The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes

Ksenia V Krasileva

Plant genomes are shaped by structural variation. Gene-size insertions and among most prominent events and can have significant effects on amplification of gene families as well as facilitate new gene fusions. Transposable elements as well as plant DNA repair machinery have overlapping contributions to these events, and often work in synergy. Activity of transposable elements is often lineage specific and can preferentially affect specific gene families, such as disease resistance genes. Once duplicated, genes themselves can serve templates for additional variation that can arise from non-allelic homologous recombination. Non-homologous DNA repair mechanisms contribute to additional variation and diversify the mechanisms of gene movement, such as through ligation of extra-chromosomal DNA fragments. Genomic processes that generate structural variation can be induced by stress and, therefore, can provide adaptive potential. This review describes mechanisms that contribute to gene-size structural variation in plants, result in gene duplication and generation of new plant genes through gene fusion.

Address
Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, United States

Corresponding author: Krasileva, Ksenia V (kseniak@berkeley.edu)

Introduction

Plant genomes evolve through mutation, recombination, as well as structural variation (SV) that ranges from large scale chromosome re-assortment to gene-size insertions. The global gene order in plant genomes is conserved in large syntenic blocks and large-scale re-arrangements are relatively rare. In contrast, gene-size insertions are more common. Genes or gene fragments can duplicate, move around the genome, and form gene fusions [1••,2,3]. Genome plasticity allows plants to adapt to new environments and played a large role in crop domestication [4]. Contiguous assemblies of complex plant genomes and pan-genome projects are now revealing the extent of SV in plants [5–8].

The SV in plant genomes has been decoded using two main approaches. First, a comparison of syntenic blocks or orthologous genes can identify genes and regions that disrupt synteny. It is often used in whole genome projects and has been applied to understand variation in gene content within and across species. Second approach involves analysis of a particular gene family or class of transposable elements. There are several plant gene families that showcase SV. Tandem duplication played significant role in lineage specific expansions of genes that determine agronomic traits, such as lectins [9], papains [10] and gluten [11,12]. The plant immune receptors of Nucleotide Binding Leucine Rich repeat domain (NLRs) vary in gene number from 30 in cucumber to 2500 in wheat [13]. NLRs are often found in gene clusters with both tandem and inverted duplications (Figure 1) and participate in gene conversions, as well as fusions with other plant genes that become attached to NLR as integrated domains (IDs) [14–16]. Although the exact mechanism of NLR-ID formation is unknown, it involves gene duplication and inter-chromosomal translocation of either NLRs or IDs [2,3]. Whether comparing whole genomes or particular gene families, studies often reveal that different mechanism have contributed to the genomic landscapes that we see today.

Transposable elements (TE) are major drivers of plant genome evolution. TEs have been shown to facilitate duplication or deletion of plant genes, affect their expression or function, and combine genes from different locations into new fusions [17]. TEs come in two major classes, retrotransposons that move through an RNA intermediate, and DNA transposons that move directly as DNA [18]. TEs can be autonomous, encoding all enzymes needed for transposition or non-autonomous encoding the recognition sequences but relying on enzymes from elsewhere in the genome. Amplification of active elements contributes to genome size. The content of TEs in plant genomes varies from 3% in the genome of aquatic plant *Utricularia gibba* [19] to over 85% in complex polyploid plants, such as wheat [20]. Long Terminal Repeat (LTR) retrotransposons contribute most to the expansion of TE content [21], constituting over 65% of 17 Gb wheat genome [20]. Other TEs, such as PacMULEs

Figure 1
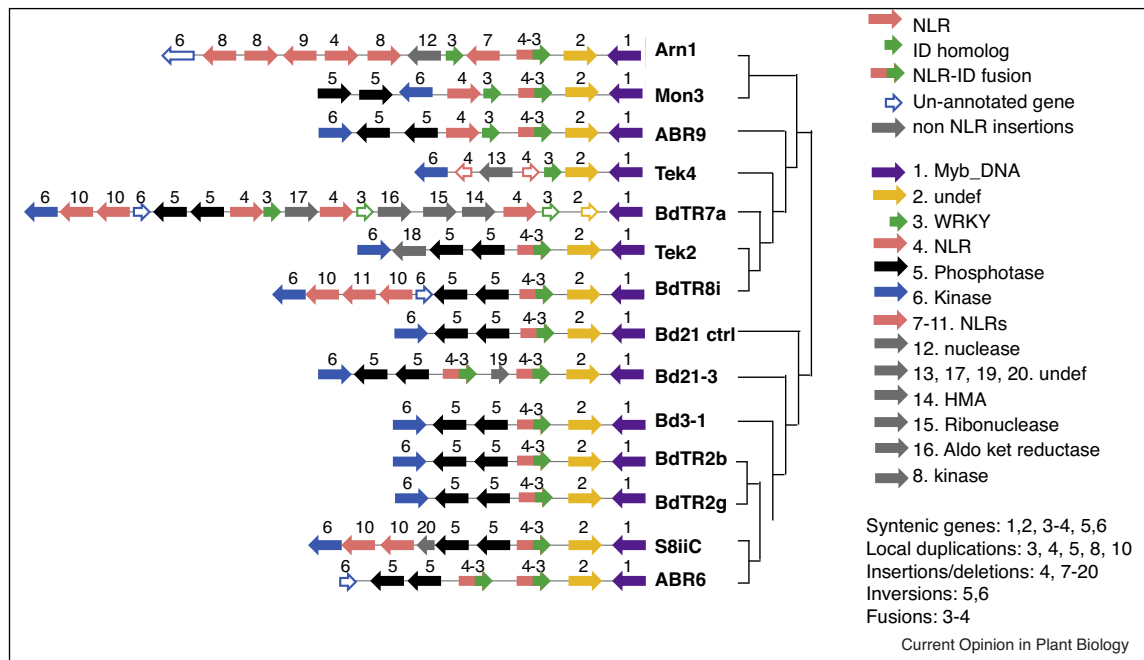
Gene duplications, insertions, and deletions in a genomic locus across 14 *Brachypodium* accessions.
The overall relatedness of 14 *Brachypodium* accessions based on whole genome analysis [5] is displayed on the right with accession names next to the branches. The analyzed 35 kb region surrounding NLR-WRKY fusion (Bradi4g09886) is located on chromosome 4 (chr4:9 395 000–9 430 000) in the reference Bd21 accession. Gene numbers correspond to orthologs/paralogs in following categories: 1–6 syntenic genes (purple, yellow, salmon-green displaying a fusion, blue and black), 7–11 exogenous NLR insertions (salmon), 12–18 exogenous non-NLR insertions (grey). Filled arrows represent annotated proteins, empty arrows are genes and gene fragments confirmed to be present in genomic assembly by BLAST, but absent in protein annotation.

and Helitrons contribute relatively small percentage to the whole genome content (from 0.1% in wheat [20] to 2.2% in maize [22]), but have a profound effect on gene composition due to their ability to trap genes and move them around the genome and facilitate novel domain combinations through gene fusions [23••,24••,25•]. Noteworthy, active TEs are also often associated with loci controlling adaptive responses, such as abiotic and biotic stress, suggesting their beneficial role in genome evolution [26].

Gene-size insertions in plants can also be a result of DNA repair mechanisms [1••,27,28••]. Repetitive DNA as well as duplicated genes provide templates for non-allelic homologous recombination (NAHR) that occurs during the loss of recombination control [29]. Another DNA repair process, non-homologous end joining (NHEJ) can ligate double stranded breaks in genome without homologous template [30]. NHEJ is a predominant process that repairs double stranded DNA breaks in plants and can lead to either deletions during the recess of broken DNA or insertions if a broken strand is filled in [1••]. Plant TEs and DNA repair machinery often work together and can produce similar SV outcomes. This is well exemplified in recent discovery of extra-chromosomal circular DNA molecules (eccDNA), that can arise from looping out of tandemly

duplicated elements in the genome or direct ligation of retrotranscribed TEs by NHEJ [31••].

This review summarizes some of the mechanisms that contribute to evolution of gene content in plant genomes, with a focus on gene-size SV, such as insertions, duplications, translocations and fusions of genes and gene fragments.

Different scales of genome re-arrangements

The SV can be divided into two groups based on the length of affected genomic segments. Large-scale SV, such as chromosomal translocations are relatively rare in plants; however, they are well documented in model species as well as crops. Large chromosomal translocations are often associated with initiation of DNA breaks that occurred during natural history of plant genomes [32] or induced by mutagenesis and T-DNA insertions [33]. Large scale translocations can affect gene expression, and genes directly at the breaks. Large-scale translocation events are often implied from disrupted synteny of large genomic blocks.

SV on a smaller, kilobase scale (0.1 kb–1 Mb) influences tandem gene duplication, translocation, and new gene

fusions. Synteny analysis between *Arabidopsis thaliana* and *Arabidopsis lyrata* estimated 420 individual genes to be affected by insertions/deletions [28••]. Analysis of 453 rice accessions revealed a total of nearly 90 000 SV of 5–500 kb in size with several thousands of SV affecting genes [6]. In wheat as many as 27% of its nearly 100 000 genes are found as tandem duplicates [20].

Such extensive SV is exciting because it allows to evaluate the plasticity of different genes and their genomic context with potential of targeted manipulation. To present a more detailed example of gene-size SV, we examined a 35 kb locus in *Brachypodium* pan-genome [5] (chr4:9 395 000–9 430 000) that contains one of validated NLR-WRKY gene fusions [2,3] (Figure 1). Our analysis of this region showed that only six genes were maintained in synteny, while thirteen genes have been inserted from elsewhere in the genome. Analysis of this region in *Brachypodium* confirms rapid evolution of the same region in rice [2]. Such extensive SV likely due to combined activity of different genomic processes, that include active TEs, replication, recombination, and DNA damage repair machinery.

Gene duplication and movement facilitated by retro-transposons

Autonomous and non-autonomous retrotransposons constitute major part of plant genomes. [18]. Among the most prominent and active TEs that have been shown to capture genes are LTR retrotransposons with characteristic long terminal repeats [34,35]. The gene capture by LTRs can occur during the reverse transcription when a template switches to a messenger RNA of another plant gene. Upon insertion back into the genome, the gene will be duplicated without its introns and will be now surrounded by LTR sequences (Figure 2a). Identification of retroposed genes affected by LTRs, therefore, is relatively straight forward. An example of gene family expansion through retro-transposition is subclade specific expansion of NLR immune receptors in peppers [34] (Table 1). Several next generation sequencing based methods have been developed to isolate and sequence active LTRs [31••,36]. Global analyses of LTR activity show that they can be induced by stress [31••,36], so the genes trapped by LTRs might have an adaptive advantage.

Tandem repeats of retrotransposons also facilitate tandem gene duplications that does not involve retroposition. In this case, terminal repeats can align to induce gene excision or gene duplication that is mediated by host NAHR. NAHR can occur due to the loss of replication control as have been well documented in other eukaryotes, including yeast [37,38] and animals [29]. Because NAHR breakpoints occur at homologous locations, they are often undetected and finding them requires specialized algorithms [29].

Gene capture and novel gene fusions inside DNA transposons

The two main DNA elements that can trap multiple gene fragments facilitating new domain architectures are Pac-MULEs and Helitrons. Pac-MULEs were first identified in rice and have been shown to contain gene fragments as well as chimeric gene fusions [24••] (Table 1). Pac-MULEs preferentially acquire GC-rich fragments and insert into 5′ ends of genes, facilitating formation of new exons and gene fusions [25•]. Pac-MULEs are Mutator family Class II DNA elements that leave terminal inverted repeats (TIRs) surrounding their insertions (Figure 2b). Identification of Pac-MULEs, therefore, is based on TIRs. Genome-wide identification of Pac-MULEs often relies on RepeatMasker [39] to identify Mutator families of TEs and subsequent custom analyses of captured genes and gene fragments.

Helitrons have been discovered relatively recently, and they do not have repeats on their ends [40] (Table 1). Instead, they begin with CTRR (most sequence start with CTAG) and end with a short palindromic repeat [40]. Notably, Helitrons can capture multiple genes and produce chimeric transcripts [41]. Upon alternative splicing, such transcripts result in novel chimeric protein fusions [42,43]. Several mechanisms of gene capture by Helitrons have been proposed [40,41,44••]. One mode by which Helitrons can capture genes involves a readthrough during transposition where the first the terminal element is either lost or not-functional and the machinery continues to read through the neighboring gene until the next terminal element is encountered (Figure 2c). The second proposed mode of gene capture is double-stranded break repair with an exogenous gene providing repair template usually through a very short matching sequence (Figure 2c). The two modes can be clearly distinguished only if a gene capture occurred very recently (e.g. the original Helitron is present upstream of parental gene). Helitrons are challenging to identify because they do not have terminal inverted repeat and do not replicate host sequence upon insertion. Genome-wide identification tools of Helitrons rely on DNA structure prediction or include local combinational variable analyses [45].

DNA transposons can also participate in generation of tandemly duplicated genes through reserve end transposition as documented for maize Ac elements [46•]. This occurs when TEs are positioned next to each other in correct orientation, which triggers an alternative form of transposition that deploying the terminal repeats of two different elements. A detailed overview of reverse end transposition is depicted in Ref. [46•].

Gene duplication and conversion through non-allelic homologous recombination

Plant TEs often work in synergy with host recombination and DNA damage machinery. Several studies have

Figure 2
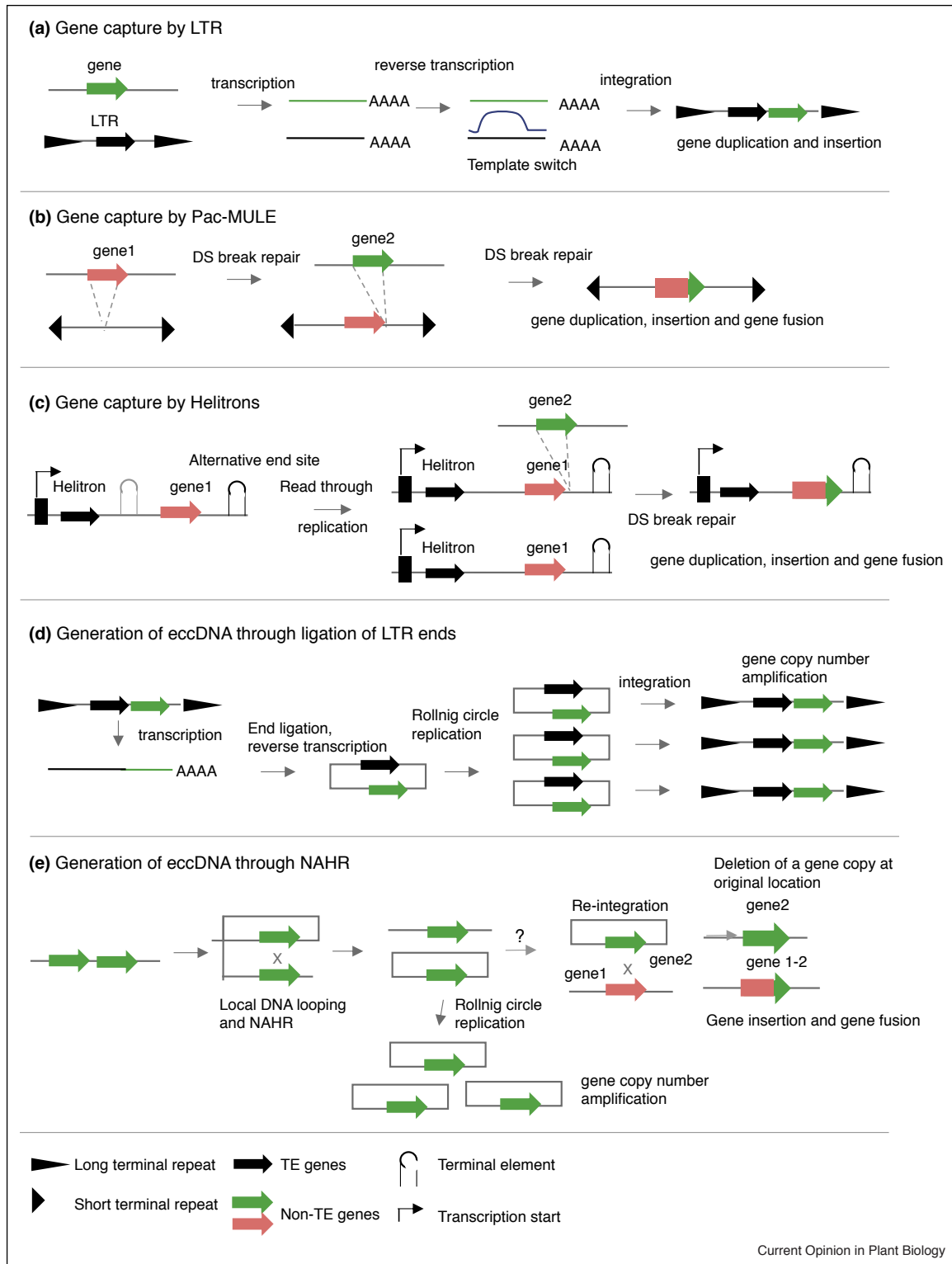
Schematic diagram of the mechanisms that lead to gene duplications and gene fusions.
(a) Gene capture by LTR during the reverse transcription stage: both LTR and other plant genes are first transcribed into mRNA; during reverse transcription of LTR, there is a template switch and mRNA of another gene is used which creates DNA chimera. Re-integration back into the genome traps a gene or its part within an LTR. **(b)** Sequential gene capture by Pac-MULE which creates new gene fusion: during DS repair, priming to another gene is used as template which leads to integration of a gene within MULE, if a process continues, another gene gets trapped,

Table 1

Known gene insertion mechanisms

Name	Consequence	Footprint	References
Transposable elements			
LTR	Gene duplication, gene movement, tandem duplications, eccDNA	No introns in moved genes, LTR signature near tandemly duplicated genes	Kim *et al*. [34]
Pack-MULEs	Gene duplication and movement, chimeric gene fusion	Terminal inverted repeats	Zhang *et al*. [46•] Jiang *et al*. [24••] Jiang *et al*. [25•]
Helitrons	Gene duplication and movement, gene fusions	LTS/RTS	Kapitonov and Jurka [40] Lai *et al*. [23••] Xiong *et al*. [45]
Host machinery			
BER + NHEJ	Tandem duplication	Tandem repeats	Vaughn and Bennetzen [57]
NAHR	Amplification of tandemly duplicated genes, gene deletions or gene conversions	Tandem repeats	Assaad and Singer [47] Jelesko *et al*. [50] Zhang *et al*. [46•] Costello *et al*. [62]
NAHR + ectopic recombination	Recombination between tandemly duplicated sequences, eccDNA formation, and gene insertion at ectopic location	Flanked by direct repeats orthologous to the origin	Woodhouse *et al*. [28••] Wicker *et al*. [1••] Salomon *et al*. [63]
NHEJ	Large scale translocations, chromosome fusions	Nucleotide recession (loss), nucleotide addition (filler)	Ma *et al*. PNAS 2004 [64] Wicker *et al*. [1••]

concluded that most gene movements in plants cannot be explained by transposable element activity alone [1••,3,28••] (Table 1). A pair of directly repeated genes even without other TEs nearby can lead to gene excision or further duplications through NAHR. Early experiments with reporter genes in *Arabidopsis* demonstrated that even in the absence of transposable elements, crossing over between repeated genic sequences can either increase or decrease copy number [47]. Recent studies showed that NAHR can affect more than one gene in a region [48]. Besides gene duplication, NAHR can lead to new chimeric combinations of repeated copies [49–52] (Table 1). In addition, NAHR can also induce large scale translocations and chromosome fusions when repeated sequences are located on different chromosomes. For schematic examples of outcomes of NAHR, see Ref. [29].

It is yet unclear if inter-chromosomal NAHR might be also guided by proximity of non-allelic homologs in 3D as could be in future determined by techniques such as Hi-C [53,54]. NAHR can often go undetected because it involves exchange of closely related sequences and requires probabilistic models [29]. Multiplex ligation-dependent probe amplification has been instrumental in identifying gene copy number in absence of contiguous genome assemblies [55].

De novo repeat generation and non-homologous DNA repair

NHEJ is one of the main DNA repair mechanisms that participates in many gene movement and gene fusions processes, either initiated by TEs, such as discussed above or during the DNA damage induced by cellular enzymes, UV and other mutagens. The outcomes of NHEJ can be large scale translocations when ends of different chromosomal segments are joined together or small SV when repair is local. NHEJ plays a critical role in gene translocations and chimeric gene fusions when the whole chromosome is undergoing catastrophy. Catastrophic reconstruction of chromosomes is critical during genome elimination and can also create new gene fusions [56]. Intriguingly, analyses of rice genomes implicated NHEJ in most cases of gene transpositions [27] (Table 1).

NHEJ is also central to generation of tandem duplications without activity of TEs. Two main mechanisms have previously been implicated in this process: replication slippage and long-patch base excision repair (BER) mechanism. Replication slippage depends on microhomology between the end of duplicated region and beginning of the donor site. Long-patch BER depends on double stranded DNA breaks with complementary overhangs that are filled in using complementary strand, and end products are ligated with NHEJ [57]. A study of natural

(Figure 2 Legend Continued) the element then can propagate new fusion. (c) Gene capture by Helitron: capture can occur either by readthrough during element replication or during DS repair, the result can trap one or more genes inside the element and create new fusions. (d) Generation of eccDNA from TEs using an example of a gene trapped within LTR from (a): After transcription and reverse transcription of and LTR with a gene trapped inside, the end of the linear DNA can be ligated by NHEJ creating circular product. eccDNA can replicate amplifying TE and any trapped gene, providing platform for mutation, and subsequently re-integrate into the genome. (e) Generation of eccDNA from tandemly duplicated gene: similar to any repeat, tandemly duplicated gene can loop out and through NAHR form a circular product leaving a deletion in the genome. The eccDNA can amplify and possibly re-integrate elsewhere in the genome, forming gene insertions and gene fusions.

gene insertions in rice showed that long-patch base excision repair mechanism is prevalent and accounts for most recent insertions [57] (Table 1).

Gene movement and amplification on eccDNA

Multiple roles of NHEJ are exemplified by generation of eccDNA. The eccDNA molecules are present in both plants and animals and vary in size from 2000 to over 20 000 base pairs [58]. The eccDNA can arise by NHEJ of reverse transcription products of retroelements (Figure 2d) [31••] or directly from genome when during DNA repair, NHEJ can also create circular DNA products from any tandemly duplicated sequences, either TEs (Figure 2e) [46•], other repeats or duplicated genes. The resulting eccDNA can replicate multiplying copies of trapped sequences and providing platform for mutation. Traditionally, targets on eccDNA can be isolated by rolling circle amplification. With advancements of sequencing technologies, the whole eccDNA repertoire can be sequenced. The circular eccDNA are protected from exonuclease and, therefore, can be separated from genomic DNA by digestion and amplified through rolling circle amplification using random primers [31••].

While most of eccDNA contains non-coding repetitive elements, further research is needed to examine genic content of eccDNA on the genome scale. Interestingly, among the genes reported to be propagated on eccDNA is *5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE* (*EPSPS*), the molecular target of glyphosate and the gene responsible for transmission of herbicide resistance in crops [59•]. The mutated *EPSPS* has been found in high copy number on autonomously replicating eccDNA, suggesting that eccDNA can also serve as an important reservoir of producing rapid somatic variation that can then be transmitted to germ line.

Comparative genomics of *A. thaliana* and *A. lyrata* proposed that recombination across tandem gene copies can lead to formation of eccDNA containing one of the repeated elements as well as a deletion in the source (Figure 2e). The circular DNA can then move to another location and recombine with a similar repeat elsewhere in the genome, including other genes. This implies that in principle, genes from eccDNA could integrate back to genome. The result could be a gene insertion or generation of novel gene fusion [5]; however, this process has yet to be demonstrated. Advances of single molecule sequencing technologies should allow now to investigate generation and evolution of eccDNA, their content and ability to integrate back to genome.

Conclusions

Gene-size insertions are one of the main drivers of small-scale SV and gene content variation in plant genomes. The same outcomes can be generated through a variety of mechanisms, involving TEs and DNA repair machinery.

While we have a solid understanding of each individual process that can generate SV, a comparative analysis of their contributions across plant lineages is only becoming possible now due to large number of high-quality genomes and pan-genome projects. Tools that can process large scale dataset and integrate different types of SV signatures are becoming available [60]. Even within a lineage, specific TE elements can be dominant [61••], have specificity toward a particular gene family or even subfamily [34] or show activity under specific stress conditions [36,59•]. Understanding when and how each process is activated, and its specificity can inform not only natural history of plant genomes but also more efficient genome editing tools.

Conflict of interest statement

Nothing declared.

Acknowledgements

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Wicker T, Buchmann JP, Keller B: **Patching gaps in plant**
•• **genomes results in gene movement and erosion of colinearity**. *Genome Res* 2010, **20**:1229-1237 http://dx.doi.org/10.1101/gr.107284.110.
This study examines gene-size insertions in grasses, and it is one of the few studies that systematically document different mechanisms of gene movement and their relative contributions. While some of the gene insertions are bordered by transposable elements, others are likely to arise from template slippage or unequal crossing over.

2. Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL *et al.*: **Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus Oryza**. *Nat Genet* 2018, **50**:285-296 http://dx.doi.org/10.1038/s41588-018-0040-0.

3. Bailey PC, Schudoma C, Jackson W, Baggs E, Dagdas G, Haerty W, Moscou M, Krasileva KV: **Dominant integration locus drives continuous diversification of plant immune receptors with exogenous domain fusions**. *Genome Biol* 2018, **19** http://dx.doi.org/10.1186/s13059-018-1392-6.

4. Dubcovsky J, Dvorak J: **Genome plasticity a key factor in the success of polyploid wheat under domestication**. *Science (80-)* 2007, **316**:1862-1866 http://dx.doi.org/10.1126/science.1143986.

5. Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, Shu S, Stritt C, Roulin AC, Schackwitz W, Tyler L *et al.*: **Extensive gene content variation in the Brachypodium distachyon pan-genome correlates with population structure**. *Nat Commun* 2017, **8**:2184.

6. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F *et al.*: **Genomic variation in 3,010 diverse accessions of Asian cultivated rice**. *Nature* 2018, **557**:43-49 http://dx.doi.org/10.1038/s41586-018-0063-9.

7. Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T et al.: **Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice**. *Nat Genet* 2018, **50**:278-284 http://dx.doi.org/10.1038/s41588-018-0041-z.

8. Hübner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, Ziegler E, Lee JS, Baute GJ, Owens GL, Grassa CJ et al.: **Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance**. *Nat Plants* 2018, **5**:54-62 http://dx.doi.org/10.1038/s41477-018-0329-0.

9. Jiang S-Y, Ma Z, Ramachandran S: **Evolutionary history and stress regulation of the lectin superfamily in higher plants**. *BMC Evol Biol* 2010, **10**:79.

10. Liu J, Sharma A, Niewiara MJ, Singh R, Ming R, Yu Q: **Papain-like cysteine proteases in Carica papaya: Lineage-specific gene duplication and expansion**. *BMC Genomics* 2018, **19** http://dx.doi.org/10.1186/s12864-017-4394-y.

11. Huo N, Dong L, Zhang S, Wang Y, Zhu T, Mohr T, Altenbach S, Liu Z, Dvorak J, Anderson OD et al.: **New insights into structural organization and gene duplication in a 1.75-Mb genomic region harboring the α-gliadin gene family in Aegilops tauschii, the source of wheat D genome**. *Plant J* 2017, **92**:571-583 http://dx.doi.org/10.1111/tpj.13675.

12. Okita TW, Cheesbrough V, Reeves CD: **Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences**. *J Biol Chem* 1985, **260**:8203-8213.

13. Baggs E, Dagdas G, Krasileva KV: **NLR diversity, helpers and integrated domains: making sense of the NLR IDentity**. *Curr Opin Plant Biol* 2017, **38**:59-67 http://dx.doi.org/10.1016/j.pbi.2017.04.012.

14. Sarris PF, Cevik V, Dagdas G, Jones JDG, Krasileva KV: **Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens**. *BMC Biol* 2016, **14** http://dx.doi.org/10.1186/s12915-016-0228-7.

15. Kroj T, Chanclud E, Michel-Romiti C, Grand X, Morel JB: **Integration of decoy domains derived from protein targets of pathogen effectors into plant immune receptors is widespread**. *New Phytol* 2016, **210**:618-626 http://dx.doi.org/10.1111/nph.13869.

16. Cesari S, Bernoux M, Moncuquet P, Kroj T, Dodds PN: **A novel conserved mechanism for plant NLR protein pairs: the "integrated decoy" hypothesis**. *Front Plant Sci* 2014, **5** http://dx.doi.org/10.3389/fpls.2014.00606.

17. Bennetzen JL: **Transposable element contributions to plant gene and genome evolution**. *Plant Mol Biol* 2000, **42**:251-269 http://dx.doi.org/10.1023/A:1006344508454.

18. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al.: **A unified classification system for eukaryotic transposable elements**. *Nat Rev Genet* 2007, **8**:973-982 http://dx.doi.org/10.1038/nrg2165.

19. Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juárez MJA, Simpson J et al.: **Architecture and evolution of a minute plant genome**. *Nature* 2013, **498**:94-98 http://dx.doi.org/10.1038/nature12132.

20. Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J et al.: **Shifting the limits in wheat research and breeding using a fully annotated reference genome**. *Science (80-)* 2018, **361**:eaar7191 http://dx.doi.org/10.1126/science.aar7191.

21. Lee S-I, Kim N-S: **Transposable elements and genome size variations in plants**. *Genomics Inform* 2014, **12**:87 http://dx.doi.org/10.5808/GI.2014.12.3.87.

22. Yang L, Bennetzen JL: **Distribution, diversity, evolution, and survival of Helitrons in the maize genome**. *Proc Natl Acad Sci U S A* 2009, **106**:19922-19927 http://dx.doi.org/10.1073/pnas.0908008106.

23. Lai J, Li Y, Messing J, Dooner HK: **Gene movement by Helitron
•• transposons contributes to the haplotype variability of maize**. *Proc Natl Acad Sci U S A* 2005, **102**:9068-9073.

The paper shows that intraspecific erosion of gene collinearity in maize can be due to movement of four genes by Helitron. The ends of Helitron that are often elusive were reconstructed from the Helitron-minus maize lines.

24. Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR: **Pack-MULE
•• transposable elements mediate gene evolution in plants**. *Nature* 2004, **431**:569-573 http://dx.doi.org/10.1038/nature02953.
This is a first systematic analysis of Pac-MULEs based on rice genome that documents its ability to capture gene fragments from different chromosomes creating chimeric gene fusions inside the element.

25. Jiang N, Ferguson AA, Slotkin RK, Lisch D: **Pack-Mutator-like
• transposable elements (Pack-MULEs) induce directional modification of genes through biased insertion and DNA acquisition**. *Proc Natl Acad Sci U S A* 2011, **108**:1537-1542 http://dx.doi.org/10.1073/pnas.1010814108.
This paper documents preference of Pac-MULEs to capture GC-rich genes and gene fragments and preferentially insert into five-ends of genes, with possibility of creating new gene fusions. The Pac-MULEs can be in part responsible for the negative GC gradient observed in grasses.

26. Quadrana L, Silveira AB, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, Colot V: **The Arabidopsis thaliana mobilome and its impact at the species level**. *eLife* 2016, **5** http://dx.doi.org/10.7554/eLife.15716.

27. Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis Goicoechea J, Liang C et al.: **Whole-genome sequencing of Oryza brachyantha reveals mechanisms underlying Oryza genome evolution**. *Nat Commun* 2013, **4** http://dx.doi.org/10.1038/ncomms2596.

28. Woodhouse MR, Pedersen B, Freeling M: **Transposed genes in
•• Arabidopsis are often associated with flanking repeats**. *PLoS Genet* 2010, **6**:e1000949.
This study provides a genome wide scan for gene insertions that happened since the divergence of *Arabidopsis thaliana* and *Arabidopsis lyrata*, with papaya as an outgroup. The results indicate the most of gene movement occurs independent of transposons. Certain gene categories, including NLRs are more prone to move. The proposed model involves circular DNA intermediate that loops out from the donor site and inserts into the destination site through non-allelic homologous recombination.

29. Parks MM, Lawrence CE, Raphael BJ: **Detecting non-allelic homologous recombination from high-throughput sequencing data**. *Genome Biol* 2015, **16** http://dx.doi.org/10.1186/s13059-015-0633-1.

30. Chang HHY, Pannunzio NR, Adachi N, Lieber MR: **Non-homologous DNA end joining and alternative pathways to double-strand break repair**. *Nat Rev Mol Cell Biol* 2017, **18**:495-506 http://dx.doi.org/10.1038/nrm.2017.48.

31. Lanciano S, Carpentier MC, Llauro C, Jobet E, Robakowska-
•• Hyzorek D, Lasserre E, Ghesquière A, Panaud O, Mirouze M: **Sequencing the extrachromosomal circular mobilome reveals retrotransposon activity in plants**. *PLoS Genet* 2017, **13**:e1006630 http://dx.doi.org/10.1371/journal.pgen.1006630.
This study sequences extra chromosomal circular DNA from *Arabidopsis* and rice to characterize the mobile elements. The methods of isolating eccDNA from plants are well described. Results suggest inducible activity of selected retrotransposable elements upon stress and destabilization of epigenetic regulation.

32. Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ et al.: **Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae**. *Proc Natl Acad Sci U S A* 2009, **106**:15780-15785 http://dx.doi.org/10.1073/pnas.0908195106.

33. Clark KA, Krysan PJ: **Chromosomal translocations are a common phenomenon in Arabidopsis thaliana T-DNA insertion lines**. *Plant J* 2010, **64**:990-1001 http://dx.doi.org/10.1111/j.1365-313X.2010.04386.x.

34. Kim S, Park J, Yeom S-I, Kim Y-M, Seo E, Kim K-T, Kim M-S, Lee JM, Cheong K, Shin H-S et al.: **New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication**. *Genome Biol* 2017, **18**:210.

35. Sakai H, Mizuno H, Kawahara Y, Wakimoto H, Ikawa H, Kawahigashi H, Kanamori H, Matsumoto T, Itoh T, Gaut BS: **Retrogenes in rice (*Oryza sativa* L. ssp. japonica) exhibit correlated expression with their source genes**. *Genome Biol Evol* 2011, **3**:1357-1368 http://dx.doi.org/10.1093/gbe/evr111.

36. Cho J, Benoit M, Catoni M, Drost H-G, Brestovitsky A, Oosterbeek M, Paszkowski J: **Sensitive detection of pre-integration intermediates of LTR retrotransposons in crop plants**. *bioRxiv* 2018 http://dx.doi.org/10.1101/317479.

37. Green BM, Finn KJ, Li JJ: **Loss of DNA replication control is a potent inducer of gene amplification**. *Science (80-)* 2010, **329**:943-946 http://dx.doi.org/10.1126/science.1190966.

38. Finn KJ, Li JJ: **Single-stranded annealing induced by re-initiation of replication origins provides a novel and efficient mechanism for generating copy number expansion via non-allelic homologous recombination**. *PLoS Genet* 2013, **9** http://dx.doi.org/10.1371/journal.pgen.1003192.

39. Smit A, Hubley R, Green P: *RepeatMasker Open-4.0.6 2013–2015*. 2017 http://www.repeatmasker.org.

40. Kapitonov VV, Jurka J: **Rolling-circle transposons in eukaryotes**. *Proc Natl Acad Sci U S A* 2001, **98**:8714-8719 http://dx.doi.org/10.1073/pnas.151269298.

41. Brunner S, Pea G, Rafalski A: **Origins, genetic organization and transcription of a family of non-autonomous Helitron elements in maize**. *Plant J* 2005, **43**:799-810 http://dx.doi.org/10.1111/j.1365-313X.2005.02497.x.

42. Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK: **Gene capture by Helitron transposons reshuffles the transcriptome of maize**. *Genetics* 2012, **190**:965-975 http://dx.doi.org/10.1534/genetics.111.136176.

43. Lynch BT, Patrick TL, Moreno JJ, Siebert AE, Klusman KM, Shodja DN, Hannah LC, Lal SK: **Differential pre-mRNA splicing alters the transcript diversity of Helitrons between the maize inbred lines**. *G3 Genes Genomes Genet* 2015, **5**:1703-1711 http://dx.doi.org/10.1534/g3.115.018630.

44. Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A:
•• **Gene duplication and exon shuffling by Helitron-like transposons generate intraspecies diversity in maize**. *Nat Genet* 2005, **37**:997-1002 http://dx.doi.org/10.1038/ng1615.
A whole genome comparison across two maize lines identified that Helitrons are responsible for eight out of nine observed gene sized insertions. Some of the transcripts contained fragments of different genes solidifying the role of Helitrons in novel gene fusions.

45. Xiong W, He L, Lai J, Dooner HK, Du C: **HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes**. *Proc Natl Acad Sci U S A* 2014, **111**:10263-10268 http://dx.doi.org/10.1073/pnas.1410006111.

46. Zhang J, Zuo T, Peterson T: **Generation of tandem direct**
• **duplications by reversed-ends transposition of maize ac elements**. *PLoS Genet* 2013, **9**:e1003691.
This is a well-documented analysis of reverse-end transposition and the role of DNA transposons in generation of tandem repeats.

47. Assaad FF, Signer ER: **Somatic and germinal recombination of a direct repeat in Arabidopsis**. *Genetics* 1992, **132**:553-566.

48. Zmienko A, Samelak-Czajka A, Kozlowski P, Szymanska M, Figlerowicz M: **Arabidopsis thaliana population analysis reveals high plasticity of the genomic region spanning MSH2, AT3G18530 and AT3G18535 genes and provides evidence for NAHR-driven recurrent CNV events occurring in this location**. *BMC Genomics* 2016, **17** http://dx.doi.org/10.1186/s12864-016-3221-1.

49. Jelesko JG, Carter K, Thompson W, Kinoshita Y, Gruissem W: **Meiotic recombination between paralogous RBCSB genes on sister chromatids of *Arabidopsis thaliana***. *Genetics* 2004, **166**:947-957.

50. Jelesko JG, Harper R, Furuya M, Gruissem W: **Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana***. *Proc Natl Acad Sci U S A* 1999, **96**:10302-10307.

51. Jelesko JG, Carter K, Kinoshita Y, Gruissem W: **Frequency and character of alternative somatic recombination fates of paralogous genes during T-DNA integration**. *Mol Genet Genomics* 2005, **274**:91-102.

52. Leister D: **Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene**. *Trends Genet* 2004, **20**:116-122.

53. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: **Topological domains in mammalian genomes identified by analysis of chromatin interactions**. *Nature* 2012, **485**:376-380 http://dx.doi.org/10.1038/nature11082.

54. Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J: **Hi-C: a comprehensive technique to capture the conformation of genomes**. *Methods* 2012, **58**:268-276 http://dx.doi.org/10.1016/j.ymeth.2012.05.001.

55. Samelak-Czajka A, Marszalek-Zenczak M, Marcinkowska-Swojak M, Kozlowski P, Figlerowicz M, Zmienko A: **MLPA-based analysis of copy number variation in plant populations**. *Front Plant Sci* 2017, **8** http://dx.doi.org/10.3389/fpls.2017.00222.

56. Tan EH, Henry IM, Ravi M, Bradnam KR, Mandakova T, Marimuthu MPA, Korf I, Lysak MA, Comai L, Chan SWL: **Catastrophic chromosomal restructuring during genome elimination in plants**. *eLife* 2015, **4** http://dx.doi.org/10.7554/eLife.06516.

57. Vaughn JN, Bennetzen JL: **Natural insertions in rice commonly form tandem duplications indicative of patch-mediated double-strand break induction and repair**. *Proc Natl Acad Sci U S A* 2014, **111**:6684-6689 http://dx.doi.org/10.1073/pnas.1321854111.

58. Cohen S, Houben A, Segal D: **Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants**. *Plant J* 2008, **53**:1027-1034 http://dx.doi.org/10.1111/j.1365-313X.2007.03394.x.

59. Koo D-H, Molin WT, Saski CA, Jiang J, Putta K, Jugulam M,
• Friebe B, Gill BS: **Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri***. *Proc Natl Acad Sci U S A* 2018, **115**:3332-3337 http://dx.doi.org/10.1073/pnas.1719354115.
This paper reports a critical example of adaptive gene evolution on eccDNA that allowed wild grass species to acquire herbicide resistance.

60. Carpentier M-C, Manfroi E, Wei F-J, Wu H-P, Lasserre E, Llauro C, Debladis E, Akakpo R, Hsing Y-I, Panaud O: **Retrotranspositional landscape of Asian rice revealed by 3000 genomes**. *Nat Commun* 2019, **10**:24.

61. Estep MC, Debarry JD, Bennetzen JL: **The dynamics of LTR**
•• **retrotransposon accumulation across 25 million years of panicoid grass evolution**. *Heredity (Edinb)* 2013, **110**:194-204 http://dx.doi.org/10.1038/hdy.2012.99.
This paper presents a detailed analysis of LTR family dynamics across *Pennisetum*, *Saccharum*, *Sorghum*, and *Zea* lineages. Their analyses indicate that independent activation of distinct LTR elements occurred in different lineages and these elements can become responsible for genome amplification. The results also suggest random activation of LTRs, with some families being more prone to re-activation and amplification.

62. Costello R, Emms D, Kelly S: **Gene duplication accelerates the pace of protein gain and loss from plant organelles**. *bioRxiv* 2018 http://dx.doi.org/10.1101/415125.

63. Salomon S, Puchta H: **Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells**. *EMBO J* 1998, **17**:6086-6095.

64. Ma J, Bennetzen JL: **Rapid recent growth and divergence of rice nuclear genomes**. *Proc Natl Acad Sci USA* 2004, **101**:12404-12410.