where the amplified region displayed switches from 2 to 6 copies (Figure 4C).

Shattering and Chromothripsis-Like Outcomes

Animal cells sustaining chromosome loss or breakage undergo a process known as chromothripsis that results in complex genomic rearrangements in localized areas, generally consisting of tens to hundreds of small pieces (Stephens et al., 2011; Korbel and Campbell, 2013). The reassembly process involves a reshuffling and loss of sections of the genome. Instead of uniform coverage, a region that has undergone chromothripsis shows oscillations from the normal copy number state of two to a copy number state of one (haploid) and occasionally three (triploid) in the context of numerous rearrangements. Analysis of the rice and maize transgenic events revealed similar oscillating copy number states in regions surrounding what appear as “impact sites” on Circos displays: large areas of genome damage with multiple lambda and plasmid fragments.

We found particularly complex rearrangements with copy number oscillations and interspersed lambda and plasmid fragments in three rice events. In rice λ -7, broken fragments (44 bp to 7858 bp) from localized regions of chromosome 3, 5, 6, 7, 9, and 11 were interlinked along with lambda and plasmid fragments in inverted and noninverted orientations (Figures 1B to 1D; Supplemental Figure 5). These patchwork assemblages are

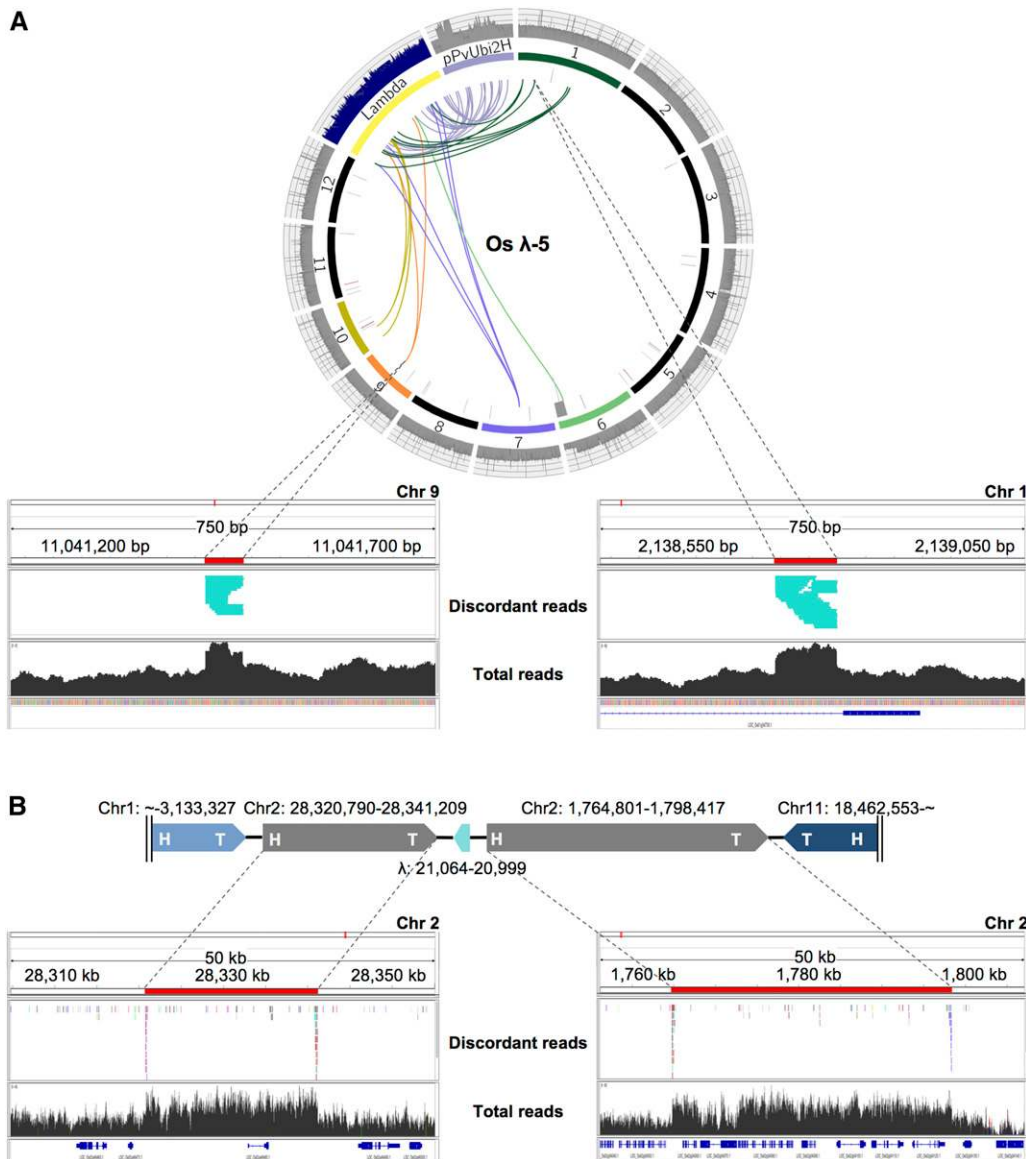


Figure 3. Evidence of HDR in Rice Transgenic Events.

(A) Circos plot of rice transgenic event λ -5 annotated as in Figure 1. The λ coverage is divided by 15. Region 2,138,442 - 2,139,257 on chromosome 1 and region 11,041,419 - 11,041,484 on chromosome 9 are displayed in IGV windows, where displaced fragments (110 bp and 66 bp) are highlighted in red. The top panels show only discordant reads (where one end maps to the fragment and the other maps to another chromosome). The bottom panels show all reads, illustrating the ~50% increase in read depth indicative of an HDR event.

(B) Complex rearrangements observed in rice event λ -8. Regions from chromosome 2 were assembled into an array with other broken fragments at an unknown location in the genome. The damaged regions of chromosome 2 were subsequently repaired as demonstrated by the ~50% increase in read depth.

presumably integrated into one or a few arrays. The damage imparted during transformation caused large swathes of the same regions on chromosomes 3, 5, 6, 7, 9, and 11 to be deleted. The combination of retained displaced fragments and deletions results in oscillating patterns between 1 and 2 copy number states (Figures 1C and 1D). Higher order oscillation patterns were identified in rice λ -8, where numerous fragments from chromosome 1 were linked with segments of chromosome 2, 4, 7, 9,

and 11 in what is likely another complex array (Figure 4A). However, in this case the read depth data indicate that the damaged regions of chromosome 1 were repaired by HDR. The combination of retained displaced fragments and repaired regions result in oscillating patterns between 2 and 3 copy number states (Figure 4B).

Similar results were found in three maize events where large deletions and duplications occurred. The sensitivity of our

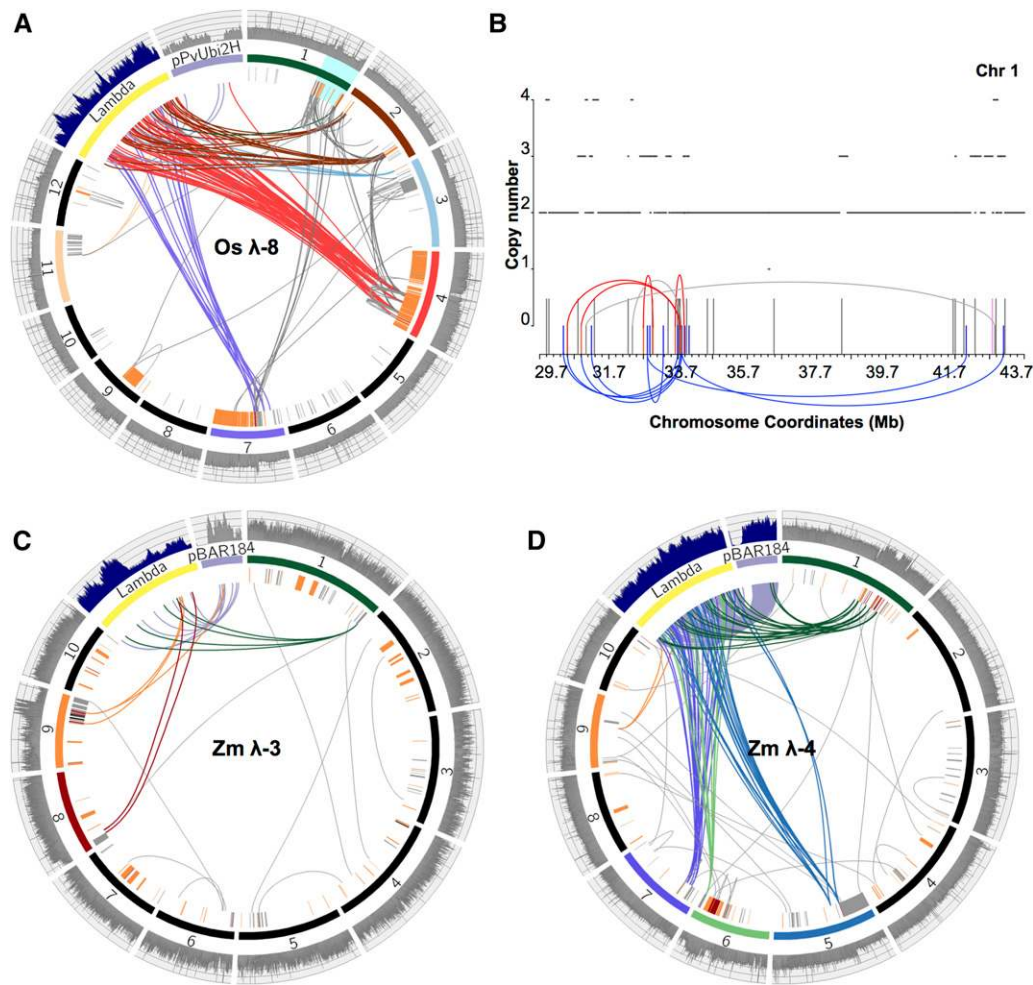


Figure 4. Chromothripsis-Like Outcomes and BFB-like Genomic Rearrangements in Rice and Maize Transgenic Events.

(A) Circos plot of rice transgenic event λ -8 annotated as in Figure 1. The coverage of λ in the histogram track is divided by 4.

(B) Copy number states of region 29.7 - 43.7 Mb on chromosome 1 (highlighted in cyan in Figure 4A) annotated as in Figure 1C.

(C) Circos plot of maize transgenic event λ -3, with coverage of λ in the histogram track divided by 5. Note the region of increased copy number states on chromosome 9 indicative of BFB.

(D) Circos plot of maize transgenic event λ -4, with the coverage of λ and plasmid in the histogram track divided by 15 and 10, respectively. Note the regions of increased copy number states on chromosome 1 and 6 indicative of BFB.

assay is significantly lower in maize because of the high repeat content and necessity of using only perfectly mapped reads. Although we can only detect a fraction of the rearrangements present, the linking patterns between displaced genomic segments and lambda and plasmid is obvious (Figures 4C and 4D; Supplemental Figure 2A). For example, maize λ -4 shows lambda and plasmid within an inter-chromosomal network including sections of chromosomes 1, 5, 6, 7, and 9, as well as evidence of copy number switching (Figure 4D).

Similar Genome Scale Disturbances in Single Plasmid Transformations

We were concerned that the linearity of lambda or the high concentration of DNA used when transforming lambda may have

led to new or extreme forms of genome damage. To test whether this was the case, we transformed rice with circular plasmids designed to knockdown (pANIC10A-OsFPGS1) or overexpress (pANIC12A-OsFPGS1) folic polyglutamate synthetase 1 (chosen for its presumptive role in regulating lignin content). Approximately 125 ng of DNA was delivered to 100 mg of callus tissue per shot, which is considerably lower than the 585 ng of DNA delivered for lambda. In addition, we only sequenced the genomes of fully regenerated plants in these experiments.

The rice lines transformed with single plasmids showed a narrower span of transgene copy numbers (ranging from 0.5 to 12.3X; Supplemental Table 5), consistent with the lower amount of DNA used in transformation (Lowe et al., 2009). However, the genome-level damage (average inter- and intra-chromosome breakages, 17.9) was nearly identical to what we observed for the lambda

transformation experiments (19.5 for rice and 17.8 for maize). In general, although there is a natural relationship between transgene copy number and the number of junctions between the plasmid or lambda and the genome (the transgenes must insert somewhere), the copy number of the transgene was not correlated with the level of collateral damage at other genomic sites. Lines with one copy of the transgene are just as likely to have sustained damage elsewhere in the genome than lines with multiple copies (Figure 5E).

As in the lambda experiments, single plasmid transformations caused large-scale deletions, inversions, duplications consistent with BFB, and rearrangement patterns indicative of chromothripsis-like processes (Figure 5; Supplemental Figures 6 and 7). For example, in event 12A-6, chromosome 4 sustained a large deletion and the remainder of the chromosome was duplicated to create a region of partial trisomy (Figure 5D). Evidence of alternating copy number states was found on chromosome 1 in event 10A-6 (Supplemental Figure 6B) and chromosome 8 in event 12A-3 (Supplemental Figure 7B).

DISCUSSION

Here we provide data showing that biolistically transformed rice and maize plants contain a wide diversity of transgene copy numbers ranging from a fraction of a single copy to as many as 51. Although it is known that lowering the amounts of input DNA (<1 ng/kb of input DNA per shot) can result in more single copy insertions (as high as 54% in maize; Lowe et al., 2009), single copy insertions are also commonly observed when higher amounts of input DNA are used to improve transformation efficiency (~10 ng/kb of input DNA per shot; Li et al., 2016; Raji et al., 2018). Seven of the 24 events we analyzed had less than 1.5 copies of the plasmid by read depth (Table 1; Supplemental Table 5). Our expectation based on prior work was that lines with multiple transgenes would contain complex arrays of broken and rearranged plasmids (Register et al., 1994; Gorbunova and Levy, 1997; Kohli et al., 1999; Svitashv et al., 2000; Jackson et al., 2001; Makarevitch et al., 2003; Shou et al., 2004). Key among the early studies was work from the Somers lab (Svitashv et al., 2002; Makarevitch et al., 2003) showing that plasmids transformed biolistically are frequently broken into small (<100 bp) pieces and scrambled with genomic segments. Our results strongly support these interpretations, illustrated most vividly by our analysis of the long lambda array in rice λ -4, which contained total of 1810 lambda fragments ranging in size from 31 to 11,387 bp (Figure 2C). The Somers group further speculated that DNA was broken randomly and rejoined at blunt ends often containing microhomology (Svitashv et al., 2002). Our more extensive analysis implicates NHEJ as the primary mechanism for rejoining broken fragments and that MMEJ and perhaps SDSA is involved as a secondary pathway.

In addition to confirming the broken and rearranged fate of transgenes following biolistic transformation, we found massive genome rearrangements on a scale that would have been difficult to anticipate. Our focus on callus tissue gave us a perspective on the outcome of transformation than might not have been visible had we worked entirely with regenerated plants. Callus is known to tolerate chromosome instability (Lee and Phillips, 1988) and is

presumably more tolerant of mutations than differentiated tissue. Likewise, our use of long linear molecules allowed us to visualize DNA rearrangements with greater ease than would have been possible with plasmids alone. Nevertheless, the same types of breakages and copy number variation were observed with single plasmid transformants assayed in regenerated plants. Most of the major events were associated with fragments of introduced DNA, implicating the microprojectiles themselves as the primary mutagens. Such damage is to be expected, because the 0.45 μ m gold beads used for rice transformation are about a quarter of the diameter of a rice nucleus (~2 μ m; Jones and Rost, 1989) and 225 times larger than the diameter of DNA. When the genome is damaged in this manner, it can be repaired in one of three ways (Figure 6):

Repair can occur by homology-directed repair such that the damaged region is completely restored to its original state (Figure 6D).

Repair can occur by NHEJ or MMEJ, where the end of any other broken DNA molecule is used as a substrate. Broken fragments of introduced DNA are a likely substrate particularly when they contain markers that are under selection. The other end of the newly joined fragment may then be ligated to a second fragment of introduced DNA or to another segment of the genome. If this process culminates by reconnecting the two pieces of the original chromosome, the result will be a "simple insertion" containing a variable number of conjoined foreign DNA fragments (Figure 6A).

Repair can be initiated by the process above but not culminate in the reconstitution of the original chromosome. The break may not be repaired at all or it may culminate in connecting of two different chromosomes. In this case there can be severe genomic consequences including large terminal deficiencies, chromosome fusion and BFB cycling, and more complex events resembling chromothripsis (Figures 6B and 6C). These dramatic chromosomal rearrangements are a natural outcome of the same processes that are used to create a simple insertion.

The stage of the cell cycle may have a significant impact on the outcome of biolistic transformation. Data from nonplant systems indicate that although NHEJ is active throughout interphase, it is particularly important in G1. In contrast, HDR is more likely in S and G2 phases after DNA replication has provided additional templates for repair (Heyer et al., 2010; Karanam et al., 2012; Ceccaldi et al., 2016). Simple insertions may be more probable when the cell is transformed in S or G2 so that NHEJ can insert the foreign DNA while HDR serves to repair extraneous damage. Simple insertions may also be an outcome of transformation during mitosis when chromosomes are distributed in the cytoplasm. DNA introduced during metaphase or anaphase might find its way into newly forming telophase nuclei, and subsequently be inserted into the genome as consequence of routine DNA repair (similar to T-DNA; van Kregten et al., 2016).

When chromosomes are broken in G1, deletions and translocations are to be expected. We observed many examples of chromosomes that were missing large terminal segments of chromosome arms (Figures 1B, 3A, 4C, 4D, 5C and 5D). The

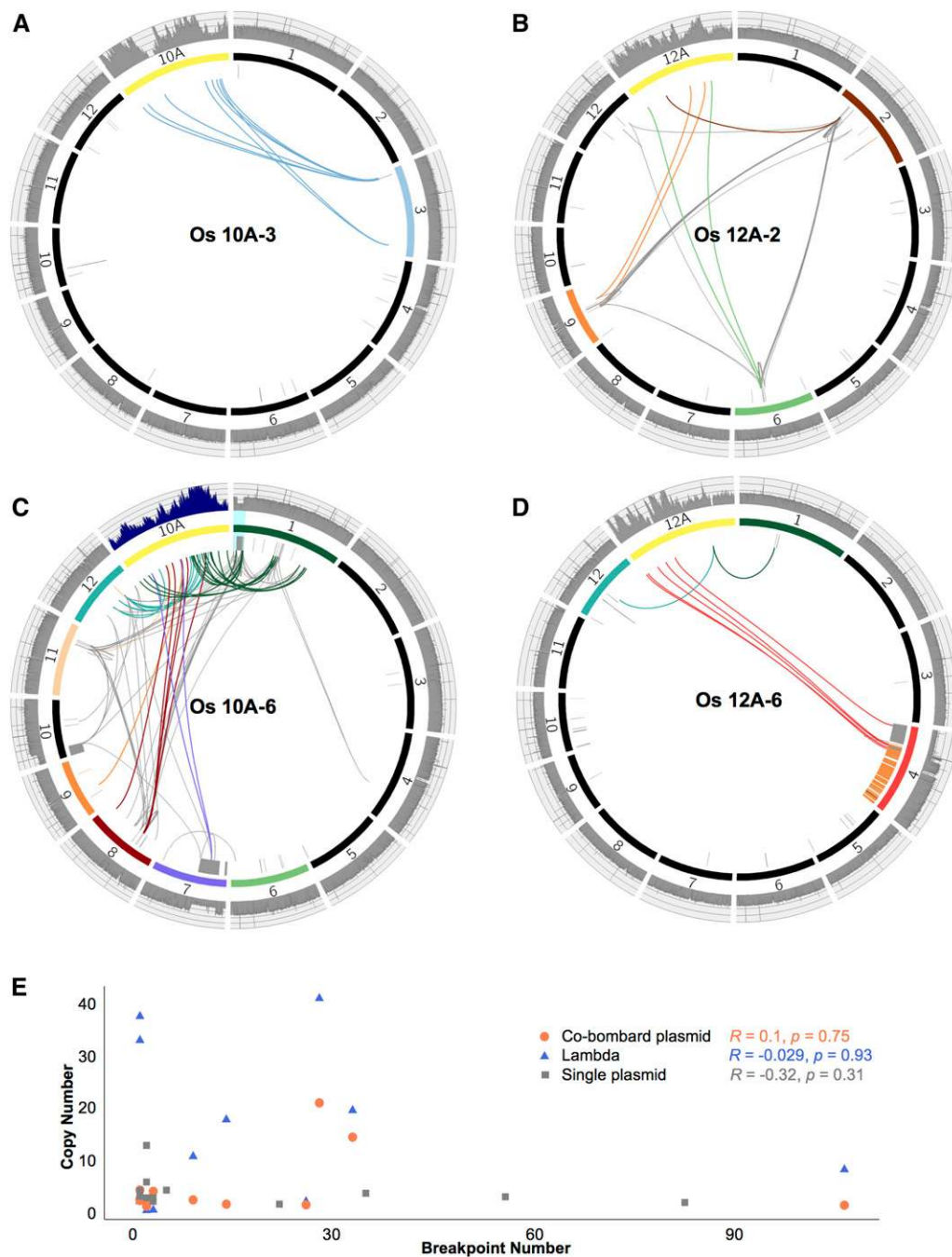


Figure 5. Similar Genomic Disturbances Following Single Plasmid Transformations. Circos plots of rice lines transformed with plasmid pANIC10A-OsFPGS1 (**A**) and (**C**) and pANIC12A-OsFPGS1 (**B**) and (**D**).

(**A**) Simple insertion.

(**B**) Complex insertion showing a network of interlinked genomic regions.

(**C**) Extensive damage with a deletion on chromosome 7 and apparent chromothripsis on chromosome 1 (coverage of the 10A plasmid is divided by 2). See Supplemental Figure 6B for a detailed view of the chromothripsis region on chromosome 1 (highlighted in cyan).

(**D**) Chromosome-scale disruption with a partially trisomic chromosome 4.

(**E**) Relationship between transgene copy number and genome breakage at sites not involving the transgene (intra- and inter-chromosomal translocations). Blue triangles and orange circles show lambda and co-bombarded plasmid from the lambda transformation events. Gray squares show data from single plasmid transformations. There are no significant correlations. Pearson correlation coefficient (**R**) and p-value are indicated.

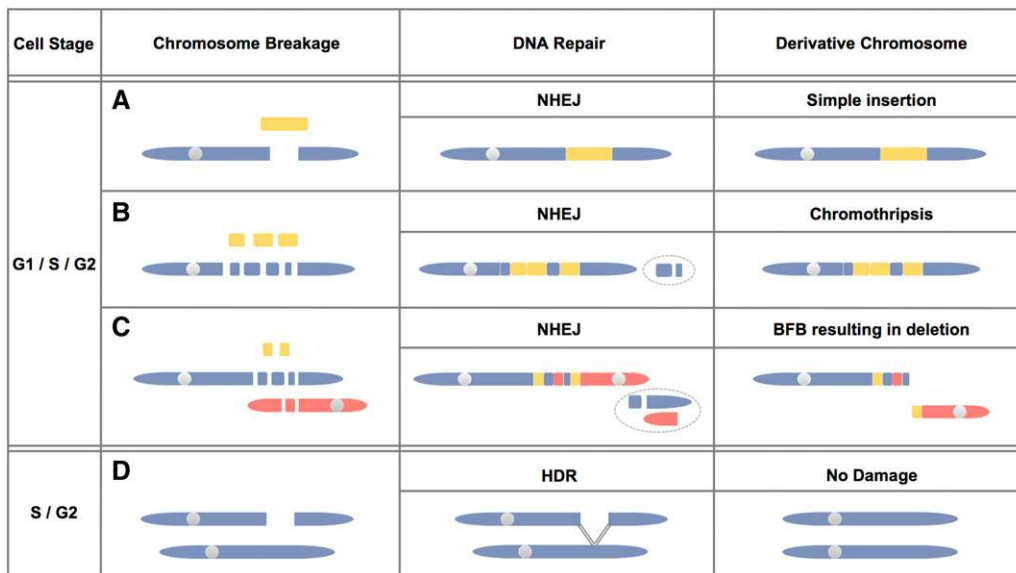


Figure 6. Models for Genomic Outcomes After Biolistic Transformation.

The stage of cell cycle may influence the outcome of biolistic transformation. The models are based on the fact that in animals and presumably plants, NHEJ is the most likely repair pathway in G1 and homology directed repair (HDR) is more likely in S and G2.

(A) Simple insertion. Fragments of introduced molecules (yellow) are ligated with broken ends of native chromosomes by NHEJ (nonhomologous end joining).

(B) Chromothripsis-like genome rearrangements. Localized regions from native genome are shattered, resulting in many double stranded breaks. Fragments of chromosomes and introduced molecules are stitched together through NHEJ, creating complex patterns that involve the loss of genomic DNA and changes in copy number state (lost regions are circled).

(C) Breakage and joining of two different chromosomes and breakage-fusion-bridge (BFB)-like genome rearrangements. When two chromosomes are broken, they can be ligated together through NHEJ. The resulting dicentric chromosome is expected to undergo BFB, which can result in stable terminal deletions.

(D) DNA damage repaired by HDR. Double stranded breaks in S or G2 phase may be repaired by HDR through recombination with an intact sister chromatid.

formation of a stable truncated chromosome requires that the end be healed by formation of a telomere, which is a process that occurs over a period of cell divisions (McClintock, 1941; Chabchoub et al., 2007). In the period when there is an unattended double strand break without a stable telomere, the break is likely to be repaired by NHEJ using any other broken chromosome. As famously described by Barbara McClintock (McClintock, 1941), the fusion of broken chromosomes can initiate a BFB cycle and amplification of genome segments on the affected chromosomes. In several cases we observed copy number states of 4, 5 and 6 that are difficult to explain by any other mechanism. We also observed partial and fully trisomic chromosomes (Figures 4A and 5D). Such large-scale chromosome abnormalities may also be the result of the tissue culture process itself (Lee and Phillips, 1988), and we cannot rule out the possibility that some of the chromosomal changes were either present before transformation or occurred after transformation. However, for most of the large duplications and deletions we observed, there was either evidence of inserted foreign DNA or evidence that the lost DNA had been fragmentation and rejoined with foreign DNA.

In addition, our analyses revealed extreme shattering and chromothripsis-like outcomes. Chromothripsis was originally described as a process whereby “tens to hundreds of genomic rearrangements occur in a one-off cellular crisis” (Stephens et al.,

2011). Our data meet this definition in a descriptive sense, but the biological underpinnings are presumably different. For cancer lines, the simplest model (as it relates to our study, for other models, see Rode et al., 2016) requires that a chromosome be partitioned from the primary nucleus, generally as a result of an error in chromosome segregation that leaves it stranded in the cytoplasm (Zhang et al., 2015). The resulting micronuclei show aberrant DNA replication (Crasta et al., 2012; Leibowitz et al., 2015; Zhang et al., 2015) and appear to have fragmented chromatin (Crasta et al., 2012). The partially degraded chromatin can then be reincorporated into the primary nucleus where it is evident as broken and reassociated fragments (Rode et al., 2016). Recent data indicate that when plants sustain errors in chromosome segregation, they too show evidence of chromosome fragmentation with oscillating copy number states confined to single chromosomes (Tan et al., 2015). In contrast, our biolistically transformed lines are not expected to undergo regular loss of chromosomes during cell division. It is possible that microprojectiles severely damage nuclei such that portions of the genome are released into the cytoplasm. Another plausible explanation is that acentric fragments formed during the repair process (Figures 6B and 6C) are lost during anaphase, become partially degraded, and are reincorporated into a nucleus during a subsequent cell division.

Taken together our data help to explain the long nearly continuous arrays of 156 bp repeats we observed following biolistic transformation of PCR products in maize (Zhang et al., 2012). At the time we were unable to determine whether the long PCR products had been transferred intact or were broken and reassembled in planta. Based on the data here it seems more likely that the PCR products were fragmented and reassembled by NHEJ to create the observed long arrays. Although we did recover simple low copy insertions, the conditions used were not ideal for recovering this type of event at high frequencies. Researchers wishing to do so would be well served to lower the amounts of DNA and consider using linearized plasmids or amplified fragments that are more likely to be inserted at low copy numbers (Fu et al., 2000; Tassy et al., 2014). Constructs as long as 53 kb have been recovered with careful selection for low copy inserts (Partier et al., 2017), although this kind of success is rare.

From a product development perspective, genomic rearrangements were initially considered to be a food/feed safety hazard (Kessler et al., 1992). To put this hazard in perspective, Anderson et al. (Anderson et al., 2016) noted that the genomic rearrangements from *Agrobacterium*-mediated transformation were an order of magnitude lower than those created by fast-neutron mutagenesis. In turn, rearrangements from fast-neutron mutagenesis were an order of magnitude lower than the standing genomic structural variations in the cultivated soybean germplasm pool, all of which has a history of safe use. The frequency of rearrangements from biolistic transformation may be more comparable to that induced by fast neutrons. Regardless, as of yet, there is no evidence that a genomic rearrangement has compromised the safety of a plant used as food (Weber et al., 2012), although its agronomic performance can be compromised. Because poor agronomic performance is not tolerated in modern cultivars and hybrids, there is a rigorous selection process that eliminates deleterious mutations during the breeding process (Glenn et al., 2017).

From a research perspective, such rearrangements may be acceptable in some cases, whereas in others it may be necessary to consider that undetected rearrangements could be influencing the phenotype. Gene editing applications are a special case where the intent is usually to make a single precise change. Although there is great appeal in directly introducing Cas9 ribonucleoproteins (Liang et al., 2017) and repair templates (Altpeter et al., 2016) for this purpose, our data suggest that there is strong likelihood that the delivery method itself will cause unintended genome damage. Until new transformation methods become available, the *Agrobacterium*-based methods that have been in regular use for decades (Gelvin, 2017) remain the superior alternative in terms of minimizing genome rearrangements.

METHODS

Rice Transformation

Rice (*Oryza sativa*) variety Taipei 309 was transformed as described previously using 0.45 μm gold beads (Phan et al., 2007). For the lambda experiments, we mixed 33 ng of the 5839 bp plasmid pPvUbi2H (Mann et al., 2012), which confers hygromycin resistance and a twofold molar excess (552 ng) of purified lambda DNA cl857 (New England Biolabs

#N3011S). This equates to 5.6 ng/kb of plasmid DNA per shot and 11.0 ng/kb of lambda per shot. After screening for lambda by PCR (forward primer 5'-GACTCTGCCGCCGTCATAAAATGG and reverse primer 5'-TCGGGAGATAGTAATTAGCATCCGCC), 14 callus lines were chosen for sequence analysis. Three of these callus lines were regenerated to mature rice plants (Supplemental Table 1).

The plasmids pANIC10A-OsFPGS1 (17,603 bp) and pANIC12A-OsFPGS1 (17,501 bp) are based on the pANIC backbone (Mann et al., 2012) with inserts designed to silence or overexpress folylpolyglutamate synthetase. In these experiments only plasmid DNA was used, delivering ~125 ng per shot. This equates to 7.1 ng/kb of plasmid per shot. All 12 lines were regenerated to plants.

Maize Transformation

Biolistic transformation of the maize (*Zea mays*) inbred Hi-II was performed by the Iowa State University Transformation Facility (Ames, IA) as previously described using 0.6 μm gold beads (Frame et al., 2000). To achieve a four molar excess of lambda DNA, we mixed 20 ng of the 7121 bp plasmid pBAR184, which confers resistance to glyphosate (Frame et al., 2000) and 528 ng of purified lambda DNA cl857 (New England Biolabs #N3011S). This equates to 2.8 ng/kb of plasmid DNA per shot and 11.5 ng/kb of lambda per shot. Ten callus lines were screened for lambda by PCR (forward primer 5'-GACTCTGCCGCCGTCATAAAATGG and reverse primer 5'-TCGGGAGATAGTAATTAGCATCCGCC) and subjected to sequence analysis. All of these lines were later regenerated to mature maize plants (Supplemental Table 1).

Library Preparation and Sequencing

DNA was extracted by the cetyltrimethyl ammonium bromide method (Clarke, 2009), and libraries were prepared using KAPA Hyper Prep kit and KAPA Single-Indexed Adapter kit for Illumina Platforms (KAPA Biosciences, KK8504 and KK8700). For the lambda experiments, 14 rice and 10 maize lines were skim sequenced at low coverage (~1 \times) using Illumina NextSeq PE35. Of those, eight rice and four maize lines were chosen for deeper sequencing using Illumina NextSeq PE75, achieving an average coverage of 20X for rice and 15X for maize. For plasmid experiments, six lines each transformed with either pANIC10A-OsFPGS1 or pANIC12A-OsFPGS1 were sequenced with Illumina NextSeq PE75 at ~20X.

Copy Number Calculation

The lambda and plasmid sequences were added to the rice (Kawahara et al., 2013) and maize reference genomes (Jiao et al., 2017) as separate chromosomes to construct concatenated genomes, which were then used as references for read alignment by BWA-mem (version 0.7.15) with default parameters (Li, 2013). For skim-sequenced lines, the mean coverage of lambda/plasmid and genome in each event was estimated as the division of the total number of reads mapped to individual sequences by their respective genome sizes. For lines sequenced at high coverage, the average read depth of lambda/plasmid and genome was calculated as the mean of per-base coverage analyzed by bedtools (version 2.26). The copy number of lambda/plasmid was then derived by multiplying the mean coverage by two, considering that the insertions are heterozygous in diploid genomes.

SV Calling

After adapter trimming by trimalore (version 0.4.4) and quality checking by fastqc (version 0.11.3) at default settings, reads were aligned to the rice/maize concatenated genomes where lambda and plasmid sequences were

added as separate chromosomes, using the Burrows-Wheeler Aligner (BWA-MEM, version 0.7.15) with default parameters. PCR duplicates were removed by Picard's MarkDuplicates (version 2.4.1), and MAPQ filter of 20 was applied. The output BAM files were analyzed for structural variants by SVDetect (version 0.7; Zeitouni et al., 2010) and Lumpy (version 0.2.13; Layer et al., 2014) to call inter-chromosomal translocations and intra-chromosomal translocations. For SVDetect, step length and window size were calculated separately for each sample and structural variants supported by fewer than two reads were filtered. For Lumpy, the mean and sd of insert sizes were calculated for each sample, with two reads set as minimum weight for a call and trim threshold set as 0. For intra-chromosomal translocations, the read cutoff for both Lumpy and SVDetect was set at 3 to increase accuracy. Structural variants in each event called (from both Lumpy and SVDetect) were filtered against those of wild type and the other events with an in-house script. Unique breakpoints were manually inspected with the Integrative Genomics Viewer (IGV) (version 2.3.81) and plotted with Circos (version 0.69; Krzywinski et al., 2009).

Data Simulation

We performed four sets of simulations by embedding shattered and reshuffled fragments from lambda or other chromosomes into the rice reference chromosome 1 sequence at random sites (Supplemental Table 2). Subsequently, a heterozygous diploid genome was constructed by concatenating the modified chromosome with reference chromosome 1. Paired-end Illumina reads were then simulated by ART (version 2.5.8; Huang et al., 2012) at coverage 10X. For ART, the Illumina sequencing system was set as NextSeq 500 v2 (75 bp), average fragment size and sd were set to 300 and 80 bp, respectively. The inter- and intra-chromosomal translocations in each simulated data set were then identified with the SV calling pipeline described above. The output of Lumpy and SVDetect was compared with the simulated data to assess detection performance.

Junction Assembly and Validation by PCR

Reads that support lambda-genome junctions identified by both SVDetect and Lumpy were assembled by SPAdes (version 3.10.0; Bankevich et al., 2012) with default parameters. The output sequences were aligned against the reference genome with NCBI BLAST (version 2.2.26) at default parameters and used as templates for primer design. The BLAST output of all rice events transformed with lambda was then subjected to analyzing microhomology at junction sites and identifying relative orientations between ligated fragments with an in-house script. Selected products of PCR were sequenced and aligned to the assemblies.

Copy Number Variation Detection

CNVnator (version 0.3.3; Abyzov et al., 2011) was used to call copy number variation on BAM files where mapping quality was set to be at least 20. The bin sizes for rice and maize genomes were set at 500 and 5000 bp, respectively. The CNVnator output was filtered by removing calls with $q0 > 0.5$ and $eval1 > 0.01$ using an in-house script. We declared copy number variation as a deletion if the copy number in specific sample is between 0.5 and 1.5, and at least 0.5 lower than that in wild type and all other samples (unaltered regions are expected to have copy numbers between 1.5 to 2.5). We declared copy number variation as duplication if the copy number in specific sample is between 2.5 and 10, and at least 0.5 higher than in wild type and all other samples. Copy number variations in nonrepetitive regions where breakpoints were identified were further inspected using IGV.

It is possible that some of the callus samples were chimeric and contained tissue from more than one independent transformation event. With use of our filtering pipeline, if a callus sample contained two deletion events

in roughly equal proportions we might have detected both, but if the proportions were not equal the less abundant one would most likely not have been detected. We would be unlikely to detect duplications if they were chimeric.

Bionano Optical Mapping

High molecular weight DNA was prepared from rice λ -4 young leaf tissue using the IrysPrep Plant Tissue DNA Isolation Kit (RE-014-05) and labeled with Nt.BspQI using the IrysPrep NRLS labeling kit (RE-012-10). Data were collected at the Georgia Genomics and Bioinformatics Core facility on a single BioNano IrysChip at 80X coverage with an average molecule length of 248 kb. The raw data were assembled with IrysView software (version 2.5.1) set to "optArgument_human," resulting in 501 BioNano genome maps with an N50 of 1.050 Mbp. The genome maps were then aligned to an *in silico*-digested BspQI *cmap* of the rice Nipponbare reference genome. Overall alignment was excellent, yielding a "Total Unique Aligned Len / Ref Len" value of 0.946, which exceeds the general recommendation of 0.85 (Udall and Dawe, 2018). Potential SVs were identified using Bionano Solve software (version 3.0.1) and analyzed individually by eye. In addition to the large insertion on chromosome 2, the SV calling software identified several other regions with small (<100 kb), potential insertions in the rice λ -4 sample relative to the Nipponbare reference chromosome (Chr3: 31,097,744, Chr12: 20,548,710, Chr1: 2,287,999, Chr3: 13,461,815, Chr3: 16,548,253, Chr6: 3,287,514, Chr7: 22,897,940). Because none of these correspond to the coordinates of lambda or plasmid junctions identified by sequence analysis, they may either represent differences between the Taipei 309 line (used for transformation) and the Nipponbare reference, errors in either assembly, or small insertions caused by biolistic transformation but not involving lambda or the plasmid.

PacBio Sequencing and Analysis

High molecular weight DNA was prepared using a modified CTAB method (Healey et al., 2014) from young leaf tissue. The single plant was one of the 23 progeny from the original λ -4 transformant and was homozygous for the large insertion on chromosome 2. The PacBio library was prepared following SMRTbell library guidelines. The library was sequenced with three single-molecule real-time (SMRT) cells to generate 10.32 Gb of long reads with N50s ranging from 16 to 18 kb. Consensus sequences were created from subread BAM files using SMRTLink (version 5.1) with parameters: min_length 50, max_length 30,000, $minPredictedAccuracy$ 0.8, $minZScore$ -3.4, $minPasses$ 0, $maxDropFraction$ 0.34, and *polish*. The derived consensus reads were then mapped to lambda sequence with Basic Local Alignment with Successive Refinement (BLASR; Chaisson and Tesler, 2012) at default settings. The BAM output file of mapped reads was converted to fasta format by samtools (version 1.3.1) and aligned against lambda full sequence using NCBI blast (version 2.2.26) at default parameters. The blast output was filtered by removing reads with E-values higher than 0.1 and lengths shorter than 30 bp, and by retaining the longest consecutive matches for each read using an in-house script.

Code Availability

The custom code required for analysis in this study is available at the GitHub repository (<https://github.com/dawelab/Genome-Rearrangements>).

Accession Numbers

Raw PacBio and Illumina sequence data are available from the National Center for Biotechnology Information Short Read Archive under BioProject PRJNA508943.

Supplemental Data

- Supplemental Figure 1.** Circos plots of additional rice lines transformed with λ and plasmid pVUbi2H.
- Supplemental Figure 2.** Additional data from maize lines transformed with λ and plasmid pBAR184.
- Supplemental Figure 3.** Linkage analysis of fragments from the 1.6 Mb array of rice λ -4 in self-pollinated progeny.
- Supplemental Figure 4.** Distributions of microhomology at junction sites and relative orientations of rejoined fragments.
- Supplemental Figure 5.** Three major intra-chromosomal SV types and the strand orientations of paired-end reads.
- Supplemental Figure 6.** Additional data from rice lines transformed with plasmid pANIC10A-OsFPGS1.
- Supplemental Figure 7.** Additional data from rice lines transformed with plasmid pANIC12A-OsFPGS1.
- Supplemental Table 1.** Copy number of lambda and co-bombarded plasmid in rice/maize transgenic genome.
- Supplemental Table 2.** Sensitivity and precision evaluation of SV detection pipeline by simulation.
- Supplemental Table 3.** Evidence of HDR in non-repetitive regions in rice transgenic events.
- Supplemental Table 4.** Evidence of HDR in non-repetitive regions in maize transgenic events.
- Supplemental Table 5.** Copy number of introduced molecules (single plasmid) and number of breakpoints in rice transgenic genome.
- Supplemental Data Set.** Copy number variation in all transgenic events.

ACKNOWLEDGMENTS

We thank Peter Lafayette, Melissa McGranahan, and Gary Orr for help with rice biolistic transformation. The Georgia Genomics and Bioinformatics Core facility carried out sequencing and optical mapping, and Jonathan Gent and Magdy Alabady provided advice on bioinformatic analyses. We also thank Jean-Michel Michno and Jonathan Gent for critically reading the manuscript. This study was supported by National Science Foundation (NSF) grants (1444514 to R.K.D. and 1400616 to N.J.N.).

AUTHOR CONTRIBUTIONS

N.J.N. and R.K.D. envisioned and oversaw the project; N.J.N. screened and chose lines for study; F.-F.F. prepared DNA and sequencing libraries; J.L. carried out bioinformatic analyses and interpreted the results; J.S. interpreted Bionano data; W.A.P. supervised rice transformation; B.A. performed rice tissue culture; J.L., N.J.N., and R.K.D. wrote the paper; J.L. and F.-F.F. produced figures.

Received August 15, 2018; revised December 13, 2018; accepted January 9, 2019; published January 16, 2019.

REFERENCES

Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize

- typical and atypical CNVs from family and population genome sequencing. *Genome Res.* **21**: 974–984.
- Altpeter, F., et al.** (2005). Particle bombardment and the genetic enhancement of crops: Myths and realities. *Mol. Breed.* **15**: 305–327.
- Altpeter, F., et al.** (2016). Advancing Crop Transformation in the Era of Genome Editing. *Plant Cell* **28**: 1510–1520.
- Anderson, J.E., Michno, J.-M., Kono, T.J.Y., Stec, A.O., Campbell, B.W., Curtin, S.J., and Stupar, R.M.** (2016). Genomic variation and DNA repair associated with soybean transgenesis: A comparison to cultivars and mutagenized plants. *BMC Biotechnol.* **16**: 41.
- Bankevich, A., et al.** (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.
- Begemann, M.B., Gray, B.N., January, E., Gordon, G.C., He, Y., Liu, H., Wu, X., Brutnell, T.P., Mockler, T.C., and Oufattole, M.** (2017). Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Sci. Rep.* **7**: 11606.
- Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Patron, N.J., and Nekrasov, V.** (2015). Editing plant genomes with CRISPR/Cas9. *Curr. Opin. Biotechnol.* **32**: 76–84.
- Campbell, P.J., et al.** (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**: 1109–1113.
- Casjens, S.R., and Hendrix, R.W.** (2015). Bacteriophage lambda: Early pioneer and still relevant. *Virology* **479–480**: 310–330.
- Ceccaldi, R., Rondinelli, B., and D'Andrea, A.D.** (2016). Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* **26**: 52–64.
- Chabchoub, E., Rodríguez, L., Galán, E., Mansilla, E., Martínez-Fernández, M.L., Martínez-Frías, M.L., Frys, J.-P., and Vermeesch, J.R.** (2007). Molecular characterisation of a mosaicism with a complex chromosome rearrangement: Evidence for coincident chromosome healing by telomere capture and neo-telomere formation. *J. Med. Genet.* **44**: 250–256.
- Chaisson, M.J., and Tesler, G.** (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): Application and theory. *BMC Bioinformatics* **13**: 238.
- Clark, K.A., and Krysan, P.J.** (2010). Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines. *Plant J.* **64**: 990–1001.
- Clarke, J.D.** (2009). Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.* **2009**: db.prot5177.
- Cluster, P.D., O'Dell, M., Metzloff, M., and Flavell, R.B.** (1996). Details of T-DNA structural organization from a transgenic Petunia population exhibiting co-suppression. *Plant Mol. Biol.* **32**: 1197–1203.
- Crasta, K., Ganem, N.J., Dagher, R., Lantermann, A.B., Ivanova, E. V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D., and Pellman, D.** (2012). DNA breaks and chromosome pulverization from errors in mitosis. *Nature* **482**: 53–58.
- Ercolano, M.R., Ballvora, A., Paal, J., Steinbiss, H.-H., Salamini, F., and Gebhardt, C.** (2004). Functional complementation analysis in potato via biolistic transformation with BAC large DNA fragments. *Mol. Breed.* **13**: 15–22.
- Frame, B.R., Zhang, H., Cocciolone, S.M., Sidorenko, L.V., Dietrich, C.R., Pegg, S.E., Zhen, S., Schnable, P.S., and Wang, K.** (2000). Production of transgenic maize from bombarded type II callus: Effect of gold particle size and callus morphology on transformation efficiency. *In Vitro Cell. Dev. Biol. Plant* **36**: 21–29.
- Fu, X., Duc, L.T., Fontana, S., Bong, B.B., Tinjuangjun, P., Sudhakar, D., Twyman, R.M., Christou, P., and Kohli, A.** (2000). Linear transgene constructs lacking vector backbone sequences generate low-copy-number transgenic plants with simple integration patterns. *Transgenic Res.* **9**: 11–19.

- Gelvin, S.B.** (2017). Integration of *Agrobacterium* T-DNA into the plant genome. *Annu. Rev. Genet.* **51**: 195–217.
- Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C.V., Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., and Voytas, D.F.** (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant J.* **89**: 1251–1262.
- Glenn, K.C., et al.** (2017). Bringing new plant varieties to market: Plant breeding and selection practices advance beneficial characteristics while minimizing unintended changes. *Crop Sci.* **57**: 2906–2921.
- Gorbunova, V., and Levy, A.A.** (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res.* **25**: 4650–4657.
- He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., and Feng, B.** (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.* **44**: e85.
- Healey, A., Furtado, A., Cooper, T., and Henry, R.J.** (2014). Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods* **10**: 21.
- Heyer, W.-D., Ehmsen, K.T., and Liu, J.** (2010). Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**: 113–139.
- Huang, W., Li, L., Myers, J.R., and Marth, G.T.** (2012). ART: A next-generation sequencing read simulator. *Bioinformatics* **28**: 593–594.
- Jackson, M.A., Anderson, D.J., and Birch, R.G.** (2013). Comparison of *Agrobacterium* and particle bombardment using whole plasmid or minimal cassette for production of high-expressing, low-copy transgenic plants. *Transgenic Res.* **22**: 143–151.
- Jackson, S.A., Zhang, P., Chen, W.P., Phillips, R.L., Friebe, B., Muthukrishnan, S., and Gill, B.S.** (2001). High-resolution structural analysis of biolistic transgene integration into the genome of wheat. *Theor. Appl. Genet.* **103**: 56–62.
- Jiao, Y., et al.** (2017). Improved maize reference genome with single-molecule technologies. *Nature* **546**: 524–527.
- Jones, T.J., and Rost, T.L.** (1989). The developmental anatomy and ultrastructure of somatic embryos from rice (*Oryza sativa* L.) scutellum epithelial cells. *Bot. Gaz.* **150**: 41–49.
- Jupe, F., Rivkin, A.C., Michael, T.P., Zander, M., and Motley, T.S.** (2018). The complex architecture of plant transgene insertions. bioRxiv.
- Karanam, K., Kafri, R., Loewer, A., and Lahav, G.** (2012). Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol. Cell* **47**: 320–329.
- Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S., Childs, K.L., and Davidson, R.M., et al.** (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice (N. Y.)* **6**: 4.
- Kaya, H., Sato, S., Tabata, S., Kobayashi, Y., Iwabuchi, M., and Araki, T.** (2000). Hosoba toge toge, a syndrome caused by a large chromosomal deletion associated with a T-DNA insertion in *Arabidopsis*. *Plant Cell Physiol.* **41**: 1055–1066.
- Kessler, D.A., Taylor, M.R., Maryanski, J.H., Flamm, E.L., and Kahl, L.S.** (1992). The safety of foods developed by biotechnology. *Science* **256**: 1747–1749.
- Klein, T.M., Kornstein, L., Sanford, J.C., and Fromm, M.E.** (1989). Genetic transformation of maize cells by particle bombardment. *Plant Physiol.* **91**: 440–444.
- Kleinboelting, N., Huep, G., Appelhagen, I., Viehoveer, P., Li, Y., and Weisshaar, B.** (2015). The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. *Mol. Plant* **8**: 1651–1664.
- Kohli, A., Griffiths, S., Palacios, N., Twyman, R.M., Vain, P., Laurie, D.A., and Christou, P.** (1999). Molecular characterization of trans-forming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. *Plant J.* **17**: 591–601.
- Korbel, J.O., and Campbell, P.J.** (2013). Criteria for inference of chromothripsis in cancer genomes. *Cell* **152**: 1226–1236.
- Krizkova, L., and Hrouda, M.** (1998). Direct repeats of T-DNA integrated in tobacco chromosome: characterization of junction regions. *Plant J.* **16**: 673–680.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A.** (2009). Circo: An information aesthetic for comparative genomics. *Genome Res.* **19**: 1639–1645.
- Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M.** (2014). LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **15**: R84.
- Lee, M., and Phillips, R.L.** (1988). The chromosomal basis of somaclonal variation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**: 413–437.
- Leibowitz, M.L., Zhang, C.-Z., and Pellman, D.** (2015). Chromothripsis: A new mechanism for rapid karyotype evolution. *Annu. Rev. Genet.* **49**: 183–211.
- Li, H.** (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN]. <https://arxiv.org/abs/1303.3997>
- Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J., and Gao, C.** (2016). Gene replacements and insertions in rice by intron targeting using CRISPR-Cas9. *Nat. Plants* **2**: 16139.
- Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., and Gao, C.** (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleo-protein complexes. *Nat. Commun.* **8**: 14261.
- Lowe, B.A., Shiva Prakash, N., Way, M., Mann, M.T., Spencer, T.M., and Boddupalli, R.S.** (2009). Enhanced single copy integration events in corn via particle bombardment using low quantities of DNA. *Transgenic Res.* **18**: 831–840.
- Makarevitch, I., Svitashv, S.K., and Somers, D.A.** (2003). Complete sequence analysis of transgene loci from plants transformed via microprojectile bombardment. *Plant Mol. Biol.* **52**: 421–432.
- Mann, D.G., Lafayette, P.R., Abercrombie, L.L., King, Z.R., Mazarei, M., Halter, M.C., Poovaiah, C.R., Baxter, H., Shen, H., Dixon, R.A., Parrott, W.A., and Neal Stewart, C., Jr.** (2012). Gateway-compatible vectors for high-throughput gene functional analysis in switchgrass (*Panicum virgatum* L.) and other monocot species. *Plant Biotechnol. J.* **10**: 226–236.
- Mardin, B.R., et al.** (2015). A cell-based model system links chromothripsis with hyperploidy. *Mol. Syst. Biol.* **11**: 828.
- McClintock, B.** (1941). The stability of broken ends of chromosomes in *Zea Mays*. *Genetics* **26**: 234–282.
- McClintock, B.** (1942). The fusion of broken ends of chromosomes following nuclear fusion. *Proc. Natl. Acad. Sci. USA* **28**: 458–463.
- McVey, M., and Lee, S.E.** (2008). MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet.* **24**: 529–538.
- Nacry, P., Camilleri, C., Courtial, B., Caboche, M., and Bouchez, D.** (1998). Major chromosomal rearrangements induced by T-DNA transformation in *Arabidopsis*. *Genetics* **149**: 641–650.

- Pannunzio, N.R., Watanabe, G., and Lieber, M.R.** (2018). Non-homologous DNA end joining for repair of DNA double-strand breaks. *J. Biol. Chem.* **293**: 10512–10523.
- Partier, A., Gay, G., Tassy, C., Beckert, M., Feuillet, C., and Barret, P.** (2017). Molecular and FISH analyses of a 53-kbp intact DNA fragment inserted by biolistics in wheat (*Triticum aestivum* L.) genome. *Plant Cell Rep.* **36**: 1547–1559.
- Pawlowski, W.P., and Somers, D.A.** (1996). Transgene inheritance in plants genetically engineered by microprojectile bombardment. *Mol. Biotechnol.* **6**: 17–30.
- Pawlowski, W.P., and Somers, D.A.** (1998). Transgenic DNA integrated into the oat genome is frequently interspersed by host DNA. *Proc. Natl. Acad. Sci. USA* **95**: 12106–12110.
- Phan, B.H., Jin, W., Topp, C.N., Zhong, C.X., Jiang, J., Dawe, R.K., and Parrott, W.A.** (2007). Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. *Transgenic Res.* **16**: 341–351.
- Raji, J.A., Frame, B., Little, D., Santoso, T.J., and Wang, K.** (2018). Agrobacterium- and Biolistic-Mediated Transformation of Maize B104 Inbred. *Methods Mol. Biol.* **1676**: 15–40.
- Register III, J.C., et al.** (1994). Structure and function of selectable and non-selectable transgenes in maize after introduction by particle bombardment. *Plant Mol. Biol.* **25**: 951–961.
- Rode, A., Maass, K.K., Willmund, K.V., Lichter, P., and Ernst, A.** (2016). Chromothripsis in cancer cells: An update. *Int. J. Cancer* **138**: 2322–2333.
- Shi, J., Gao, H., Wang, H., Lafitte, H.R., Archibald, R.L., Yang, M., Hakimi, S.M., Mo, H., and Habben, J.E.** (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnol. J.* **15**: 207–216.
- Shou, H., Frame, B.R., Whitham, S.A., and Wang, K.** (2004). Assessment of transgenic maize events produced by particle bombardment or Agrobacterium-mediated transformation. *Mol. Breed.* **13**: 201–208.
- Stephens, P.J., et al.** (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**: 27–40.
- Storchová, Z., and Kloosterman, W.P.** (2016). The genomic characteristics and cellular origin of chromothripsis. *Curr. Opin. Cell Biol.* **40**: 106–113.
- Svitashev, S., Ananiev, E., Pawlowski, W.P., and Somers, D.A.** (2000). Association of transgene integration sites with chromosome rearrangements in hexaploid oat. *Theor. Appl. Genet.* **100**: 872–880.
- Svitashev, S.K., Pawlowski, W.P., Makarevitch, I., Plank, D.W., and Somers, D.A.** (2002). Complex transgene locus structures implicate multiple mechanisms for plant transgene rearrangement. *Plant J.* **32**: 433–445.
- Svitashev, S., Young, J.K., Schwartz, C., Gao, H., Falco, S.C., and Cigan, A.M.** (2015). Targeted mutagenesis, precise gene editing, and site-specific gene insertion in maize using Cas9 and guide RNA. *Plant Physiol.* **169**: 931–945.
- Takano, M., Egawa, H., Ikeda, J.E., and Wakasa, K.** (1997). The structures of integration sites in transgenic rice. *Plant J.* **11**: 353–361.
- Tan, E.H., Henry, I.M., Ravi, M., Bradnam, K.R., Mandakova, T., Marimuthu, M.P., Korf, I., Lysak, M.A., Comai, L., and Chan, S.W.** (2015). Catastrophic chromosomal restructuring during genome elimination in plants. *eLife* **4**: e06516.
- Tassy, C., Partier, A., Beckert, M., Feuillet, C., and Barret, P.** (2014). Biolistic transformation of wheat: increased production of plants with simple insertions and heritable transgene expression. *Plant Cell Tissue Organ Cult.* **119**: 171–181.
- Udall, J.A., and Dawe, R.K.** (2018). Is it ordered correctly? Validating genome assemblies by optical mapping. *Plant Cell* **30**: 7–14.
- van Kregten, M., de Pater, S., Romeijn, R., van Schendel, R., Hooykaas, P.J.J., and Tijsterman, M.** (2016). T-DNA integration in plants results from polymerase- θ -mediated DNA repair. *Nat. Plants* **2**: 16164.
- Verma, P., and Greenberg, R.A.** (2016). Noncanonical views of homology-directed DNA repair. *Genes Dev.* **30**: 1138–1154.
- Weber, N., Halpin, C., Curtis Hannah, L., Jez, J.M., Kough, J., and Parrott, W.** (2012). Crop Genome Plasticity and Its Relevance to Food and Feed Safety of Genetically Engineered Breeding Stacks. *Plant Physiol.* **160**: 1842–1853.
- Zakov, S., Kinsella, M., and Bafna, V.** (2013). Detecting Breakage Fusion Bridge cycles in tumor genomes—an algorithmic approach. *Proc. Natl. Acad. Sci. USA* **110**: 5546–5551.
- Zeitouni, B., Boeva, V., Janoueix-Lerosey, I., Loeillet, S., Legoix-né, P., Nicolas, A., Delattre, O., and Barillot, E.** (2010). SVDetect: A tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics* **26**: 1895–1896.
- Zhang, C.-Z., Spektor, A., Cornils, H., Francis, J.M., Jackson, E.K., Liu, S., Meyerson, M., and Pellman, D.** (2015). Chromothripsis from DNA damage in micronuclei. *Nature* **522**: 179–184.
- Zhang, H., Phan, B.H., Wang, K., Artelt, B.J., Jiang, J., Parrott, W. A., and Dawe, R.K.** (2012). Stable integration of an engineered megabase repeat array into the maize genome. *Plant J.* **70**: 357–365.
- Zhu, C., Wu, J., and He, C.** (2010). Induction of chromosomal inversion by integration of T-DNA in the rice genome. *J. Genet. Genomics* **37**: 189–196.