

## Review Article

## Current status of structural variation studies in plants

Yuxuan Yuan<sup>1,2</sup> , Philipp E. Bayer<sup>1</sup> , Jacqueline Batley<sup>1</sup>  and David Edwards<sup>1,\*</sup> <sup>1</sup>School of Biological Sciences and Institute of Agriculture, The University of Western Australia, Perth, WA, Australia<sup>2</sup>School of Life Sciences and State Key Laboratory for Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong SAR, China

Received 18 February 2021;

revised 31 May 2021;

accepted 3 June 2021.

\*Correspondence (Tel +61 8 6488 2415; fax

+61 8 6488 1380; email

dave.edwards@uwa.edu.au)

**Keywords:** structural variation, DNA sequencing, optical mapping, gene expression, phenotypic variation, breeding.

## Summary

Structural variations (SVs) including gene presence/absence variations and copy number variations are a common feature of genomes in plants and, together with single nucleotide polymorphisms and epigenetic differences, are responsible for the heritable phenotypic diversity observed within and between species. Understanding the contribution of SVs to plant phenotypic variation is important for plant breeders to assist in producing improved varieties. The low resolution of early genetic technologies and inefficient methods have previously limited our understanding of SVs in plants. However, with the rapid expansion in genomic technologies, it is possible to assess SVs with an ever-greater resolution and accuracy. Here, we review the current status of SV studies in plants, examine the roles that SVs play in phenotypic traits, compare current technologies and assess future challenges for SV studies.

## Introduction

Structural variations (SVs) are genetic differences between individuals, which can lead to gene loss, gene duplication and the generation of novel genes, therefore, leading to phenotypic variations in a species. An SV is defined as a region of DNA that has a change in sequence length, copy number, orientation or chromosomal location between individuals (Escaramis *et al.*, 2015). Generally, an SV can be classified as a deletion, an insertion, a copy number variation (CNV), an inversion or a translocation. In contrast to single nucleotide polymorphisms (SNPs) and small indels (insertions and deletions), SVs are considered to be longer (>50 bp) and can have a greater influence on gene expression and protein function than SNPs (Chiang *et al.*, 2017).

In early plant genomic studies, the limitation of technologies and the lack of high-quality reference genome assemblies prevented the comprehensive exploration of SVs in plants. Many plants have large and complex genomes, with polyploidy occurring in up to 80% of plant species (Meyers and Levin, 2006), making the identification of SVs in plant genomes a challenge. Recent advances in genomic technologies, particularly long-read sequencing and whole-genome mapping, promise the production of high-quality plant genome and pangenome assemblies and access to a broad range of SVs to assess their potential role in plant phenotypic variation.

Although improvements in DNA sequencing, whole-genome mapping and novel algorithms have made it feasible to characterize SVs on a genome-wide scale with higher accuracy, the reports of SV studies in plants are still limited. Significant effort and resources are needed to comprehensively decipher the association between SVs and agronomic traits to support plant improvement, supporting the economies and food security. Here, we discuss the current progress and challenges of SV studies in

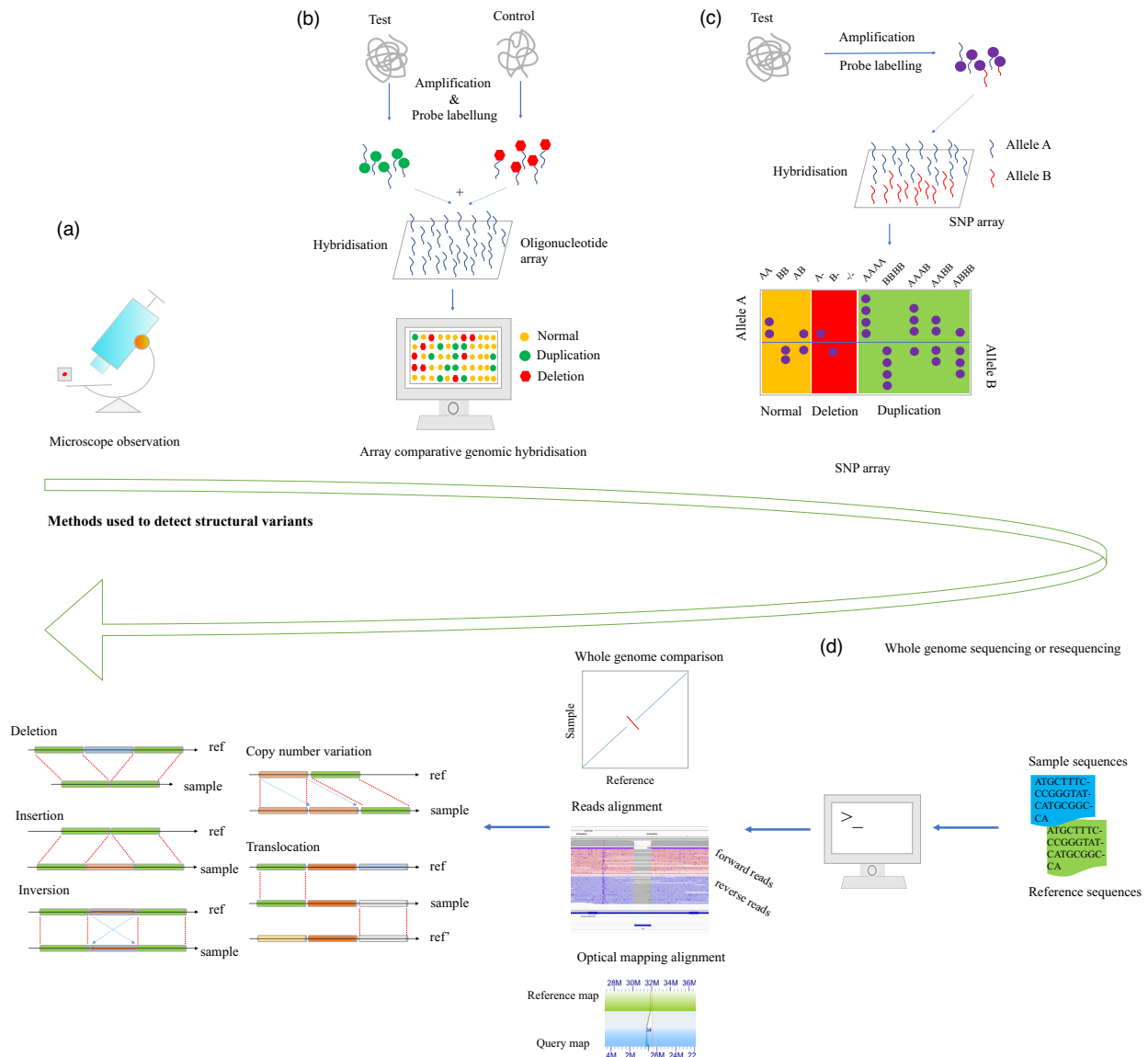
plants and the potential to apply knowledge of SVs to improve crop varieties.

## Limitations of early technologies for SV identification

Before the widespread use of molecular markers and DNA sequencing, SVs were characterized by microscopes at a karyotype level, with a resolution >3 Mb (Figure 1a) (Feuk *et al.*, 2006). Due to the low throughput and the limited resolution of microscopic observation however, there are few novel SV studies using microscopic techniques, and they are now mostly applied to confirm known SVs. Approximately 15 years ago, the advent of hybridization-based microarray approaches made it possible to perform SV studies with a greater resolution and lower cost than microscopic methods. Two commonly used methods were array comparative genomic hybridization (array-CGH) and SNP arrays.

Array-CGH can efficiently detect CNVs at multiple genomic loci (Figure 1b) and has been applied to diverse studies, including gene discovery, epigenetic modification and chromatin conformation (Bejjani and Shaffer, 2006). Nevertheless, array-CGH cannot detect balanced SVs (e.g. reciprocal translocations and inversions) or absolute copy numbers of a DNA segment, because it detects genetic imbalances between two individual genomes, where a sample has more or less of a specific genetic material than others (Escaramis *et al.*, 2015). Additionally, array-CGH is specifically designed for diploid individuals and it is not sensitive to higher degrees of ploidy (>2 sets of chromosomes). A well-assembled reference is also essential during the design of the array (Park *et al.*, 2010).

In contrast to array-CGH, SNP arrays are more sensitive to allele-specific CNVs and can help to identify large-scale CNVs in diverse populations (Figure 1c; Alkan *et al.*, 2011). However, SNP arrays provide a poor signal-to-noise ratio compared with array-



**Figure 1** Methods used to identify structural variations from the past to the present. The figure lists commonly used methods to identify SVs from the early (a) microscope observation, (b) array comparative genomic hybridization and (c) SNP array to the current (d) DNA sequencing.

CGH due to the smaller target size (Hester *et al.*, 2009). As with array-CGH, SNP arrays cannot be used to detect insertions. The number of SVs detected is dependent on the density or presence and absence of SNPs in the target genome regions. Moreover, SNP arrays were initially designed for diploid samples and struggle to characterize repeat-rich and duplicated regions. The design of SNP arrays also depends on the quality of the reference genome assemblies. Furthermore, the breakpoints of SVs cannot be easily detected by SNP arrays or array-CGH.

### Current technologies for SV identification

With advances in DNA sequencing, whole-genome analysis has become viable for a wide range of plant species. Examining whole genomes by DNA sequencing has permitted SV characterization at the nucleotide level, and the detection of inversions and translocations, as well as recombination breakpoints, has become more efficient.

Initially, single-end reads (sequenced only in one direction from a DNA strand) were used. With the expansion of sequencing methods, paired reads, sequenced from both forward and reverse orientations from complementary DNA strands, with a known approximate distance between the pairs, have been used to overcome the challenges of associating short single reads with regions of the genome. However, the short length (<600 bp) of these reads still poses challenges for the characterization of repetitive regions (Michael and VanBuren, 2020), and thus, the accuracy of SV detection based on short sequence reads is limited.

Recent advances in long-read sequencing and high-throughput chromosome conformation capture (Hi-C) technologies offer solutions to overcome some of the problems associated with short sequence reads. Hi-C read pairs can physically span the entire chromosomes and can be applied to detect large-scale SVs (Ho *et al.*, 2020), while long-read sequencing, comprising synthetic long-read sequencing and single-molecule long-read

sequencing (Goodwin *et al.*, 2016) can average 10 to >100 kb in length to resolve SVs that cannot readily be assayed using short reads by short-read sequencing. The previous high error rates (5–15%), low throughput and relatively high cost of single-molecule long-read sequencing have limited their application (Yuan *et al.*, 2017). However, with reducing costs and the continued advances in sequencing technology and computational algorithms, more accurate data (accuracy >99%), such as PacBio HiFi reads and Oxford Nanopore R10.3 reads, have been produced, which could further improve the accuracy of genome analysis, particularly for haplotype-aware genome assembly and SV studies (Wenger *et al.*, 2019).

Optical mapping in nanochannels is complementary to DNA sequencing and provides an approach for large-scale SV detection (Yuan *et al.*, 2020). DNA from the plant species is nicked or directly labelled by specific enzymes such as *Nt. BspQI*, *Nb. BssSI* and *DLE-1*, and strands are loaded and stretched in nanochannels, labelled by fluorescence and scanned in an optical mapping device (Lam *et al.*, 2012). The fluorescence images produced are then converted into single-molecule maps based on nicked enzyme site positions. The average length of single-molecule maps is around 225 kb (Shelton *et al.*, 2015), and thus, optical mapping can capture larger genomic structural variation that is not easily detected by DNA sequencing.

## Strategies for SV identification

There are two commonly used strategies to detect SVs using DNA sequencing (Figure 1d). One is to directly compare *de novo* genome assemblies, and the other is to use the information from mapping reads to a reference, such as paired reads (PR), read depth (RD) and split reads (SR) to detect SVs (Escaramis *et al.*, 2015).

Since the release of the *Arabidopsis thaliana* genome assembly in 2000 (Arabidopsis Genome, 2000), approximately 450 plant genomes have been assembled (<https://www.plabipd.de>). The continued increase in high-quality genome assemblies makes SV characterization in plants more reliable. Whole-genome comparison can identify SVs by comparing the genome of one individual to another. Several tools have been developed for this purpose, including Mauve (Darling *et al.*, 2004), MUMMER (Kurtz *et al.*, 2004), LASTZ (Harris, 2007), Assemblytics (Nattestad and Schatz, 2016), paftools (Li, 2018), SyRI (Goel *et al.*, 2019) and SVIM-asm (Heller and Vingron, 2020). However, due to the difficulty and expense of producing high-quality genome assemblies, and the challenge of differentiating between real genomic differences and assembly or annotation artefacts (Bayer *et al.*, 2018; Bayer *et al.*, 2017), the application of whole-genome comparison in SV detection is limited (Wala *et al.*, 2018), while SV analysis using read mapping is more common.

In principle, paired reads can be used to detect all kinds of SVs, as SVs change the paired read mapping patterns (Ye *et al.*, 2016). Briefly, when aligning to an insertion, the distance between paired reads will be increased compared to the average insert size, while for a deletion, the distance between paired reads will reduce compared to the average. If an inversion occurs, the orientation of reads can be reversed. Translocations can also be detected using the information from mapped paired reads, as the reads may map to different chromosomal locations. For CNVs, read mapping can lead to increased or decreased mapped read depth depending on the copy number of the target genome regions. However, due to the short-read length, repetitiveness

and complexity of plant genomes, up to 89% of SVs have been reported to be false positives, which needs comprehensive filtration to ensure robust results (Sedlazeck *et al.*, 2018). Although short sequence reads can be less efficient for SV detection than longer reads, they are still applied to characterize SVs due to their relatively low cost. To facilitate such analysis, many tools have been developed for using short reads to detect SVs (Table 1).

With continued advances in DNA sequencing and algorithms, long DNA sequence reads have increasingly been adopted for SV detection. Compared to short-read-based mapping approaches, long sequence reads can more accurately identify SVs, particularly in complex regions that cannot be spanned by short sequence reads (Sedlazeck *et al.*, 2018; Spielmann *et al.*, 2018). Long sequence reads are particularly useful for insertion detection, which can be challenging using short sequence reads. For example, in a human SV study, Huddleston *et al.* (2017) used PacBio long sequence reads and detected 1967 novel SVs that had been missed in previous studies. Using 10× Genomics reads, Wong *et al.* (2018) found that short sequence reads were inefficient in large-scale insertion detection, with 1842 unique insertions having been missed. In plants, a chromosome-level assembly of *A. thaliana* Nd-1 using PacBio long sequence reads revealed 385 genes initially identified in *A. thaliana* Col-0, having at least two copies in Nd-1 (Pucker *et al.*, 2019).

Although long-read sequencing has provided improved resolution in detecting SVs that may not readily be identified by short-read sequencing, both technologies are inefficient in large-scale SV detection. In these cases, optical mapping and Hi-C technologies afford useful solutions. By mapping the physical locations of nicking sites in reference and query genomes, optical mapping can detect large-scale variations in genome structure (Cao *et al.*, 2014). However, although optical maps are long, the accuracy of SVs detected by optical mapping is highly dependent on the quality of the reference genome and the density of nicking sites (Yuan *et al.*, 2018). In contrast, Hi-C detects large-scale SVs based on 3D chromatin structure, while the coverage of Hi-C reads can support the accuracy of SV detection (Bickhart *et al.*, 2017).

With the increasing understanding of SVs between individuals of the same species, there has been a growth in the production of pangenome references which aim to capture presence and absence variations (Golicz *et al.*, 2016a). A pangenome describes the whole gene set in a species, involving genes present in all individuals (core genes) and genes present only in some individuals (variable or dispensable genes; Bayer *et al.*, 2020; Danilevicz *et al.*, 2020; Golicz *et al.*, 2020). First applied to the studies of microorganisms (Read *et al.*, 2013; Tettelin *et al.*, 2005), pangenome studies have been extended to more complex organisms including plants, and the definition has also been expanded to include all genomic elements, not just expressed genes. Several plant pangenomes have been analysed, including wheat (Montenegro *et al.*, 2017; Walkowiak *et al.*, 2020), barley (Jayakodi *et al.*, 2020), maize (Hirsch *et al.*, 2014; Hufford *et al.*, 2021; Lu *et al.*, 2015; Unterseer *et al.*, 2017), rice (Schatz *et al.*, 2014; Sun *et al.*, 2017; Wang *et al.*, 2018), soybean (Li *et al.*, 2014; Liu *et al.*, 2020; Valliyodan *et al.*, 2021), *Brassica rapa* (Lin *et al.*, 2014), *Brassica oleracea* (Golicz *et al.*, 2016b), *Brassica napus* (Hurgobin *et al.*, 2018; Song *et al.*, 2020), chickpea (Varshney *et al.*, 2019), grapevine (Magris *et al.*, 2015), *Medicago truncatula* (Zhou *et al.*, 2014), *Arabidopsis thaliana* (Cao *et al.*, 2011; Jiao and Schneeberger, 2020), pigeonpea (Zhao *et al.*, 2020), *Brachypodium distachyon* (Vogel *et al.*, 2016), cultivated

**Table 1** Software used to detect structural variations

Software	Language	SV calling type					Data type	References
		Insertion	Deletion	Inversion	CNV	Translocation		
ETCHING	C and C++		✓	✓		✓	PE	Choi <i>et al.</i> (2020)
Scpluscnv	R	✓	✓	✓	✓	✓	PE	Lopez <i>et al.</i> (2020)
CONY	R				✓		PE	Wei and Huang (2020)
cuteSV	Python	✓	✓	✓		✓	PB; ONT	Jiang <i>et al.</i> (2020)
NanoVar	C++; Python; C; shell	✓	✓	✓			ONT	Tham <i>et al.</i> (2019)
SVIM	Python	✓	✓	✓			PB; ONT	Heller and Vingron (2019)
PBSV	Python	✓	✓	✓	✓	✓	PB	PacificBiosciences (2018)
Sniffles	C++; C; HTML	✓	✓	✓		✓	PB; ONT	Sedlazeck <i>et al.</i> (2018)
Picky	Perl	✓	✓	✓		✓	PB; ONT	Gong <i>et al.</i> (2018)
NanoSV	Python; shell	✓	✓	✓		✓	PB; ONT	Cretu Stancu <i>et al.</i> (2017)
SVachra	Ruby	✓	✓	✓		✓	PE; MP	Hampton <i>et al.</i> (2017)
PSSV	R	✓	✓	✓		✓	PE	Chen <i>et al.</i> (2017)
Seeksv	C++	✓	✓	✓			PE	Liang <i>et al.</i> (2017)
novoBreak	Perl; shell		✓	✓		✓	PE	Chong <i>et al.</i> (2017)
Manta	C++; Python	✓	✓				PE	Chen <i>et al.</i> (2016)
SoftSV	C++		✓	✓		✓	PE	Bartenhagen and Dugas (2016)
SV-STAT	Shell; Perl	✓	✓	✓		✓	PE; SE	Davis <i>et al.</i> (2016)
MUMdex	C++	✓	✓	✓		✓	PE	Andrews <i>et al.</i> (2016)
MetaSV	Python; HTML; Shell	✓	✓	✓		✓	PE	Mohiyuddin <i>et al.</i> (2015)
BreakMer	Python	✓	✓	✓		✓	SE	Abo <i>et al.</i> (2015)
Genome STRiP2	Java; R		✓	✓			PE	Handsaker <i>et al.</i> (2015)
Hydra-multi	C++; Python; Shell; Perl		✓	✓		✓	PE	Lindberg <i>et al.</i> (2015)
Ulysses	Python; R	✓	✓	✓		✓	MP	Gillet-Markowska <i>et al.</i> (2015)
LUMPY	C; C++; Python; Shell		✓	✓	✓	✓	PE	Layer <i>et al.</i> (2014)
Scalpel	Perl; C++	✓	✓				PE	Narzisi <i>et al.</i> (2014)
Gustaf	C++		✓	✓		✓	PE; SE	Trappe <i>et al.</i> (2014)
PBHoney	Python	✓	✓	✓		✓	PB	English <i>et al.</i> (2014)
Socrates	Java		✓	✓		✓	PE; SE	Schroder <i>et al.</i> (2014)
FACTERA	Perl		✓	✓		✓	PE	Newman <i>et al.</i> (2014)
SMuFin	C	✓	✓	✓		✓	PE	Moncunill <i>et al.</i> (2014)
CNVeM	C				✓		PE	Wang <i>et al.</i> (2013)
Breakpointer	Fortran; Python					✓	PE	Drier <i>et al.</i> (2013)
Bellerophon	Perl					✓	PE	Hayes and Li (2013)
PeSV-Fisher	Python		✓	✓		✓	PE; MP	Escaramis <i>et al.</i> (2013)
RetroSeq	Perl	✓					PE	Keane <i>et al.</i> (2013)
SOAPindel	Perl; C++	✓	✓				PE	Li <i>et al.</i> (2013)
cn.MOPS	R				✓		PE; SE	Klambauer <i>et al.</i> (2012)
Magnolya	Python				✓		PE	Nijkamp <i>et al.</i> (2012)
Cortex	C	✓	✓				PE; SE	Iqbal <i>et al.</i> (2012)
CNVnorma	R				✓		PE	Gusnanto <i>et al.</i> (2012)
Control-FREEC	C++				✓		PE; SE	Boeva <i>et al.</i> (2012)
cnvHiTSeq	Java				✓		PE	Bellos <i>et al.</i> (2012)
CLEVER	C++	✓	✓				PE	Marschall <i>et al.</i> (2012)
Delly	C++		✓	✓		✓	PE	Rausch <i>et al.</i> (2012)
GASVPro	Java; C++; perl; python	✓	✓	✓		✓	PE	Sindi <i>et al.</i> (2012)
PRISM	N/A		✓	✓			PE	Jiang <i>et al.</i> (2012)
SVMIner	C++; Perl	✓	✓				PE	Hayes <i>et al.</i> (2012)
BIC-seq	Perl; R				✓		PE; SE	Xi <i>et al.</i> (2011)
ReadDepth	R				✓		PE	Miller <i>et al.</i> (2011)
CNVnator	C++; Perl				✓		PE; SE	Abyzov <i>et al.</i> (2011)
JointSLM	R				✓		PE; SE	Magi <i>et al.</i> (2011)
Clipcrop	JavaScript	✓	✓	✓		✓	PE	Suzuki <i>et al.</i> (2011)
CREST	Perl		✓	✓		✓	PE; SE	Wang <i>et al.</i> (2011)
inGAP-sv	Java	✓	✓	✓		✓	PE	Qi and Zhao (2011)
Splitread	C; Shell	✓	✓				PE	Karakoc <i>et al.</i> (2011)
rSW-seq	C				✓		SE	Kim <i>et al.</i> (2010)

Table 1 Continued

Software	Language	SV calling type					Data type	References
		Insertion	Deletion	Inversion	CNV	Translocation		
cnD	D				✓		PE	Simpson <i>et al.</i> (2010)
CNVr	Shell; C				✓		PE	Medvedev <i>et al.</i> (2010)
SVMerge	Perl; Shell	✓	✓	✓		✓	PE; SE	Wong <i>et al.</i> (2010)
SVDetect	Perl	✓	✓	✓		✓	PE; MP	Zeitouni <i>et al.</i> (2010)
VariationHunter	N/A	✓					PE	Hormozdiari <i>et al.</i> (2010)
NovelSeq	C++	✓					PE	Hajirasouliha <i>et al.</i> (2010)
SLOPE	C++	✓	✓			✓	PE; SE	Abel <i>et al.</i> (2010)
BreakSeq	Python; Perl	✓	✓				PE	Lam <i>et al.</i> (2010)
mrCaNaVaR	C		✓		✓		PE	Alkan <i>et al.</i> (2009)
CNV-seq	Perl; R				✓		PE	Xie and Tammi (2009)
RDXplorer	Shell; C				✓		SE	Yoon <i>et al.</i> (2009)
BreakDancer	Perl; C++	✓	✓	✓		✓	PE	Chen <i>et al.</i> (2009)
MoDIL	N/A	✓	✓				PE	Lee <i>et al.</i> (2009)
PEMer	Python; Perl; C++	✓	✓	✓		✓	PE	Korbel <i>et al.</i> (2009)
Pindel	C++; Perl; Python; Shell	✓	✓	✓			PE	Ye <i>et al.</i> (2009)

Data type: PE – paired end; SE – single end; MP – mate pair. PB: PacBio; ONT: Oxford nanopore; N/A: not available.

pepper (Ou *et al.*, 2018), sesame (Yu *et al.*, 2019), sunflower (Hubner *et al.*, 2019), tomato (Alonge *et al.*, 2020; Gao *et al.*, 2019), apple (Sun *et al.*, 2020) and poplar (Pinosio *et al.*, 2016). The methods to study PAVs in a pangenome are similar to these described here for SV detection, and with the further improvement of genomic technologies and algorithms, the study of pangenomes in plants will be more common.

## Current status of SV studies in plants

Structural variation studies in plants are increasing and are being applied to understand genomic changes during evolution, domestication and breeding. Recently, several pangenome studies have been conducted for different plant species, and PAV diversity has been investigated. For example, in a wheat pangenome study, Montenegro *et al.* (2017) used 18 wheat cultivars to identify PAVs associated with important agronomic traits, including response to environmental stress and defence response. In this study, they also demonstrated that the reference genome cultivar, Chinese Spring, poorly represented modern wheat lines. Recently, Walkowiak *et al.* (2020) studied 15 representative wheat cultivars collected from around the world and found that a translocation that occurred in some of the cultivars between chromosomes 5B and 7B is selectively neutral during breeding. In a subsequent study using 538 wheat lines, they found that the translocation occurred in 66% of the selected lines (Walkowiak *et al.*, 2020). In a recent barley pangenome study, Jayakodi *et al.* (2020) found that a large-scale inversion (~10 Mb) on chromosome 2H is frequently found in germplasm from northern Europe. Golicz *et al.* (2016b) reported that SVs affected the presence of flowering time genes such as *FLOWERING LOCUS C (FLC)* in *Brassica oleracea*. Through the pangenome study of *Brassica oleracea* and *Brassica napus*, Bayer *et al.* (2019) and Dolatabadian *et al.* (2020) revealed that disease resistance genes show diverse PAV patterns among different *Brassica*

accessions and that this seems to be a common feature of plant pangenomes (Dolatabadian *et al.*, 2017). By examining 725 tomato accessions, Gao *et al.* (2019) discovered 4873 genes demonstrating PAV and identified a rare allele deletion in the *TomLoxC* promoter that affects the flavour of tomato. In a 3000 rice genome project, Fuentes *et al.* (2019) demonstrated that rice genomic regions with frequent SVs were enriched in stress response genes.

Zhang *et al.* (2015) reported that SVs affected the coding regions of 1676 cucumber genes, and they found that genes in deleted regions were associated with histone methylation and abiotic stress response, while duplicated genes were often involved in the reproductive process. Genes encoded in inversion regions played an important role in the response to chemical stimulus, and genes in insertion regions were related to histone acetylation.

With recent development and application of long-read sequencing and optical mapping, SV studies in plants have been further refined, and numerous high-quality SV studies have been reported (Table 2). For example, Michael *et al.* (2018) assembled one *Arabidopsis* genome and found that, compared to the Col-0 genome assembly, the new Oxford Nanopore genome assembly has 4280 SVs with a total length of 9.5 Mb, among which, repeat-related SVs account for 58%, followed by insertions and deletions (31%). Zhou *et al.* (2019) studied 50 *grapevine* cultivars and 19 wild relatives and found that inversions and translocations have strong associations with selection. In a soybean study, Xie *et al.* (2019) compared wild and cultivated soybeans using optical mapping and confirmed a large inversion at the *I* locus that can affect seed coat colour during domestication. Using PacBio long-read sequencing, Song *et al.* (2020) *de novo* assembled eight canola genomes and found 77.2 – 149.6 Mb PAVs among these accessions. After a PAV-based genome-wide association study (GWAS), they identified three *FLC* genes that are related to ecotype differentiation. Recently, two SV studies by Liu *et al.*

**Table 2** Recent structural variations studies in plants

Species	Methods	Major SV findings	References
Melon	Short-read alignment	A 1,070-bp deletion at 23.85 kb upstream of MELO3C019694 was found which might impair the transcriptional regulation of this gene	Zhao <i>et al.</i> (2019)
<i>Setaria viridis</i>	Whole-genome comparison	Approximately 22% of the genes were variable genes	Mamidi <i>et al.</i> (2020)
<i>Brassica nigra</i>	Long-read alignment	Approximately 6000–7000 SVs found in the two <i>B. nigra</i> accessions and among the SVs 63.4–70% were deletions	Perumal <i>et al.</i> (2020)
Eggplant	Short-read alignment	Asymmetric SV accumulation was found in potential regulatory regions of protein-coding genes among the different eggplant genomes	Wei <i>et al.</i> (2020)
Peach	Short-read alignment	A 9-bp insertion in Prupe.4G186800 had an association with early fruit maturity; a 487-bp deletion in the promoter of PpMYB10.1 was associated with flesh colour around the stone; a 1.67 Mb inversion was highly associated with fruit shape; a gene adjacent to the inversion breakpoint of PpOFP1 regulated flat shape formation	Guo <i>et al.</i> (2020)
Banana	Fluorescence <i>in situ</i> hybridization (FISH)	Large differences in chromosome structure discriminated individual banana accessions	Simonikova <i>et al.</i> (2020)
Maize	Whole-genome comparison, short-read alignment, and FISH	A 1.8 Mb duplication on the Gametophyte factor1 locus which was for unilateral cross-incompatibility; increased copy number of carotenoid cleavage dioxygenase 1 ( <i>ccd1</i> ) in A188 was associated with elevated expression during seed development	Lin <i>et al.</i> (2020)
Peanut	Short-read alignment, whole-genome comparison	<i>A. hypogaea</i> showed more enrichment of deletions and insertions in the upstream regions of the coding sequences than <i>A. monticola</i>	Yin <i>et al.</i> (2020)
<i>Brassica napus</i>	whole-genome comparison	77.2–149.6 Mb sequences showed PAV patterns, which included more than 9.5% of the genes	Song <i>et al.</i> (2020)
Rice	Short-read alignment, whole-genome comparison and long-read alignment	The site-frequency spectrum of SVs was skewed towards lower frequency variants than synonymous SNPs; peaks of SV divergence were enriched for known domestication genes	Kou <i>et al.</i> (2020)
Maize	Whole-genome comparison	21.9% of the polymorphic SVs showed low linkage disequilibrium with nearby SNPs; A new significant locus for oil concentration and long-chain fatty acid composition (C18_1, C18_2 and C20_1) on chromosome 4 was found associating with SVs	Yang <i>et al.</i> (2019)
<i>Solanum pimpinellifolium</i>	Whole-genome comparison and long-read alignment	SVs overlapping genes played a role in breeding traits such as fruit weight and lycopene content; SVs contribute to complex regulatory networks, such as fruit quality traits	Wang <i>et al.</i> (2020)
Rice	Whole-genome comparison and long-read alignment	organelle-to-nucleus DNA transfers resulted in numerous SVs that participated in the nuclear genome divergence of rice species and subspecies	Ma <i>et al.</i> (2020)
Banyan tree	Whole-genome comparison	A chromosome fusion event found in FmChr03, FhChr03 and FhChr07, which was followed by two inversions. Genes within the rearranged regions of FmChr03 and FhChr03 have an association with plant immunity	Zhang <i>et al.</i> (2020)
Apple	Short-read alignment	PAV genes were highly associated with pollination, signal transduction and response to stress	Sun <i>et al.</i> (2020)
<i>Brassica napus</i>	Long-read alignment	SVs played a role in <i>B. napus</i> eco-geographical adaptation and disease resistance	Chawla <i>et al.</i> (2020)
Tomato	Long-read alignment	SVs could change tomato gene dosage and expression levels modified fruit flavour, size and production	Alonge <i>et al.</i> (2020)
Sunflower	Whole-genome comparison	SVs had associations with flowering time and seed size	Todesco <i>et al.</i> (2020)
Soybean	Whole-genome comparison	PAV was a major contributor to driving genome size variation. A 10-kb PAV of a hydrophobic protein-encoding gene may be responsible for seed lustre	Liu <i>et al.</i> (2020)
Wheat	Short-read alignment	23% of the genes were variable and 330 genes were absent from the reference. Variable genes tended to be enriched in functions like protein phosphorylation and protein catabolic process	De Oliveira <i>et al.</i> (2020)

(2020) and Alonge *et al.* (2020) used long-read sequencing to study the role of SVs in plants. In Liu *et al.* (2020), more than 776 000 SVs were discovered in 26 representative soybean accessions. They also identified a 10-kb PAV on chromosome 15 that has a significant association with seed lustre. Alonge *et al.*

(2020) performed a ‘panSV’ study using Oxford Nanopore data for 100 tomato accessions with 238 490 SVs identified. After associating SVs with QTL involved in the metabolism of guaiacol and fruit weight, different haplotypes were resolved that had been missed in previous GWAS.

## Conclusions and perspective

Structural variation represents an important part of genetic diversity in plants and plays a role in phenotypic variation. The limitations of technology and methods used to analyse SVs have previously hindered our understanding of the extent and importance of these variations. With recent advances in DNA sequencing and optical mapping, together with the development of advanced bioinformatics tools, the study of SVs in plants is becoming more common, and there is an increasing awareness that SVs are as important as SNPs and small indels (Wellenreuther *et al.*, 2019).

Although the current technology and methods have dramatically increased the resolution of SV identification, false positives remain. Filtration and further validation are required to make SV detection more reliable. New computational algorithms for SV calling, particularly using long-read sequencing and long-range genomic information, are expected to be developed for plant genome data, which considers different ploidy levels and genome repetitiveness. Refined SV identification pipelines are also needed to further increase sensitivity. Machine learning approaches may be developed to integrate SV calls from different algorithms to reduce false positives. With the improved accuracy and read length in long-read sequencing, haplotype-aware plant genome assemblies are expected to be produced, which could support a detailed mining of allelic or heterozygous variation and the hidden genes missing in current linear assemblies. Techniques, such as Strand-seq (Falconer *et al.*, 2012), can be applied to further assess allele-based SVs, particularly inversions. Other techniques, such as short-read, long-read and direct RNA-sequencing may be useful to check the accuracy of SV identification through gene expression. With more individuals studied, improved visualization of SVs between different individuals, such as using a genome graph, can be produced to display SVs. However, more efforts are needed to solve the problems in using a graph genome, such as finding an efficient way to easily switch sequence coordinates between assemblies.

Mining SVs or genes altered by SVs can be useful for breeding. Currently, there are few methods to directly link SVs with particular phenotypes; therefore, SV-specific genome-wide association study approaches are needed to efficiently associate SVs with phenotypes. Genome editing using such as CRISPR/Cas system provides a way to validate or induce SVs of interest in plants to produce advanced crop varieties (Zhang *et al.*, 2018). Building pangenomes or genus-wide pangenomes (Khan *et al.*, 2020) provides a useful way of mining SV-related genes. To further benefit plant breeding, SV-phenotype-related databases for different species are needed. By searching such databases, breeders and crop researchers can identify candidate SVs that can be used in their breeding programmes to produce improved varieties.

## Acknowledgement

Y.Y. was supported by the China Scholarship Council (CSC). P.B. acknowledges the support of the Forrest Research Foundation. We thank Armin Scheben for his editing of this manuscript.

## Funding

This work was funded by the Australian Research Council (ARC) (grant no: LP160100030).

## Conflict of interest

There is no conflict of interest to disclose.

## Author contributions

Y.Y. drafted this manuscript and prepared all figures and tables. P.B., J.B. and D.E. edited this manuscript. All authors read and approved this manuscript.

## References

- Abel, H.J., Duncavage, E.J., Becker, N., Armstrong, J.R., Magrini, V.J. and Pfeifer, J.D. (2010) SLOPE: a quick and accurate method for locating non-SNP structural variation from targeted next-generation sequence data. *Bioinformatics*, **26**, 2684–2688.
- Abo, R.P., Ducar, M., Garcia, E.P., Thorner, A.R., Rojas-Rudilla, V., Lin, L., Sholl, L.M. *et al.* (2015) BreakMer: detection of structural variation in targeted massively parallel sequencing data using kmers. *Nucleic Acids Res.* **43**, e19.
- 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.
- Alkan, C., Coe, B.P. and Eichler, E.E. (2011) Genome structural variation discovery and genotyping. *Nat. Rev. Genet.* **12**, 363–376.
- Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., Kitzman, J.O. *et al.* (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat. Genet.* **41**, 1061–1067.
- Alonge, M., Wang, X., Benoit, M., Soyk, S., Pereira, L., Zhang, L., Suresh, H. *et al.* (2020) Major impacts of widespread structural variation on gene expression and crop improvement in Tomato. *Cell*, **182**, 145–161.
- Andrews, P.A., Iossifov, I., Kendall, J., Marks, S., Muthuswamy, L., Wang, Z., Levy, D. *et al.* (2016) MUMdex: MUM-based structural variation detection. *bioRxiv*.
- Arabidopsis Genome, I. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.
- Bartenhagen, C. and Dugas, M. (2016) Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms. *Brief. Bioinform.* **17**, 51–62.
- Bayer, P.E., Edwards, D. and Batley, J. (2018) Bias in resistance gene prediction due to repeat masking. *Nat. Plants*, **4**, 762–765.
- Bayer, P.E., Golicz, A.A., Tirnaz, S., Chan, C.K., Edwards, D. and Batley, J. (2019) Variation in abundance of predicted resistance genes in the Brassica oleracea pangenome. *Plant Biotechnol. J.* **17**, 789–800.
- Bayer, P.E., Golicz, A.A., Scheben, A., Batley, J. and Edwards, D. (2020) Plant pan-genomes are the new reference. *Nat. Plants*, **6**, 914–920.
- Bayer, P.E., Hurgobin, B., Golicz, A.A., Chan, C.K., Yuan, Y., Lee, H., Renton, M. *et al.* (2017) Assembly and comparison of two closely related Brassica napus genomes. *Plant Biotechnol. J.* **15**, 1602–1610.
- Bejjani, B.A. and Shaffer, L.G. (2006) Application of array-based comparative genomic hybridization to clinical diagnostics. *J. Mol. Diagn.* **8**, 528–533.
- Bellos, E., Johnson, M.R. and Coin, L.J. (2012) cnvHiTSeq: integrative models for high-resolution copy number variation detection and genotyping using population sequencing data. *Genome Biol.* **13**, R120.
- Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S., Lee, J. *et al.* (2017) Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650.
- Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappel, J., Schleiermacher, G., Janoueix-Lerosey, I. *et al.* (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, **28**, 423–425.
- Cao, H., Hastie, A.R., Cao, D., Lam, E.T., Sun, Y., Huang, H., Liu, X. *et al.* (2014) Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *Gigascience*, **3**, 34.

- Cao, J., Schneeberger, K., Ossowski, S., Gunther, T., Bender, S., Fitz, J., Koenig, D. et al. (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* **43**, 956–963.
- Chawla, H.S., Lee, H., Gabur, I., Vollrath, P., Tamilselvan-Nattar-Amutha, S., Obermeier, C., Schiessl, S.V. et al. (2020) Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant Biotechnol. J.* **19**, 240–250.
- Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D. et al. (2009) BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods*, **6**, 677–681.
- Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Kallberg, M., Cox, A.J. et al. (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*, **32**, 1220–1222.
- Chen, X., Shi, X., Hilakivi-Clarke, L., Shajahan-Haq, A.N., Clarke, R. and Xuan, J. (2017) PSSV: a novel pattern-based probabilistic approach for somatic structural variation identification. *Bioinformatics*, **33**, 177–183.
- Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T. et al. (2017) The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699.
- Choi, M.-H., Sohn, J.-I., Yi, D., Menon, A.V., Kim, Y.J., Kyung, S., Shin, S.-H. et al. (2020) Ultra-fast prediction of somatic structural variations by reduced read mapping via pan-genome-k-mer sets. *bioRxiv*, 2020.2010.2025.354456
- Chong, Z., Ruan, J., Gao, M., Zhou, W., Chen, T., Fan, X., Ding, L. et al. (2017) novoBreak: local assembly for breakpoint detection in cancer genomes. *Nat. Methods*, **14**, 65–67.
- Cretu Stancu, M., van Roosmalen, M.J., Renkens, I., Nieboer, M.M., Middelkamp, S., de Ligt, J., Pregno, G. et al. (2017) Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat. Commun.* **8**, 1326.
- Danilevicz, M.F., Tay Fernandez, C.G., Marsh, J.I., Bayer, P.E. and Edwards, D. (2020) Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* **54**, 18–25.
- Darling, A.C., Mau, B., Blattner, F.R. and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–1403.
- Davis, C.F., Ritter, D.I., Wheeler, D.A., Wang, H., Ding, Y., Dugan, S.P., Bainbridge, M.N. et al. (2016) SV-STAT accurately detects structural variation via alignment to reference-based assemblies. *Source Code Biol. Med.* **11**, 8.
- De Oliveira, R., Rimbart, H., Balfourier, F., Kitt, J., Dymant, E., Vrana, J., Dolezel, J. et al. (2020) Structural variations affecting genes and transposable elements of chromosome 3B in wheats. *Front. Genet.* **11**, 891.
- Dolatabadian, A., Bayer, P.E., Tirnaz, S., Hurgobin, B., Edwards, D. and Batley, J. (2020) Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* **18**, 969–982.
- Dolatabadian, A., Patel, D.A., Edwards, D. and Batley, J. (2017) Copy number variation and disease resistance in plants. *Theor. Appl. Genet.* **130**, 2479–2490.
- Drier, Y., Lawrence, M.S., Carter, S.L., Stewart, C., Gabriel, S.B., Lander, E.S., Meyerson, M. et al. (2013) Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability. *Genome Res.* **23**, 228–235.
- English, A.C., Salerno, W.J. and Reid, J.G. (2014) PBHoney: identifying genomic variants via long-read discordance and interrupted mapping. *BMC Bioinformatics*, **15**, 180.
- Escaramis, G., Docampo, E. and Rabionet, R. (2015) A decade of structural variants: description, history and methods to detect structural variation. *Brief Funct. Genomics*, **14**, 305–314.
- Escaramis, G., Tornador, C., Bassaganyas, L., Rabionet, R., Tubio, J.M.C., Martínez-Fundichely, A., Cáceres, M. et al. (2013) PeSV-Fisher: identification of somatic and non-somatic structural variants using next generation sequencing data. *PLoS One*, **8**, e63377.
- Falconer, E., Hills, M., Naumann, U., Poon, S.S., Chavez, E.A., Sanders, A.D., Zhao, Y. et al. (2012) DNA template strand sequencing of single-cells maps genomic rearrangements at high resolution. *Nat. Methods*, **9**, 1107–1112.
- Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nat. Rev. Genet.* **7**, 85–97.
- Fuentes, R.R., Chebotarov, D., Duitama, J., Smith, S., De la Hoz, J.F., Mohiyuddin, M., Wing, R.A. et al. (2019) Structural variants in 3000 rice genomes. *Genome Res.* **29**, 870–880.
- Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D.M., Burzynski-Chang, E.A. et al. (2019) The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* **51**, 1044–1051.
- Gillet-Markowska, A., Richard, H., Fischer, G. and Lafontaine, I. (2015) Ulysses: accurate detection of low-frequency structural variations in large insert-size sequencing libraries. *Bioinformatics*, **31**, 801–808.
- Goel, M., Sun, H., Jiao, W.B. and Schneeberger, K. (2019) SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277.
- Golicz, A.A., Batley, J. and Edwards, D. (2016a) Towards plant pangenomics. *Plant Biotechnol. J.* **14**, 1099–1105.
- Golicz, A.A., Bayer, P.E., Barker, G.C., Edger, P.P., Kim, H., Martinez, P.A., Chan, C.K. et al. (2016b) The pangenome of an agronomically important crop plant *Brassica oleracea*. *Nat. Commun.* **7**, 13390.
- Golicz, A.A., Bayer, P.E., Bhalla, P.L., Batley, J. and Edwards, D. (2020) Pangenomics comes of age: from bacteria to plant and animal applications. *Trends Genet.* **36**, 132–145.
- Gong, L., Wong, C.H., Cheng, W.C., Tjong, H., Menghi, F., Ngan, C.Y., Liu, E.T. et al. (2018) Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nat. Methods*, **15**, 455–460.
- Goodwin, S., McPherson, J.D. and McCombie, W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **17**, 333–351.
- Guo, J., Cao, K., Deng, C., Li, Y., Zhu, G., Fang, W., Chen, C. et al. (2020) An integrated peach genome structural variation map uncovers genes associated with fruit traits. *Genome Biol.* **21**, 258.
- Gusnanto, A., Wood, H.M., Pawitan, Y., Rabbitts, P. and Berri, S. (2012) Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics*, **28**, 40–47.
- Hajirasouliha, I., Hormozdiari, F., Alkan, C., Kidd, J.M., Birol, I., Eichler, E.E. and Sahinalp, S.C. (2010) Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. *Bioinformatics*, **26**, 1277–1283.
- Hampton, O.A., English, A.C., Wang, M., Salerno, W.J., Liu, Y., Muzny, D.M., Han, Y. et al. (2017) SVachra: a tool to identify genomic structural variation in mate pair sequencing data containing inward and outward facing reads. *BMC Genom.* **18**, 691.
- Handsaker, R.E., Van Doren, V., Berman, J.R., Genovese, G., Kashin, S., Boettger, L.M. and McCarroll, S.A. (2015) Large multiallelic copy number variations in humans. *Nat. Genet.* **47**, 296–303.
- Harris, R.S. (2007) *Improved Pairwise Alignment of Genomic DNA*. The Pennsylvania State University. <http://www.bx.psu.edu/~rsharris/lastz>
- Hayes, M. and Li, J. (2013) Bellerophon: a hybrid method for detecting interchromosomal rearrangements at base pair resolution using next-generation sequencing data. *BMC Bioinformatics*, **14**(Suppl 5), S6.
- Hayes, M., Pyon, Y.S. and Li, J. (2012) A model-based clustering method for genomic structural variant prediction and genotyping using paired-end sequencing data. *PLoS One*, **7**, e25881.
- Heller, D. and Vingron, M. (2019) SVIM: structural variant identification using mapped long reads. *Bioinformatics*, **35**, 2907–2915.
- Heller, D. and Vingron, M. (2020) SVIM-asm: Structural variant detection from haploid and diploid genome assemblies. *bioRxiv*, 2020.2010.2027.356907.
- Hester, S.D., Reid, L., Nowak, N., Jones, W.D., Parker, J.S., Knudtson, K., Ward, W. et al. (2009) Comparison of comparative genomic hybridization technologies across microarray platforms. *J. Biomol. Tech.* **20**, 135–151.
- Hirsch, C.N., Foerster, J.M., Johnson, J.M., Sekhon, R.S., Muttoni, G., Vaillancourt, B., Penagaricano, F. et al. (2014) Insights into the maize pangenome and pan-transcriptome. *Plant Cell*, **26**, 121–135.
- Ho, S.S., Urban, A.E. and Mills, R.E. (2020) Structural variation in the sequencing era. *Nat. Rev. Genet.* **21**, 171–189.
- Hormozdiari, F., Hajirasouliha, I., Dao, P., Hach, F., Yorukoglu, D., Alkan, C., Eichler, E.E. et al. (2010) Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery. *Bioinformatics*, **26**, i350–i357.



- Hubner, S., Bercovich, N., Todesco, M., Mandel, J.R., Odenheimer, J., Ziegler, E., Lee, J.S. *et al.* (2019) Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance. *Nat Plants*, **5**, 54–62.
- Huddleston, J., Chaisson, M.J.P., Steinberg, K.M., Warren, W., Hoekzema, K., Gordon, D., Graves-Lindsay, T.A. *et al.* (2017) Discovery and genotyping of structural variation from long-read haploid genome sequence data. *Genome Res.* **27**, 677–685.
- Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J., Ricci, W.A. *et al.* (2021) *De novo* assembly, annotation, and comparative analysis of 26 diverse maize genomes. *bioRxiv*, 2021.2001.2014.426684.
- Hurgobin, B., Golicz, A.A., Bayer, P.E., Chan, C.K., Tirnaz, S., Dolatabadian, A., Schiessl, S.V. *et al.* (2018) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid Brassica napus. *Plant Biotechnol. J.* **16**, 1265–1274.
- Iqbal, Z., Caccamo, M., Turner, I., Flicek, P. and McVean, G. (2012) *De novo* assembly and genotyping of variants using colored de Bruijn graphs. *Nat. Genet.* **44**, 226–232.
- Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V.S., Gundlach, H., Monat, C., Lux, T. *et al.* (2020) The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, **588**, 284–289.
- Jiang, T., Liu, Y., Jiang, Y., Li, J., Gao, Y., Cui, Z., Liu, Y. *et al.* (2020) Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol.* **21**, 189.
- Jiang, Y., Wang, Y. and Brudno, M. (2012) PRISM: pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics*, **28**, 2576–2583.
- Jiao, W.B. and Schneeberger, K. (2020) Chromosome-level assemblies of multiple Arabidopsis genomes reveal hotspots of rearrangements with altered evolutionary dynamics. *Nat. Commun.* **11**, 989.
- Karakoc, E., Alkan, C., O’Roak, B.J., Dennis, M.Y., Vives, L., Mark, K., Rieder, M.J. *et al.* (2011) Detection of structural variants and indels within exome data. *Nat. Methods*, **9**, 176–178.
- Keane, T.M., Wong, K. and Adams, D.J. (2013) RetroSeq: transposable element discovery from next-generation sequencing data. *Bioinformatics*, **29**, 389–390.
- Khan, A.W., Garg, V., Roorkiwal, M., Golicz, A.A., Edwards, D. and Varshney, R.K. (2020) Super-Pangenome by Integrating the Wild Side of a Species for Accelerated Crop Improvement. *Trends Plant Sci.* **25**, 148–158.
- Kim, T.M., Luquette, L.J., Xi, R. and Park, P.J. (2010) rSW-seq: algorithm for detection of copy number alterations in deep sequencing data. *BMC Bioinformatics*, **11**, 432.
- Klambauer, G., Schwarzbauer, K., Mayr, A., Clevert, D.A., Mitterecker, A., Bodenhofer, U. and Hochreiter, S. (2012) cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **40**, e69.
- Korbel, J.O., Abyzov, A., Mu, X.J., Carriero, N., Cayting, P., Zhang, Z., Snyder, M. *et al.* (2009) PEMer: a computational framework with simulation-based error models for inferring genomic structural variants from massive paired-end sequencing data. *Genome Biol.* **10**, R23.
- Kou, Y., Liao, Y., Toivainen, T., Lv, Y., Tian, X., Emerson, J.J., Gaut, B.S. *et al.* (2020) Evolutionary genomics of structural variation in Asian rice (*Oryza sativa*) domestication. *Mol. Biol. Evol.* **37**, 3507–3524.
- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. and Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.
- Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P. *et al.* (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776.
- Lam, H.Y., Mu, X.J., Stutz, A.M., Tanzer, A., Cayting, P.D., Snyder, M., Kim, P.M. *et al.* (2010) Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28**, 47–55.
- 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, S., Hormozdiari, F., Alkan, C. and Brudno, M. (2009) MoDIL: detecting small indels from clone-end sequencing with mixtures of distributions. *Nat. Methods*, **6**, 473–474.
- Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100.
- Li, S., Li, R., Li, H., Lu, J., Li, Y., Bolund, L., Schierup, M.H. *et al.* (2013) SOAPindel: efficient identification of indels from short paired reads. *Genome Res.* **23**, 195–200.
- Li, Y.H., Zhou, G., Ma, J., Jiang, W., Jin, L.G., Zhang, Z., Guo, Y. *et al.* (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotechnol.* **32**, 1045–1052.
- Liang, Y., Qiu, K., Liao, B., Zhu, W., Huang, X., Li, L., Chen, X. *et al.* (2017) Seeksv: an accurate tool for somatic structural variation and virus integration detection. *Bioinformatics*, **33**, 184–191.
- Lin, G., He, C., Zheng, J., Koo, D.-H., Le, H., Zheng, H., Tamang, T.M. *et al.* (2020) Chromosome-level Genome Assembly of a Regenerable Maize Inbred Line A188. *bioRxiv*, 2020.2009.2009.289611.
- Lin, K., Zhang, N., Severing, E.I., Nijveen, H., Cheng, F., Visser, R.G., Wang, X. *et al.* (2014) Beyond genomic variation—comparison and functional annotation of three Brassica rapa genomes: a turnip, a rapid cycling and a Chinese cabbage. *BMC Genom.* **15**, 250.
- Lindberg, M.R., Hall, I.M. and Quinlan, A.R. (2015) Population-based structural variation discovery with Hydra-Multi. *Bioinformatics*, **31**, 1286–1289.
- Liu, Y., Du, H., Li, P., Shen, Y., Peng, H., Liu, S., Zhou, G.A. *et al.* (2020) Pan-Genome of Wild and Cultivated Soybeans. *Cell*, **182**, 162–176.
- Lopez, G., Egolf, L.E., Giorgi, F.M., Diskin, S.J. and Margolin, A.A. (2020) Sypluscnv: analysis and visualization of complex structural variation data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btaa878>. Epub ahead of print. PMID: 33051644.
- Lu, F., Romy, M.C., Glaubitz, J.C., Bradbury, P.J., Elshire, R.J., Wang, T., Li, Y. *et al.* (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* **6**, 6914.
- Ma, X., Fan, J., Wu, Y., Zhao, S., Zheng, X., Sun, C. and Tan, L. (2020) Whole-genome *de novo* assemblies reveal extensive structural variations and dynamic organelle-to-nucleus DNA transfers in African and Asian rice. *Plant J.* **104**, 596–612.
- Magi, A., Benelli, M., Yoon, S., Roviello, F. and Torricelli, F. (2011) Detecting common copy number variants in high-throughput sequencing data by using JointSLM algorithm. *Nucleic Acids Res.* **39**, e65.
- Magris, G., Pinosio, S., Marroni, F. and Gaspero, G.D. (2015) From one to the many genomes of a plant: the evolution of the grapevine pan-genome. In: *Plant & Animal Genome Conference XXIII*. San Diego, USA.
- Mamidi, S., Healey, A., Huang, P., Grimwood, J., Jenkins, J., Barry, K., Sreedasyam, A. *et al.* (2020) A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci. *Nat. Biotechnol.* **38**, 1203–1210.
- Marschall, T., Costa, I.G., Canzar, S., Bauer, M., Klau, G.W., Schliep, A. and Schonhuth, A. (2012) CLEVER: clique-enumerating variant finder. *Bioinformatics*, **28**, 2875–2882.
- Medvedev, P., Fiume, M., Dzamba, M., Smith, T. and Brudno, M. (2010) Detecting copy number variation with mated short reads. *Genome Res.* **20**, 1613–1622.
- Meyers, L.A. and Levin, D.A. (2006) On the abundance of polyploids in flowering plants. *Evolution*, **60**, 1198–1206.
- Michael, T.P., Jupe, F., Bemm, F., Motley, S.T., Sandoval, J.P., Lanz, C., Loudet, O. *et al.* (2018) High contiguity Arabidopsis thaliana genome assembly with a single nanopore flow cell. *Nat. Commun.* **9**, 541.
- Michael, T.P. and VanBuren, R. (2020) Building near-complete plant genomes. *Curr. Opin. Plant Biol.* **54**, 26–33.
- Miller, C.A., Hampton, O., Coarfa, C. and Milosavljevic, A. (2011) ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads. *PLoS One*, **6**, e16327.
- Mohiyuddin, M., Mu, J.C., Li, J., Bani Asadi, N., Gerstein, M.B., Abyzov, A., Wong, W.H., *et al.* (2015) MetaSV: an accurate and integrative structural-variant caller for next generation sequencing. *Bioinformatics*, **31**, 2741–2744.
- Moncunill, V., Gonzalez, S., Bea, S., Andrieux, L.O., Salaverria, I., Royo, C., Martinez, L. *et al.* (2014) Comprehensive characterization of complex structural variations in cancer by directly comparing genome sequence reads. *Nat. Biotechnol.* **32**, 1106–1112.

- Montenegro, J.D., Golicz, A.A., Bayer, P.E., Hurgobin, B., Lee, H.T., Chan, C.-K.-K., Visendi, P. et al. (2017) The pangenome of hexaploid bread wheat. *Plant J.* **90**, 1007–1013.
- Narzisi, G., O’Rawe, J.A., Iossifov, I., Fang, H., Lee, Y.H., Wang, Z., Wu, Y. et al. (2014) Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods*, **11**, 1033–1036.
- Nattestad, M. and Schatz, M.C. (2016) Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**, 3021–3023.
- Newman, A.M., Bratman, S.V., Stehr, H., Lee, L.J., Liu, C.L., Diehn, M. and Alizadeh, A.A. (2014) FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. *Bioinformatics*, **30**, 3390–3393.
- Nijkamp, J.F., van den Broek, M.A., Geertman, J.M., Reinders, M.J., Daran, J.M. and de Ridder, D. (2012) De novo detection of copy number variation by co-assembly. *Bioinformatics*, **28**, 3195–3202.
- Ou, L., Li, D., Lv, J., Chen, W., Zhang, Z., Li, X., Yang, B. et al. (2018) Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses. *New Phytol.* **220**, 360–363.
- PacificBiosciences (2018) *pbsv*. <https://github.com/PacificBiosciences/pbsv>
- Park, H., Kim, J.I., Ju, Y.S., Gokcumen, O., Mills, R.E., Kim, S., Lee, S. et al. (2010) Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. *Nat. Genet.* **42**, 400–405.
- Perumal, S., Koh, C.S., Jin, L., Buchwaldt, M., Higgins, E.E., Zheng, C., Sankoff, D. et al. (2020) A high-contiguity *Brassica nigra* genome localizes active centromeres and defines the ancestral *Brassica* genome. *Nat. Plants*, **6**, 929–941.
- Pinosio, S., Giacomello, S., Faivre-Rampant, P., Taylor, G., Jorge, V., Le Paslier, M.C., Zaina, G. et al. (2016) Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Mol. Biol. Evol.* **33**, 2706–2719.
- Pucker, B., Holtgrawe, D., Stadermann, K.B., Frey, K., Huettel, B., Reinhardt, R. and Weissshaar, B. (2019) A chromosome-level sequence assembly reveals the structure of the *Arabidopsis thaliana* Nd-1 genome and its gene set. *PLoS One*, **14**, e0216233.
- Qi, J. and Zhao, F. (2011) inGAP-sv: a novel scheme to identify and visualize structural variation from paired end mapping data. *Nucleic Acids Res.* **39**, W567–W575.
- Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V. and Korbel, J.O. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, **28**, i333–i339.
- Read, B.A., Kegel, J., Klute, M.J., Kuo, A., Lefebvre, S.C., Maumus, F., Mayer, C. et al. (2013) Pan genome of the phytoplankton *Emiliania* underpins its global distribution. *Nature*, **499**, 209–213.
- Schatz, M.C., Maron, L.G., Stein, J.C., Hernandez Wences, A., Gurtowski, J., Biggers, E., Lee, H. et al. (2014) Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. *Genome Biol.* **15**, 506.
- Schroder, J., Hsu, A., Boyle, S.E., Macintyre, G., Cmero, M., Tothill, R.W., Johnstone, R.W. et al. (2014) Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. *Bioinformatics*, **30**, 1064–1072.
- Sedlazeck, F.J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A. and Schatz, M.C. (2018) Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods*, **15**, 461–468.
- Shelton, J.M., Coleman, M.C., Herndon, N., Lu, N., Lam, E.T., Anantharaman, T., Sheth, P., et al. (2015) Tools and pipelines for BioNano data: molecule assembly pipeline and FASTA super scaffolding tool. *BMC Genom.* **16**, 734.
- Simonikova, D., Nemeckova, A., Cizkova, J., Brown, A., Swennen, R., Dolezel, J. and Hribova, E. (2020) Chromosome painting in cultivated bananas and their wild relatives (*Musa* spp.) reveals differences in chromosome structure. *Int. J. Mol. Sci.*, **21**, 7915.
- Simpson, J.T., McIntyre, R.E., Adams, D.J. and Durbin, R. (2010) Copy number variant detection in inbred strains from short read sequence data. *Bioinformatics*, **26**, 565–567.
- Sindi, S.S., Onal, S., Peng, L.C., Wu, H.T. and Raphael, B.J. (2012) An integrative probabilistic model for identification of structural variation in sequencing data. *Genome Biol.* **13**, R22.
- Song, J.M., Guan, Z., Hu, J., Guo, C., Yang, Z., Wang, S., Liu, D. et al. (2020) Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants*, **6**, 34–45.
- Spielmann, M., Lupianez, D.G. and Mundlos, S. (2018) Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467.
- Sun, C., Hu, Z., Zheng, T., Lu, K., Zhao, Y., Wang, W., Shi, J. et al. (2017) RPAN: rice pan-genome browser for approximately 3000 rice genomes. *Nucleic Acids Res.* **45**, 597–605.
- Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A. et al. (2020) Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* **52**, 1423–1432.
- Suzuki, S., Yasuda, T., Shiraishi, Y., Miyano, S. and Nagasaki, M. (2011) ClipCrop: a tool for detecting structural variations with single-base resolution using soft-clipping information. *BMC Bioinformatics*, **12**(Suppl 14), S7.
- Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.I., Angiuoli, S.v. et al. (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ‘pan-genome’. *Proc. Natl Acad. Sci. USA*, **102**, 13950–13955.
- Tham, C.Y., Tirado-Magallanes, R., Goh, Y., Fullwood, M.J., Koh, B.T.H., Wang, W., Ng, C.H. et al. (2019) NanoVar: accurate characterization of patients’ genomic structural variants using low-depth nanopore sequencing. *bioRxiv*, 662940.
- Todesco, M., Owens, G.L., Bercovich, N., Legare, J.S., Souidi, S., Burge, D.O., Huang, K. et al. (2020) Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, **584**, 602–607.
- Trappe, K., Emde, A.K., Ehrlich, H.C. and Reinert, K. (2014) Gustaf: Detecting and correctly classifying SVs in the NGS twilight zone. *Bioinformatics*, **30**, 3484–3490.
- Unterseer, S., Seidel, M.A., Bauer, E., Haberer, G., Hochholdinger, F., Opitz, N. and Marcon, C. et al. (2017) European Flint reference sequences complement the maize pan-genome. *bioRxiv*.
- Valliyodan, B., Brown, A.V., Wang, J., Patil, G., Liu, Y., Otyama, P.I., Nelson, R.T. et al. (2021) Genetic variation among 481 diverse soybean accessions, inferred from genomic re-sequencing. *Sci Data*, **8**, 50.
- Varshney, R.K., Thudi, M., Roorkival, M., He, W., Upadhyaya, H.D., Yang, W., Bajaj, P. et al. (2019) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* **51**, 857–864.
- Vogel, J.P., Gordon, S., Contreras-Moreira, B., Marais, D.L.D., Burgess, D., Schackwitz, W., Tyler, L. et al. (2016) The pan-genome of *Brachypodium distachyon*, capturing the full genetic complement of a plant species. In: *Plant & Animal Genome Conference XXIV*. San Diego, USA.
- Wala, J.A., Bandopadhyay, P., Greenwald, N.F., O’Rourke, R., Sharpe, T., Stewart, C., Schumacher, S. et al. (2018) SvABA: genome-wide detection of structural variants and indels by local assembly. *Genome Res.* **28**, 581–591.
- Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H. et al. (2020) Multiple wheat genomes reveal global variation in modern breeding. *Nature*, **588**, 277–283.
- Wang, J., Mullighan, C.G., Easton, J., Roberts, S., Heatley, S.L., Ma, J., Rusch, M.C. et al. (2011) CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods*, **8**, 652–654.
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M. et al. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, **557**, 43–49.
- Wang, X., Gao, L., Jiao, C., Stravrovadis, S., Hosmani, P.S., Saha, S., Zhang, J. et al. (2020) Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat. Commun.* **11**, 5817.
- Wang, Z., Hormozdiari, F., Yang, W.Y., Halperin, E. and Eskin, E. (2013) CNVnM: copy number variation detection using uncertainty of read mapping. *J. Comput. Biol.* **20**, 224–236.
- Wei, Q., Wang, J., Wang, W., Hu, T., Hu, H. and Bao, C. (2020) A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Hortic. Res.* **7**, 153.
- Wei, Y.C. and Huang, G.H. (2020) CONY: A Bayesian procedure for detecting copy number variations from sequencing read depths. *Sci. Rep.* **10**, 10493.

- Wellenreuther, M., Merot, C., Berdan, E. and Bernatchez, L. (2019) Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Mol. Ecol.* **28**, 1203–1209.
- Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Concepcion, G.T., Ebler, J. *et al.* (2019) Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **37**, 1155–1162.
- Wong, K., Keane, T.M., Stalker, J. and Adams, D.J. (2010) Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biol.* **11**, R128.
- Wong, K.H.Y., Levy-Sakin, M. and Kwok, P.Y. (2018) De novo human genome assemblies reveal spectrum of alternative haplotypes in diverse populations. *Nat. Commun.* **9**, 3040.
- Xi, R., Hadjipanayis, A.G., Luquette, L.J., Kim, T.M., Lee, E., Zhang, J., Johnson, M.D. *et al.* (2011) Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proc. Natl Acad. Sci. USA*, **108**, E1128–E1136.
- Xie, C. and Tammi, M.T. (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*, **10**, 80.
- Xie, M., Chung, C.Y., Li, M.W., Wong, F.L., Wang, X., Liu, A., Wang, Z. *et al.* (2019) A reference-grade wild soybean genome. *Nat. Commun.* **10**, 1216.
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., Huang, J. *et al.* (2019) Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* **51**, 1052–1059.
- Ye, K., Hall, G. and Ning, Z. (2016) Structural variation detection from next generation sequencing. *J. Next Gen. Seq. Appl.* **51**, 7.
- Ye, K., Schulz, M.H., Long, Q., Apweiler, R. and Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, **25**, 2865–2871.
- Yin, D., Ji, C., Song, Q., Zhang, W., Zhang, X., Zhao, K., Chen, C.Y. *et al.* (2020) Comparison of arachis monticola with diploid and cultivated tetraploid genomes reveals asymmetric subgenome evolution and improvement of peanut. *Adv Sci (Weinh)*, **7**, 1901672.
- Yoon, S., Xuan, Z., Makarov, V., Ye, K. and Sebat, J. (2009) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* **19**, 1586–1592.
- Yu, J., Golicz, A.A., Lu, K., Dossa, K., Zhang, Y., Chen, J., Wang, L. *et al.* (2019) Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars. *Plant Biotechnol. J.* **17**, 881–892.
- Yuan, Y., Bayer, P.E., Batley, J. and Edwards, D. (2017) Improvements in Genomic Technologies: Application to Crop Genomics. *Trends Biotechnol.* **35**, 547–558.
- Yuan, Y., Chung, C.Y. and Chan, T.F. (2020) Advances in optical mapping for genomic research. *Comput. Struct. Biotechnol. J.* **18**, 2051–2062.
- Yuan, Y., Milec, Z., Bayer, P.E., Vrána, J., Doležel, J., Edwards, D., Erskine, W. *et al.* (2018) Large-scale structural variation detection in subterranean clover subtypes using optical mapping. *Front. Plant Sci.* **9**, 971.
- Zeitouni, B., Boeva, V., Janoueix-Lerosey, I., Loeillet, S., Legoix-ne, P., Nicolas, A., Delattre, O., *et al.* (2010) SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics*, **26**, 1895–1896.
- Zhang, X., Wang, G., Zhang, S., Chen, S., Wang, Y., Wen, P., Ma, X. *et al.* (2020) Genomes of the banyan tree and pollinator wasp provide insights into Fig-Wasp coevolution. *Cell*, **183**, 875–889.
- Zhang, Y., Massel, K., Godwin, I.D. and Gao, C. (2018) Applications and potential of genome editing in crop improvement. *Genome Biol.* **19**, 210.
- Zhang, Z., Mao, L., Chen, H., Bu, F., Li, G., Sun, J., Li, S. *et al.* (2015) Genome-Wide Mapping of Structural Variations Reveals a Copy Number Variant That Determines Reproductive Morphology in Cucumber. *Plant Cell*, **27**, 1595–1604.
- Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., Ruggieri, V. *et al.* (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* **51**, 1607–1615.
- Zhao, J., Bayer, P.E., Ruperao, P., Saxena, R.K., Khan, A.W., Golicz, A.A., Nguyen, H.T. *et al.* (2020) Trait associations in the pangenome of pigeon pea (*Cajanus cajan*). *Plant Biotechnol. J.* **18**, 1946–1954.
- Zhou, P., Silverstein, K.A., Stupar, R.M., Bharti, A.K., Farmer, A.D., Mudge, J., May, G.D. *et al.* (2014) The Medicago pan-genome reveals large-scale variation. In: *Plant & Animal Genome Conference XXII*. San Diego, USA.
- Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D. *et al.* (2019) The population genetics of structural variants in grapevine domestication. *Nat Plants*, **5**, 965–979.