Theoretical and Applied Genetics (2019) 132:733–750
https://doi.org/10.1007/s00122-018-3233-0

REVIEW ARTICLE

Connecting genome structural variation with complex traits in crop plants

Iulian Gabur[1] · Harmeet Singh Chawla[1] · Rod J. Snowdon[1] · Isobel A. P. Parkin[2]

Received: 15 August 2018 / Accepted: 7 November 2018 / Published online: 17 November 2018

Abstract

Key message **Structural genome variation is a major determinant of useful trait diversity. We describe how genome analysis methods are enabling discovery of trait-associated structural variants and their potential impact on breeding.**

Abstract As our understanding of complex crop genomes continues to grow, there is growing evidence that structural genome variation plays a major role in determining traits important for breeding and agriculture. Identifying the extent and impact of structural variants in crop genomes is becoming increasingly feasible with ongoing advances in the sophistication of genome sequencing technologies, particularly as it becomes easier to generate accurate long sequence reads on a genome-wide scale. In this article, we discuss the origins of structural genome variation in crops from ancient and recent genome duplication and polyploidization events and review high-throughput methods to assay such variants in crop populations in order to find associations with phenotypic traits. There is increasing evidence from such studies that gene presence–absence and copy number variation resulting from segmental chromosome exchanges may be at the heart of adaptive variation of crops to counter abiotic and biotic stress factors. We present examples from major crops that demonstrate the potential of pangenomic diversity as a key resource for future plant breeding for resilience and sustainability.

Introduction: the discovery of structural variation

With rapidly increasing sophistication in genome analysis technologies, there is growing evidence that genome-wide structural variation (SV) is a major factor underlining observed phenotypic variation in eukaryotic organisms. The first report of genic SV affecting a phenotype dates back more than 80 years, when Bridges (1936) discovered that a duplication of the *Bar* gene is associated with small eyes in the fruit fly, *Drosophila*. Genomic rearrangements have been studied extensively in humans due to their association with a range of diseases. Particularly copy number variation (CNV), an important class of structural variation, has been discovered to be causal for various autoimmune disorders (Mamtani et al. 2010), including susceptibility to human immunodeficiency virus (HIV) infection (Gonzalez 2005), Parkinson's disease (Singleton 2003) and Alzheimer's disease (Rovelet-Lecrux et al. 2006; Escaramís et al. 2015). Mounting evidence supporting the importance of SV in human genetics led to the study of the same phenomena in animal species, where numerous examples have been discovered for a role of SV in important traits, for example in mice (Keane et al. 2014), cattle (Fadista et al. 2010), pigs (Esteve-Codina et al. 2013), sheep (Liu et al. 2013) and horses (Ghosh et al. 2014; Wang et al. 2014). Structural variations were initially thought to be rare in plants, but this perspective changed dramatically with the realization that almost all flowering plants derived from multiple rounds of ancient or recent polyploidization (Vialette-Guiraud et al. 2011; Van de Peer et al. 2009; Alix et al. 2017). The ability to generate reference genome sequences even for complex crop plant genomes (Edwards et al. 2013) combined with decreased costs associated with de novo genome assembly and resequencing have accelerated the study of SV (Voss-Fels and Snowdon 2016). Numerous recent reports have clearly demonstrated that both small and large genomic

Communicated by Rajeev K. Varshney.

Iulian Gabur and Harmeet Singh Chawla contributed equally to this work.

✉ Rod J. Snowdon
rod.snowdon@agrar.uni-giessen.de

[1] Department of Plant Breeding, Justus Liebig University, Heinrich-Buff-Ring 26-32, 35392 Giessen, Germany

[2] Agriculture and Agri-Food Canada, 107 Science Place, Saskatoon, SK S7N OX2, Canada

rearrangements can cause major phenotypic variance affecting an array of important traits in crops (Saxena et al. 2014; Neik et al. 2017; Żmieńko et al. 2014; Schiessl et al. 2017a).

Diversity of structural variants

Genome structural variants occur in diverse forms including translocations, inversions, insertion/deletion polymorphisms (InDels), copy number variation (CNV), or simple variation in microsatellite repeat number. Traditionally, InDels have been defined as short presence/absence nucleotide polymorphism ranging from 1 to 50 bp in length, whereas a variable number of copies for larger DNA segments, ranging from a few hundred bp to several kb, is generally referred to as CNV. Gene CNV represents the most intensively studied class of SV associated directly with trait variation, whereby variants affecting intergenic regions, splicing variants and/or regulatory factors could also infer SV–trait associations. This can be mainly attributed to their ease of detection using simple molecular biology methods. Presence–absence variation (PAV) represents an extreme form of CNV where whole genomic segments are deleted from individuals within a population (Saxena et al. 2014). Different kinds of SV can occur independently or simultaneously, resulting in complex genome alterations. Many important crop genomes arose from multiple polyploidy events, in some cases involving widespread recombination among homoeologous (related but non-homologous) chromosomes. Such exchanges can result in both reciprocal or non-reciprocal exchanges. The latter, often referred to as homoeologous non-reciprocal transpositions, or HNRT (Parkin et al. 1995; Pires et al. 2004; Gaeta and Chris Pires 2010), can lead to loss or gain of DNA fragments on related chromosome homoeologues and consequently to PAV and CNV. As described in more detail later in this review, examples in recent allopolyploids like *Brassica napus* have demonstrated that this kind of exchange during early rounds of polyploidization can be a key driver of modern crop genome diversity and phenotypic plasticity (Chalhoub et al. 2014; Samans et al. 2017; Hurgobin et al. 2017).

Origins of SV

Various cellular mechanisms can trigger generation of SV during meiotic or mitotic cell division. SV events are caused by recombination errors, like non-allelic homologous recombination (NAHR) (Lupski 1998), DNA break repair errors, such as non-homologous end joining (NHEJ) (Moore and Haber 1996), or replication errors, including fork stalling and template switching (FoSTeS) (Lee et al. 2007) and microhomology-mediated break-induced replication (MMBIR) (Hastings et al. 2009). NHEJ can be triggered by misguided fusion of double-strand breaks in DNA, often resulting in insertions and/or deletions; however, in rare cases NHEJ might also generate translocations (McVey and Lee 2008). FoSTeS/MMBIR is another cellular mechanism causing major structural variations (for example large rearrangements, inversions, duplications and translocations) ranging in size from a few kb to several Mb and involves fork stalling and polymerase switching at a nearby single-stranded DNA (Stankiewicz and Lupski 2010). The most likely cause of much of the CNVs observed in plants is NAHR, which is largely the result of misalignment in genomic regions housing highly identical sequences, such as repetitive DNA, leading to duplication or deletion of genomic segments and thus copy number variants. Segmental duplications appear when highly homologous genomic regions (more than 95%) are physically positioned at distances from a few kb to some Mb from one another. Furthermore, depending on the orientation of the homology, NAHR could also cause deletions (upstream orientation on the same chromosome), inversion (downstream orientation on the same chromosome) and translocation (located on different chromosomes) (Sharp et al. 2006). The abundance of repetitive sequences in plant genomes varies widely, with published frequencies ranging from around 10% in Arabidopsis (The Arabidopsis Genome Initiative 2000) to more than 85% (in wheat) (Appels et al. 2018). The prevalence of repetitive DNA, in particular in larger crop genomes, could promote the generation of dosage effects for particular sets of genes, partly explaining the large adaptive phenotypic variation existing within the plant kingdom.

Changes in ploidy can also lead to generation of SV in plants. The majority of angiosperms studied to date show evidence of polyploidization and/or whole-genome duplication in their evolutionary history, and most modern crop species have undergone recent genome duplication events that are now known to have played a significant role in dictating their path to adaptation (Fig. 1). Some major crops contain multiple copies of entire chromosomes from spontaneous genome duplication of the same species, for example autopolyploid potato (*Solanum tuberosum*; $2n = 4x = 48$), while others arose from interspecific hybridization of subgenomes among distinct, yet related species, for example allohexaploid wheat (*Triticum aestivum*; $2n = 6x = 42$) or allotetraploid oilseed rape/canola (*Brassica napus*; $2n = 4x = 38$). Many older crop species arose by ancient duplications and paleopolyploidization. For example, the diploid cabbage species *Brassica oleracea* ($2n = 2x = 18$) and *Brassica rapa* ($2n = 2x = 20$) represent paleohexaploids which have returned to a diploid state by genome fractionation (Lagercrantz et al. 1996; Tang et al. 2012; Parkin et al. 2014).

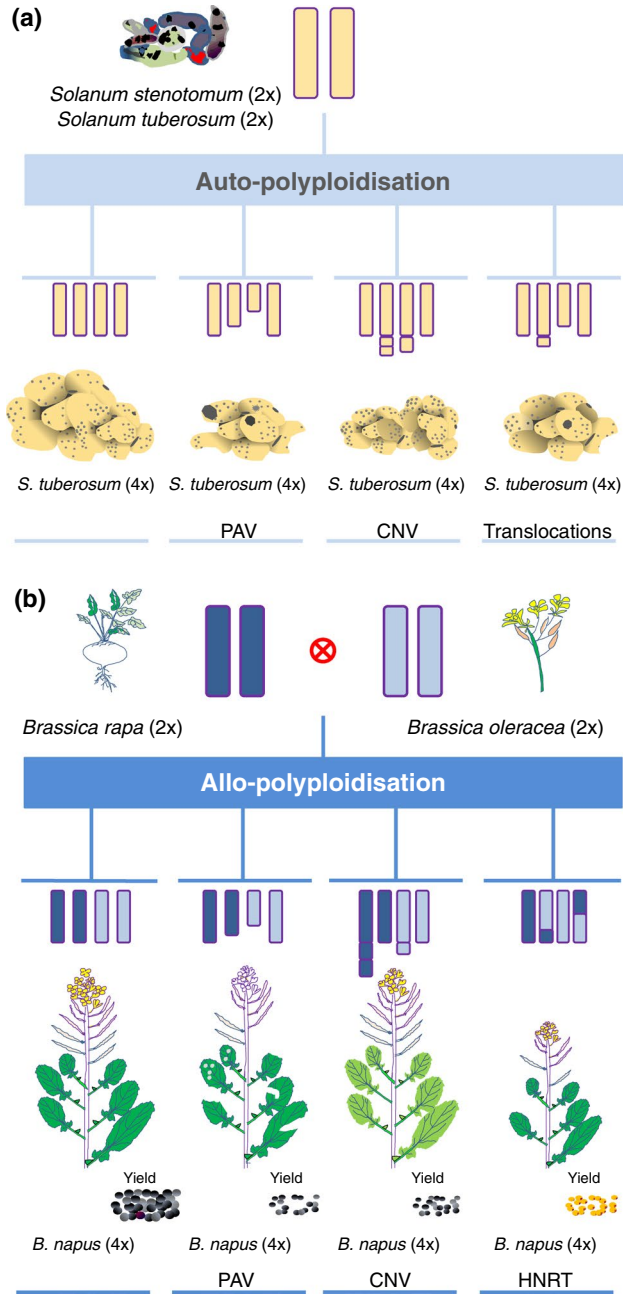

Fig. 1 Origins of different kinds of structural variants in autopolyploid and allopolyploid crops from segmental chromosome rearrangements, illustrated by the coloured bars with examples from **a** autopolyploid potato (*Solanum tuberosum*, 2n = 4x = 48) and **b** allopolyploid rapeseed/canola (*Brassica napus*, 2n = 4x = 38). Autotetraploid *S. tuberosum* arose from a spontaneous genome duplication (auto-polyploidization) of the diploid progenitor *S. stenotomum*/*S. tuberosum* (2n = 2x = 24), while *B. napus* arose from interspecific hybridization between the diploid progenitor species *B. oleracea* (2n = 2x = 18) and *B. rapa* (2n = 2x = 20). SV linked to adaptive and agronomic diversity is represented as presence–absence variation (PAV), copy number variation (CNV), translocations and homoeologous non-reciprocal transpositions (HNRT)

Visualization of large-scale SV

Classical cytology first identified evidence of large-scale chromosomal aberrations in cereals (e.g. Sears 1939), which were later confirmed as translocations using early molecular marker technologies (e.g. Gale and Devos 1998). Comparative genomic hybridization (CGH) was one of the very first methods to visualize large-scale SVs. With CGH it is possible to detect and map relative DNA sequence copy number between genomes, by hybridizing fluorescently labelled DNA from each source genome to metaphase chromosome spreads or genome-wide sequence arrays. An increase or decrease in copy number of genomic DNA (corresponding to segmental SV) can then be detected by measuring the florescence ratios between the two coloured fluorophores (Kallioniemi et al. 1992). The resolution of CGH via in situ hybridization is relatively low, with segmental SV events only visible at megabase scale, whereas array-based CGH (aCGH) can resolve smaller SV events down to a few kilobases in size. For example, Yu et al. (2011) were able to detect 641 CNVs ranging from 1.1 to 180.7 kb between two rice cultivars using aCGH. Large SV events can also be visualized directly at the chromosome level using molecular cytogenetic techniques such as fluorescence in situ hybridization (FISH) and genomic in situ hybridization (GISH) (Xiong et al. 2011; Chester et al. 2012; Horn et al. 2002; Snowdon 2007). These techniques allow physical analysis of chromosomes using chromosome arm ratios, mapping of heterochromatic regions, bacterial artificial chromosome probes (BAC-FISH) containing specific repeat sequences or molecular karyotyping of chromosome-specific probes (Xiong and Pires 2011). FISH has been used successfully in maize to analyse B chromosome non-disjunction due to *r-X1* deficiency (Tseng et al. 2017). Furthermore, the technique has been used to map homoeologous exchanges associated with agronomic traits in polyploid crop genomes. Stein et al. (2017) used BAC-FISH to identify homoeologous exchanges between two *B. napus* chromosomes associated with a QTL for seed fibre content. In potato, FISH was used to identify CNV associated with plant growth and developmental traits (Iovene et al. 2013), while Ali et al. (2016) used FISH in wheat to validate introgression of alien DNA segments that led to mosaic virus resistance. Before the availability of cheap, high-throughput genome sequencing, hybridization methods provided a relatively simple and low cost option for visualizing large SV events at the chromosomal level; however, finer resolution is required for detection of smaller SV events. With sufficient sequence depth, approaches based on next-generation sequencing (NGS) technologies provide an ideal solution.

Sequencing-based SV detection

Next-generation sequencing (NGS) approaches have accelerated the process of assembling plant reference genomes to a speed and accuracy that was unimaginable a decade ago. Furthermore, availability of methods to detect single nucleotide differences between genomes using whole-genome sequencing data, high-coverage exome sequence data or sequence capture data has been a major breakthrough in deciphering complex SV (Chen et al. 2008; Schiessl et al. 2017b). One of the key advantages of NGS-based methods for SV detection is the resolution that can be achieved by using such approaches. To date, sequenced reference genomes of varying quality are available for over 200 plant species, including most major crops (see http://www.plabipd.de/timeline_view.ep for an up-to-date overview of published plant reference genomes). As the quality of reference genome assemblies for more complex genomes continues to improve, for example by utilization of new methods like chromatin conformation associations (e.g. Mascher et al. 2017 or long-read single-molecule sequencing (Jiao et al. 2017), our ability to utilize genome-wide or targeted resequencing techniques for SV analysis in large populations will become even more powerful. Early whole-genome resequencing studies in major crops with relatively simple genomes, like sorghum, used reference-based read-mapping approaches to identify genome-wide SNP and small-scale SV. For example, Mace et al. (2013) identified 1.9 million InDels, including specific gene PAV associated with domestication and breeding, in high-coverage resequencing data from 44 genetically and geographically diverse *Sorghum bicolor* accessions. Different approaches have been developed for characterization of SVs from NGS reads, including combinations of read depth (RD), paired read (PR) and split read (SR) analysis along with de novo sequence assemblies in order to address more complex genomic re-arrangements.

Algorithms for RD analysis rely on the density of sequenced reads aligned to a locus in a reference genome for CNV identification (Alkan et al. 2009; Li and Olivier 2013). In crops, RD approaches focused mainly on calling of large SV, for example in resequencing data from sorghum (Zheng et al. 2011) or rapeseed (Samans et al. 2017). RD-based methods can detect deletions and duplications very effectively. However, limitations of the read length and the quality and coverage of the available reference genomes reduce the efficacy of this approach for detecting insertions or translocations. Furthermore, it should be noted that RD-based approaches are highly sensitive to library preparation methods. For example, PCR amplification during the sequencing library preparation can lead to either over-representation or complete absence of certain genomic regions that are difficult to amplify, which would be interpreted as duplication and deletion events, respectively, using an RD sensing algorithm. Therefore, stringent normalization of read depth is required to remove any kind of bias introduced by the library preparation. These limitations can be somewhat overcome by paired-end sequencing of single DNA fragments (paired reads). Since the sequencing library is enriched for a particular fragment size, the distance between the paired reads is pre-defined. Any insertion or deletion between the paired reads would result in a deviation from the expected mapping distance; hence, abnormally mapped read pairs might indicate the presence of SV (Korbel et al. 2007). Different types of SV can be mapped using paired reads, for example deletions or insertions (when paired reads align further apart or closer than expected), inversions (when the orientation of paired reads is inverted) or translocations (when each of the paired reads maps to different chromosomes). This approach is still highly dependent on the read coverage, size and number of repetitive elements in the genome and the quality of the reference genome, and paired read methods are best suited for detection of medium-sized insertions and deletions. However, they might not be the best choice for identification of small insertion or deletion events, due to the difficulty in distinguishing small deviations in read–pair distance from technical errors. Split read algorithms provide an alternative which also makes use of paired-end sequencing. Split read algorithms depend on accurate mapping of one of the reads from a pair, while the other read maps only partially to the reference genome (Ye et al. 2009). When reads align right across a SV break point, precise calling of breakpoints can be achieved. However, with short read NGS technologies this type of approach is only useful for detecting small-sized SV (Ye et al. 2009; Schröder et al. 2014). New opportunities to overcome these problems using long-read sequencing are described later in this review.

One major bottleneck of the methods described above is the availability of high-quality reference genomes. De novo genome sequence assembly provides the optimal method for fine-scale SV detection, but until now assembly based pangenome approaches have been largely prevented by high cost and time constraints (Hajirasouliha et al. 2010). However, costs can be significantly reduced using reduced-representation sequencing approaches which only address part of the whole genome. Reduced-representation sequencing can be achieved either by selection of restriction fragments for sequencing or by designing baits to capture certain interesting regions of the genome. Whole-exome sequencing is an example of such an approach which reduces computing and sequencing costs by focusing only on protein-coding regions. This reduces the capacity to detect large SV, but can potentially identify causal CNV when sequencing coverage

is sufficient. Exome capture has not yet been used extensively in crops, but recently a capture array was developed for barley to assay species-wide sequence diversity and SV (Mascher et al. 2013). Alternatively, targeted gene sequencing provides opportunities to capture sequence variants for specific panels of target genes, for example for QTL regions (e.g. Clarke et al. 2013) or specific biological pathways (Schiessl et al. 2017a). However, sequence capture does rely on hybridization capture and amplification steps which raise costs of library preparation and can also lead to normalization problems which must be dealt with during data analysis.

Because each method has limitations, a pragmatic approach is to use a combination of SV detection methods (Escaramís et al. 2015; Alkan et al. 2011). However, accurate and unique alignment of short sequence reads to a reference assembly is the foundation of almost every SV detection pipeline. This is extremely challenging in the case of polyploids due to the high homology between their subgenomes. The majority of the crop species reference genomes published to date are themselves based on short read sequencing, containing in some cases thousands of contigs and scaffolds that are not assembled to chromosome level due to the repetitive and complex nature of most crop genomes. The development of third-generation sequencing technologies which generate long-range sequences and enable longer, contiguous scaffolds provide new opportunities for reliable, cost-effective *de novo* assembly at whole-chromosome level (Jiao and Schneeberger 2017). In various research applications, long-read sequencing has become an efficient alternative for SV mapping and phasing. The long-read sequencing platforms from Pacific BioSciences (Menlo Park, CA, USA) and Oxford Nanopore Technologies (Oxford, UK) can provide read lengths ranging from 10 to 150 kb (Schmidt et al. 2017) depending on the DNA library quality. The sequencing error rate for both these platforms is higher compared to short read methods like Illumina sequencing. However, because the sequencing errors are randomly distributed this limitation can be overcome by increasing the depth of sequencing (Schiessl et al. 2018). By spanning rearrangement endpoints and providing more accurate reference assemblies, both of these sequencing platforms enable discovery of complex SV events which were extremely challenging to detect using only short-read methods (English et al. 2015; Chaisson et al. 2014).

Alternative technologies such as optical mapping (Bionano Genomics, San Diego, CA USA) (Lam et al. 2012) or linked-read technologies (10x Genomics, Pleasanton, CA, USA) (Mostovoy et al. 2016), which allow long distances to be effectively spanned in complex genomes, have also contributed substantially to SV detection. Even in challenging polyploid crop genomes, combinations of these different approaches can provide base pair resolution to the study of SV. Unfortunately, these high-resolution techniques are still relatively expensive, meaning that high-throughput analysis in large populations is still prohibited by cost. Until this changes, cheaper high-throughput methods like comparative genomic hybridization (CGH), single nucleotide polymorphism (SNP) arrays or real-time multiplex PCR may be viable alternatives in order to study trait-associated variants in large populations.

Analysing SV using SNP genotyping arrays

High-density SNP arrays provide a popular and cost-effective solution to analyse genetic differences among many individuals within a species. Presently the most popular platform for SNP genotyping is the Infinium™ assay from Illumina (San Diego, CA USA), which relies on hybridization of genomic fragments to probe sequences anchored in flowcells on a glass surface, with probes designed to capture approximately 50 nucleotides of unique, non-polymorphic sequence adjacent to a pre-determined SNP site. This is followed by a single-base extension using hapten-labelled nucleotides and generation of fluorescence signal by adding fluorescently labelled antibodies (Mason et al. 2017). Development of algorithms that can detect SV by quantifying the relative light intensities generated during a SNP call has been an area of research for many years. In human genetics, algorithms like QuantiSNP (Colella et al. 2007) and Birdsuite (Korn et al. 2008) use the fluorescent signal intensity of one allele relative to the other to infer a duplication or deletion event. For polyploid crop plants, the R package "gsrc" (for "genome-wide structural rearrangement calling") was developed to call rearrangements using SNP intensity information (Grandke et al. 2016). A wide range of SNP arrays have been developed in multiple crops for use in crop breeding and genetic research (Voss-Fels and Snowdon 2016).

Despite the widespread use of SNP arrays, there are some inherent problems associated with them when it comes to SV detection. The greatest problem is the limited power of detection of small SVs, due to poor resolution and ascertainment bias due to the pre-determined design of the arrays. PCR-based methods provide a simple and cheap alternative to SNP arrays, especially for detection of SVs at ultra-high resolution; however, prior knowledge of regions of interest is required and throughput is limited. Quantitative real-time PCR (qRT-PCR) and digital PCR (dPCR) are further methods capable of efficiently identifying small SV (InDels) but also translocations, inversions and CNV (Schiessl et al. 2017a; Qian et al. 2016; Ma and Chung 2001). Genes for different traits affected by PAV, CNV and InDel polymorphisms have been validated in a number of major crops using PCR, for example flowering time genes (Schiessl et al. 2017a), lignin biosynthesis genes (Stein et al. 2017)

and a stay-green gene (Qian et al. 2016) in rapeseed, copies of the boron toxicity tolerance gene *Bot1* in barley (Sutton et al. 2007), the aluminium tolerance gene *MATE1* in maize genotypes (Maron et al. 2013) and InDels in the wheat photoperiodicity genes *Ppd-A1a Ppd-B1a* (Nishida et al. 2013).

Crop pangenomes as a future reference paradigm

The unprecedented low cost and high throughput of DNA sequencing today makes it possible to generate genome sequence data for hundreds or thousands of individuals within a species. This provides a new wealth of data to discover genomic re-arrangements in crop genomes in the form of CNVs and PAVs. Insights into genomic SV have conclusively established that a single reference assembly cannot reflect the entire diversity within a species. This gave rise to the concept of pangenomes, which ideally represent all structural genome diversity present in a species. Originally coined for analysis of bacterial genomes (Tettelin et al. 2005), the pangenome concept was first adapted to plants after comparative sequencing of grass genomes revealed widespread structural variation on a previously unknown scale (Morgante et al. 2007). Since the first crop genomes became available, the pangenome concept has been investigated at many different levels especially in maize (Morgante et al. 2007; Springer et al. 2009; Lai et al. 2010; Chia et al. 2012; Hirsch et al. 2014). Most pangenome analyses so far have focused primarily on differences in gene content between individuals in a species; however, as de novo genome assemblies become more feasible in more complex organisms there is growing scope for assembly scale pangenome analysis.

A pangenome for any species is considered to comprise a so-called "core" genome, comprising all genes common to all individuals within the species, along with a "dispensable" genome consisting of partially shared genic regions that are present in some individuals but absent from others (Tettelin et al. 2005). To maximize discovery and coverage of the dispensable genome component, a pangenome should ideally include data from a broad range of individuals representative of all diversity present in the species. Pangenomic description of SV is best achieved by assembly based approaches, but due to their prohibitive cost for large and complex genomes the early pangenomes for most crop species have been generated by genomic resequencing (or in some cases skim sequencing) of representative diversity and analysis by techniques to detect and place SV in existing reference assemblies. This type of approach generally comprises three major steps: resequencing reads are mapped to a high-quality reference assembly, unmapped reads are independently assembled into additional contigs and these are inserted at the appropriate positions in the original assembly using end alignments and/or genetic mapping data. Although faster and cheaper than a de novo assembled pangenome, this method relies strongly on a high-quality reference assembly and can only capture SV in contigs that are able to be successfully assembled and placed from unmapped reads. Nevertheless, such approaches can provide cost-effective opportunities to efficiently capture genic CNV and PAV (Golicz et al. 2016a; Montenegro et al. 2017; Zhou et al. 2017).

To date there are only a handful of studies in which crop pangenomes have been created by de novo assembly of diverse individuals. The most comprehensive study so far was a pangenomic analysis of genomic variation in cultivated and wild rice (Zhao et al. 2018) in which whole-genome de novo assemblies were generated for 66 diverse genotypes chosen to represent a panel of 1529 accessions across the *Oryza sativa–Oryza rufipogon* species complex. The resulting rice pangenome identified 26,372 core genes and 16,208 dispensable genes, enabling associations of SV signatures across the pangenome to domestication sweeps and other signals of natural and artificial selection. Interestingly, several important known genes which were not observed in the original Nipponbare reference genome sequence, including the submergence tolerance1 genes *Sub1A* (Xu et al. 2006) *SNORKEL1* and *SNORKEL2* (Hattori et al. 2009), and the phosphorus-deficiency tolerance gene *Pstol* (Gamuyao et al. 2012) were discovered in the pangenome sequence (Zhao et al. 2018).

These findings reflect observations from Samans et al. (2017) in allotetraploid *B. napus* that genes involved in stress adaptation responses are particularly prevalent among genome structural variants resulting in CNV and PAV, underlining the key role of SV in crop adaption and breeding selection. Similarly, in hexaploid wheat, for which the first high-quality whole-genome reference assembly was recently published (The International Wheat Genome Sequencing Consortium 2014), a resequencing-based pangenome study including 18 wheat cultivars revealed an average of 128,656 genes per cultivar, of which 64% were found to be present in all cultivars and 49,952 genes were dispensable (Appels et al. 2018). Again, annotation of the variable set of genes revealed an enrichment for genes involved in environmental stress and defence response. Assembly based approaches to pangenome analysis will further refine these initial studies as they become more feasible with new assembly strategies and long-read sequencing capabilities. In the foreseeable future, assembly based pangenome analysis is likely to become the method of choice for generating and analysing reference genome data, even in crops with large, complex genomes like those of barley (Stein and Mascher 2019) or wheat. In a pangenome analysis based on *de novo* assemblies for wild relatives of soybean (*Glycine soja*), Li et al. (2014) found high variation in a dispensable genome

comprising around 20% of all assembled sequences, with CNV and mutations in dispensable genes showing evidence for positive selection and a strong influence on important agronomical traits. McHale et al. (2012) found previously that CNV and PAV between wild and domesticated soybean affect over 800 genes involved in biotic stress resistance, and detailed assemblies of wild vs. cultivated forms can deliver important sequence information with regard to potentially important genes that may be absent from reference cultivars.

In contrast to pangenome assembly approaches, which can miss genes not picked up by algorithms for prediction of open reading frames (ORF), some authors advocate the use of transcriptomics-based approaches as a cost-effective way to circumvent this problem. For example, He et al. (2015) introduced the concept of an ordered transcriptome for the allopolyploid *B. napus* based on gene models from its diploid progenitors *B. rapa* and *B. oleracea*, and the homoeologous diploid pan-transcriptomes as a reference to visualize SV in genetically diverse *B. napus* accessions using mRNAseq data (He et al. 2016). Such approaches provide a clear visual impression of the high degree of SV in recent polyploid crop genomes (Fig. 2). Hirsch et al. (2014) took a transcriptomics-based approach to assemble a pangenome for maize. Using this approach, they succeeded in identifying 8681 representative transcript assemblies (longest transcripts within the respective loci) which did not map to the B73 reference, 83% of which mapped only in subsets of 503 investigated lines and can be considered as dispensable genes.

Lu et al. (2015) used an alternative approach for an assembly independent pangenome analysis in maize, using linkage information to map 26 million sequence tags generated by reduced-representation sequencing of 14,129 maize inbred lines. A total of 4.4 million tags with high-confidence map

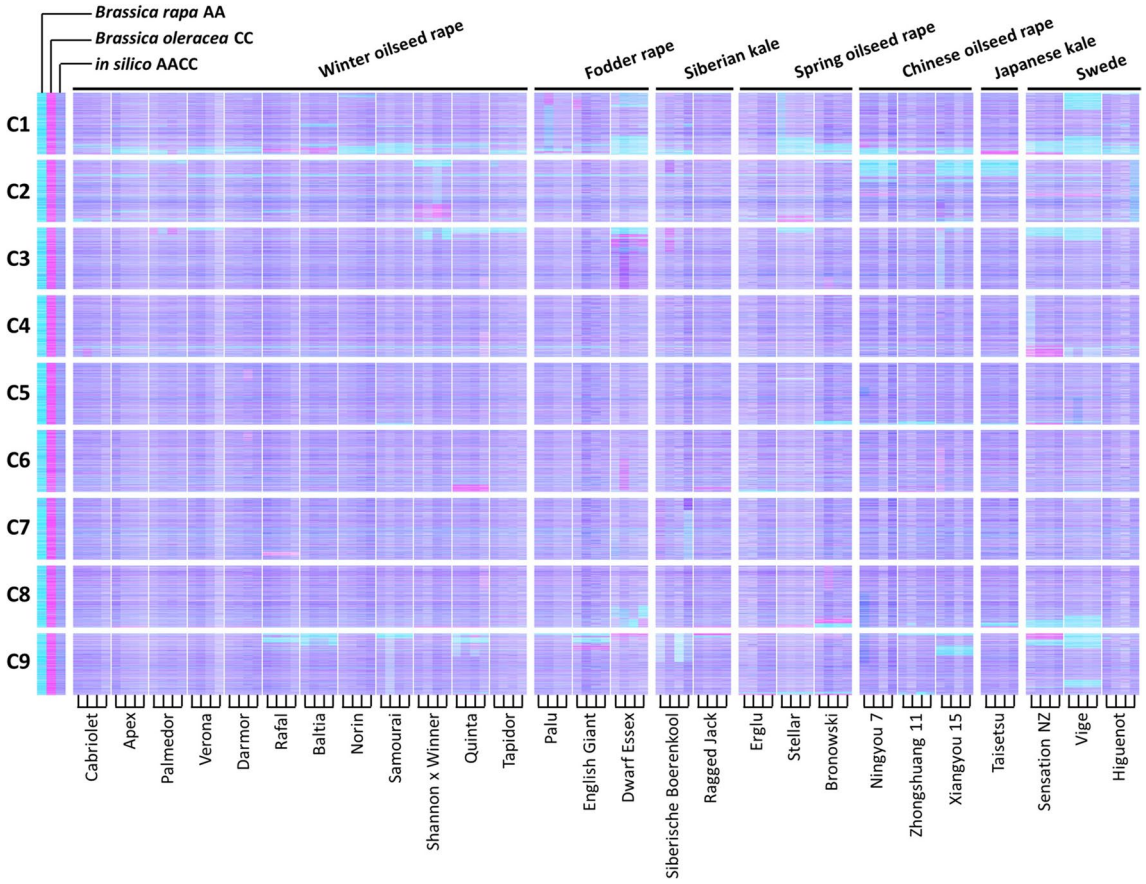

Fig. 2 Visualization of extensive structural variation (SV) caused by homoeologous genome exchanges between the A and C subgenomes of the allopolyploid crop species *Brassica napus* (oilseed rape), using Transcriptome Display Tile Plots derived from mRNAseq reads mapped to an ordered pan-transcriptome. The relative transcript abundance of homoeologous gene pairs is represented in CMYK colour space, with cyan component representing transcript abundance of the A-subgenome copy and magenta component representing transcript abundance of the C-subgenome copy. The pairs are plotted in *Brassica* C genome order (chromosomes denoted C1 to C9) for four biological replicates of each of 27 accessions of *B. napus* and controls comprising parental species and their in silico combination. Image reproduced from He et al. (2015; https://doi.org/10.1111/pbi.12657) under the terms of the Creative Commons Attribution licence 2.0

positions were selected as anchors for a high-density pangenome map. One quarter of these anchors represented PAV and showed enriched associations with phenotypic traits, providing a basis to discover genes where SV is involved in maize adaptation and agronomy. This example shows the power of combinatory approaches involving low-cost, high-throughput sequencing and population genetic analysis to define and analyse SV. Such techniques can potentially also be applied in species without extensive genomics resources.

The Brassicaceae (Cruciferae) family represents an important crop model for studying polyploidy and genomic structural re-arrangements (Mason and Snowdon 2016). Present-day allopolyploid *Brassica* crops originated by interspecific hybridization between different diploid progenitors, for example *B. napus* was formed by hybridization between *B. oleracea* and *B. rapa*. Because the diploids are themselves closely related paleopolyploids with high homoeology between their genomes, synthetic hybrids among them undergo extensive genome restructuring due to inter-homoeologue pairing during the early rounds of meiosis, leading to extensive SV (e.g. Samans et al. 2017; Zou et al. 2018). It might be reasonable to hypothesize that corresponding processes during ancient polyploidization had a similar influence on genome-wide SV and adaptive diversity in *Brassica* spp., giving rise to substantial PAV and CNV observed in present-day diploid cabbage species: such events have been found to have particularly profound effects on genes involved in biotic stress responses in *B. oleracea* (Golicz et al. 2016b) or phenylpropanoid biosynthesis in *B. rapa* (Lin et al. 2014).

Structural variation and trait diversity in major crops: key examples

Table 1 provides examples for demonstrated associations of SV to plant phenotypes in crop species. Wheat is one of the most complex plant genomes due to its large size and polyploid nature. As a result, there has been considerable delay in the detailed genomic analysis of wheat. However, early genetic mapping studies already showed that rearrangements on a number of chromosomes impact numerous important genes for resistance and adaptation traits (Nelson et al. 1995). It was also known for some time that some genes duplicated via polyploidy have remained unaltered, whereas others have been deleted or rendered non-functional by transposon insertions or premature stop codons (Gu et al. 2004). Major translocations in wheat have been associated with specific geographical regions (Riley et al. 1967; Belay and Merker 2004, 2006; Ma et al. 2015) and associated with adaptive and biotic resistance traits (Liu et al. 2016; Law and Worland 2006). With growing access to gene and sequence data, the influence of

Table 1 Examples for structural variations with demonstrated effects on agronomic traits in different crop species

Species	Type of variant	Traits associated	Reference
Barley (*Hordeum vulgare*)	CNV	Boron toxicity tolerance	Sutton et al. (2007)
	CNV	Disease resistance	Muñoz-Amatriaín et al. (2013)
Maize (*Zea mays*)	PAV, CNV	Domestication	Springer et al. (2009)
	CNV	Disease response, heterosis	Beló et al. (2010)
	CNV	–	Swanson-Wagner et al. (2010)
	CNV	Breeding selection	Jiao et al. (2012)
	CNV	Aluminium tolerance	Maron et al. (2013)
Rice (*Oryza sativa*)	PAV, CNV	Grain size, disease resistance	Xu et al. (2012)
	CNV	Disease resistance	Yang et al. (2013), Yu et al. (2013)
	InDel	Root system architecture	Uga et al. (2013)
Soybean (*Glycine max*)	PAV, CNV	Stress responses	Haun et al. (2010), McHale et al. (2012)
	CNV	Disease resistance	Lee et al. (2015)
Sorghum (*Sorghum bicolor*)	PAV, CNV	Disease resistance	Zheng et al. (2011), Mace et al. (2014)
Wheat (*Triticum aestivum*)	CNV	Vernalization, flowering time	Díaz et al. (2012), Würschum et al. (2015)
	CNV	Plant height	Li et al. (2012)
	PAV	Heading date	Nishida et al. (2013)
	CNV	Frost tolerance	Sieber et al. (2016)
	CNV	Winter hardiness	Würschum et al. (2016)
Oilseed rape (*Brassica napus*)	PAV, CNV	Flowering time	Schiessl et al. (2017b)
	HE	Seed fibre	Stein et al. (2017)
	PAV	Stay-green	Qian et al. (2016)
	PAV	Disease resistance	Gabur et al. (2018)

CNV due to polyploidization on adaptive traits like flowering time has been elucidated in more detail. For example, Díaz et al. (2012) found that variation in flowering behaviour in commercial wheat cultivars resulted from CNV for the photoperiodicity gene *Ppd-B1* and the vernalization gene *Vrn-A1*, rather than direct DNA mutations. An increase in the copy number of *Ppd-B1* was found to be associated with an early-flowering, day-length neutral phenotype, whereas plants with a higher *Vrn-A1* copy number exhibited an increased vernalization requirement. In another example, Würschum et al. (2016) reported that copy number variation of *C-repeat Binding Factor* (CBF) genes at the *Fr-A2* locus was the pivotal component for winter hardiness in a panel of 407 European winter wheat cultivars.

In addition to inter-homoeologue chromosome exchanges, interspecific hybrids of wheat with related grasses have led to rich cytogenetic stocks with segmental chromosome insertions or translocations, with a particular focus on resistance traits (Friebe et al. 1996; Wulff and Moscou 2014). One of the most well-known events is the 1BS/1RS translocation from rye, which increased drought adaptation and promoted yield performance of spring wheat in dryland production systems (Villareal et al. 1995; Reeves et al. 1999). However, the exact molecular basis of this improvement is still elusive. On the other hand, gene CNV has also been shown to shape other important phenotypic traits such as plant height in wheat. In cultivars carrying the semi-dwarfing genes *Reduced height* (*Rht*)*-B1b* and *Rht-D1b*, previously uncharacterized CNV polymorphisms of *Rht-D1* were reported to be causal for extreme dwarf phenotype, while a 90 bp insertion in *Rht-B*1 also contributed to severe dwarfism (Pearce et al. 2011). Another critical factor affecting wheat yield is photosynthetic activity determined by chlorophyll content. CNV in the wheat cytokinin oxidase gene *Tackx4*, which influences chlorophyll content and chloroplast stability via modulation of cytokinin concentration, was found to be associated with the chlorophyll content after anthesis as well as grain weight in 102 wheat varieties (Chang et al. 2016).

In barley, several studies have described gene copy number polymorphisms associated with environmental adaptation. As in wheat, CNV in the *H. vulgare CBF* orthologue at the *Frost Resistant-2 locus* (*FR-2*) was found to confer frost tolerance, with an increase in *CBF* coding sequences in winter barley compared to spring forms (Knox et al. 2010; Francia et al. 2016). Similarly, dosage effects from an increase in the number of *H. vulgare* boron transporter (*Bot1*) gene copies were found to confer boron toxicity tolerance (Sutton et al. 2007). Muñoz-Amatriaín et al. (2013) found that CNV between the barley cultivars Barke and Morex was particularly prominent for disease resistance proteins and protein kinases, while increased levels of CNVs were observed for wild accessions in comparison with cultivated barley. As for the examples mentioned above, these studies suggest a key role of SV in conferring the genome plasticity needed for adaption of barley to diverse environmental conditions.

In oilseed rape/canola, anomalies in marker segregation in mapping populations displaying otherwise normal patterns of inheritance (Parkin et al. 1995; Sharpe et al. 1995; Udall et al. 2005) provided the first evidence for exchange of genetic material between homoeologous chromosomes. Detailed elucidation revealed that in the most extreme cases such chromosomal rearrangements can range up to ~ 40 Mb in length, effectively involving whole chromosomes (Higgins et al. 2018). Interestingly, all evidence thus far shows subgenomic bias in direction of exchanges, with loss of the C genome and concomitant gain of the A genome being far more prevalent (Samans et al. 2017; Higgins et al. 2018). Early studies already suggested an important adaptive role, with a well-documented exchange between *B. napus* chromosomes A07 and C06 being associated with higher seed yields (Osborn et al. 2003). The ubiquity of such events, which have shaped the modern *B. napus* genome, was confirmed with the sequencing of the genome reference. Fixed homoeologous exchange events were found to be shared among cultivars due to intentional or inadvertent selection during allopolyploidization and/or breeding, and they thus underlie loci for a number of important traits (Chalhoub et al. 2014). One event led to loss of the C genome copy of a MYB28 transcription factor on chromosome C02 that was replaced with a non-functional A-genome copy of the same gene; this SV defined one of the strongest loci controlling the low glucosinolate phenotype that has underpinned the global success of canola as a major crop. A similar event involving the same *B. napus* chromosomes (A02/C02) created CNV for the flowering time gene *FLC* which leads to accelerated flowering in annual types. Methods to catalogue homoeologous exchanges in *B. napus* from read depth data (Samans et al. 2017) also led to the discovery of PAV underlying a QTL for seed fibre in *B. napus* (Stein et al. 2017).

A comprehensive study of natural genetic variation in homologues of 35 flowering time regulation genes in diverse *B. napus* morphotypes identified an extensive range of structural variation and potential associations to phenotypes related to flowering and secondary processes (Schiessl et al. 2017b). Different homoeologues of the vernalization response gene *Flowering Locus C* (*FLC*), the photosynthetic regulator *Phytochrome A* (*PHYA*) and the hormone *Gibberellic Acid 3-oxidase 1* (*GA3ox1*) all showed CNV and PAV associated with the derivation of *B. napus* morphotypes, again demonstrating the importance of SV on genes involved in human agricultural selection. The use of genome-wide SNP arrays to catalogue SV in multiparental segregating *B. napus* populations enables the inclusion of SV polymorphisms in genome-wide association studies. Gabur et al. (2018) utilized segregating PAV identified by

"single nucleotide absence polymorphism" (SNAP) markers (or "missing" SNP data) to discover a strong involvement of SV in the quantitative control of disease resistance in *B. napus*. Further, in combination with genome sequencing data from mapping parents the QTL could be delineated to small PAV spanning just one or a few potential candidate genes. The success of this study and the continued discovery of SVs as determinants in the control of key agronomic traits suggests that the discovery of SV should become a standard tool in future genetic analyses of crop traits.

In soybean, the world's primary pulse crop, self-pollination together with genetic bottlenecks during domestication have eroded the genetic diversity within the species (Hyten et al. 2006), with sequence polymorphism among soybean accessions typically as low as one SNP per 1000 bases. Therefore, it might also be reasonable to assume a low level of genomic SV. However, this assumption was shattered by Anderson et al. (2014) who found that a panel of 41 soybean accessions contained almost 1528 genes affected by SV. Interestingly, genes exhibiting CNVs were enriched for resistance genes with nucleotide-binding site (NBS) or NBS-leucine-rich repeat (LRR) domains, suggesting involvement of CNVs in interactions with plant pathogens. A well-known example was described by Cook et al. (2012), who found that a 31 kb sequence fragment containing an amino acid transporter, an α-SNAP protein and a WI12 (wound-inducible domain) protein that each contribute to soybean cyst nematode (SCN), one of the most devastating pathogens of soybean, was present in ten tandem copies in resistant cultivars but only a single copy in susceptible cultivars. A similar gene CNV was also reported by Liu et al. (2017) who also showed that CNV of multiple genes present in a single DNA fragment contribute towards SCN resistance.

There have been numerous examples of genome structural variations underlying commercially important traits in many vegetable species. Hardigan et al. (2016) studied genome-wide SV in homozygous clones of diploid potato (*S. tuberosum*), finding that almost 30% of the genes were tolerant to deletion or duplication, with an impact of SV on performance. As in other crop species, there was evidence that PAV and CNV impacted gene clusters in potato involved in environmental stress responses. The authors concluded that CNV may drive adaptation of potato through evolution of important pathways involved in stress responses. SVs have also been reported to play an important role in controlling several traits in tomato. Tranchida-Lombardo et al. (2018) reported over 200 deletions by resequencing and assembling of two tomato landraces. Many of these deletions were found to be localized in the genes annotated for ripening, shelf life and quality of the fruit. In cucumber (*Cucumis sativus*), a model system for sex determination studies in plants, Zhang et al. (2015) constructed a nucleotide-resolution SV map which revealed SVs in their coding regions of over 1600 genes. Using this SV map, they were able to prove that the sex determination in cucumber is controlled by CNV in four genes at the *Female* (*F*) locus.

Approximately 85% of the maize genome is composed of transposable elements (Schnable et al. 2009), which contribute significantly to genomic re-arrangements and gene PAV. In a recent effort to create a newer reference genome assembly for maize, more than 3000 SVs were detected by comparing optical maps for two inbred lines Ki11 and W22 to the B73 reference. The individual SV events ranged from 1 kb to over 1 Mb in length, with an average length of about 20 kb (Jiao et al. 2017). Because this phenomenon has been studied extensively in maize, it is not surprising that a plethora of agronomically interesting traits have been found be to be controlled by PAV in maize, ranging from abiotic and biotic stress responses to plant architecture and heterosis. For example, Wang et al. (2016) reported that an insertion in the promoter region of the *ZmVPP1* gene induces drought-dependent expression of *ZmVPP1* in drought-tolerant genotypes. The PAV in *ZmVPP1* was also associated with enhancement of photosynthetic efficiency and root development under both stress and non-stress conditions. CNV events also play a role in stress resistance responses in maize. For example, the *rp1* locus, responsible for race-specific resistance to the common rust fungus, is a hotspot for unequal crossovers leading to gain, loss or duplications in this NBS-LRR gene family. This creates a diverse haplotype makeup at the *rp1* locus, translating to variable resistance responses to various rust races (Richter et al. 1995). The same authors found a similar example for a CNV in a wall-associated kinase (*ZmWAK*) gene, which was found to confer resistance to head smut in maize. Interestingly, the responsible WAK gene was absent in many modern maize lines but present in wild relatives.

Although there is considerable indirect evidence for a role of PAV in stress responses, there is still only a handful of examples for their importance in adaptive traits in maize. Maron et al. (2013) reported association of a rare CNV in the multidrug and toxin exudation gene *MATE1* in maize to aluminium tolerance. Strikingly, the geographical origin of maize lines carrying three copies of *MATE1* coincided with highly acidic soil, implicating this CNV as an important SV conferring maize adaptation to a specific environment.

A DNA segment present or absent exclusively in germplasm adapted to a particular type of environmental cue might be indicative of the fact that genic PAV in this segment play a crucial role in adaptation. An array of INDELs in the 5′ regulatory region of the *FLOWERING LOCUS T* homologue, *LanFTc1* was found to be associated with differential vernalization response, flowering time, in narrow-leafed lupin (Taylor et al. 2018). Genotypes with no deletion exhibited an early flowering behaviour and a reduced or no response to vernalization. Such a catalogue

of structural variations could serve as the basis for the necessary environmental plasticity needed for designing the future crops adapted to wide range of environments. Darracq et al. (2018) identified in total 88 Mb of DNA in a French maize inbred line that was absent in an American inbred line, and contained 395 putative coding genes. Evidence was also observed for de novo SV in European maize alongside ancient SV, demonstrating ongoing adaptive genome evolutionary dynamics. Annotation of the novel genes revealed putative roles in biotic and biotic stress responses, in biosynthetic processes, in development, in protein synthesis and in chromatin remodelling. Intriguingly, expression of most of the novel genes was restricted to particular conditions or tissues, leading to a conclusion that at least some of the genes from the dispensable part of the genome might be involved in environmental adaptation.

The realization that heterotic pools in maize breeding programs can have vastly different gene content gave new insight into the impact of SV on heterosis. Springer et al. (2009) compared the genome structures of two maize inbred lines by comparative genomic hybridization and found that a copious amount of genomic sequences exhibited copy number differences between the two genomes. Sun et al. (2018) underlined the extent of genome-wide SV in maize by assembling the genome of Mo17 and comparing it to the B73 reference assembly. This confirmed that almost 10% of the annotated genes were exclusive to one or the other accession, while more than 20% were found to show substantial structural variation. It was also hypothesized that these SVs might be involved in heterosis and genome evolution. Furthermore, many sequences annotated as single-copy genes were found to be present in one genome but completely absent from the other genome. Although the contributing mechanisms for heterosis are still not completely elucidated and may differ from crop to crop, there is good reason to believe that fixation of complementary PAV in different heterotic pools can play an important role in exploitation of additive heterosis in hybrid breeding.

The huge diversity in rice, combined with well-defined phylogeny available for the genus *Oryza*, the small size of the genome and vast genome sequence resources make it an ideal candidate for studying effects of genomic SV on traits. Bai et al. (2016) generated a CNV map, at single nucleotide resolution for 50 rice accessions, comprising 9196 deletions compared to the Nipponbare reference genome. More than 2000 annotated genes were reported to be affected by CNV. Uga et al. (2013) identified the gene *Deep Rooting 1* (*DRO1*) as a key regulator of root system architecture with a profound effect on yield under different water regimes in rice. A japonica upland rice (Kinandang Patong) containing a full-length *DRO1* copy was found to have a deeper root system architecture, whereas the indica lowland rice cultivar IR64 carried a truncated copy due to a 1 bp InDel. This results in shallower roots due to the introduction of a premature stop codon in *DRO1* (Uga et al. 2013). Yu et al. (2011) and Yao et al. (2015) both found enrichment for disease resistance or defence response genes among dispensable genes in large rice resequencing studies. Wang et al. (2015) reported CNV at the *Grain Length on Chromosome 7* (GL7) locus associated with regulation of grain dimensions in rice. A tandem duplication of a DNA fragment within the *GL7* locus lead was found to cause upregulation of *GL7* expression and suppression of its negative regulator, thereby resulting in an increased grain length and improved grain appearance.

Outlook

As more and more genome data become available for major crops, our insight into the profound importance of SV on trait diversity continues to grow. Understanding the contribution of gene copy number and presence–absence variation to important traits will be an important factor in improving the accuracy and efficacy of many new genetic technologies in plant breeding, from genomic selection to genome editing strategies. Just a decade ago, the notion that full, high-quality reference genome assemblies for any major crop could be generated reasonably simply and quickly, at low cost, was barely conceivable. Today this is (almost) a reality, and a new era of high-throughput pangenomic analyses is set to dominate crop genetic studies in the immediate future. Although current costs of third-generation DNA sequencing technologies and chromatin conformation technologies for scaffold improvement are still high, and computational bottlenecks associated with creation of reference assemblies need to be overcome, the plummeting cost of long-read sequencing and improvement in computational algorithms and hardware could make de novo genome assembly more routine in the foreseeable future, even for complex polyploid crop genomes. One aspect of interest for breeders in a changing world is associations of SV with ecogeographical adaptations, abiotic stress adaptation and biotic stress responses. This knowledge opens fascinating new opportunities to learn from adaptive evolution of polyploid crop species in order to improve crop resilience against biotic and abiotic stress constraints in the face of climate change. From a broader perspective, studies of SV in model and crop plants derived from recent polyploidy have revealed an involvement of gene dosage and/or PAV in a wide number of different traits under natural and human (breeding) selection and showed how genome rearrangements resulting from de novo polyploidization might even be used to generate new variation for breeding. Such examples underline the role of SV as a key driver of genetic diversity for future breeding of sustainable, resilient and healthy crops. Novel methods to detect,

assay, harness and select for useful SV events will therefore be a valuable future resource for crop breeding.

Author contribution statement IG and HSC drafted the manuscript. IAP and RJS contributed text sections and edited the content.

Funding Deutscher Akademischer Austauschdienst and Deutsche Forschungsgemeinschaft (Grant No. SN14/17-1).

Compliance with ethical standards

Conflict of interest All authors jointly state that there is no conflict of interest. IAPP and RJS serve on the editorial board for this journal, but this is not considered to constitute a conflict of interest.

References

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Ann Bot 120(2):183–194. https://doi.org/10.1093/aob/mcx079

Ali N, Heslop-Harrison JP, Ahmad H, Graybosch RA, Hein GL, Schwarzacher T (2016) Introgression of chromosome segments from multiple alien species in wheat breeding lines with wheat streak mosaic virus resistance. Heredity 117:114–123. https://doi.org/10.1038/hdy.2016.36

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 41(10):1061–1067. https://doi.org/10.1038/ng.437

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12(5):363–376. https://doi.org/10.1038/nrg2958

Anderson JE, Kantar MB, Kono TY, Fu F, Stec AO, Song Q, Cregan PB, Specht JE, Diers BW, Cannon SB, McHale LK, Stupar RM (2014) A roadmap for functional structural variants in the soybean genome. G3 (Bethesda) 4(7):1307–1318. https://doi.org/10.1534/g3.114.011551

Appels R, Eversole K, Feuillet C, Keller B, Rogers J, Stein N, Pozniak CJ, Choulet F, Distelfeld A, Poland J, Ronen G, Sharpe AG, Pozniak C, Barad O, Baruch K, Keeble-Gagnère G, Mascher M, Ben-Zvi G, Josselin A-A, Himmelbach A, Balfourier F, Gutierrez-Gonzalez J, Hayden M, Koh C, Muehlbauer G, Pasam RK, Paux E, Rigault P, Tibbits J, Tiwari V, Spannagl M, Lang D, Gundlach H, Haberer G, Mayer KFX, Ormanbekova D, Prade V, Šimková H, Wicker T, Swarbreck D, Rimbert H, Felder M, Guilhot N, Kaithakottil G, Keilwagen J, Leroy P, Lux T, Twardziok S, Venturini L, Juhász A, Abrouk M, Fischer I, Uauy C, Borrill P, Ramirez-Gonzalez RH, Arnaud D, Chalabi S, Chalhoub B, Cory A, Datla R, Davey MW, Jacobs J, Robinson SJ, Steuernagel B, van Ex F, Wulff BBH, Benhamed M, Bendahmane A, Concia L, Latrasse D, Alaux M, Bartoš J, Bellec A, Berges H, Doležel J, Frenkel Z, Gill B, Korol A, Letellier T, Olsen O-A, Singh K, Valárik M, van der Vossen E, Vautrin S, Weining S, Fahima T, Glikson V, Raats D, Číhalíková J, Toegelová H, Vrána J, Sourdille P, Darrier B, Barabaschi D, Cattivelli L, Hernandez P, Galvez S, Budak H, Jones JDG, Witek K, Yu G, Small I, Melonek J, Zhou R, Belova T, Kanyuka K, King R, Nilsen K, Walkowiak S, Cuthbert R, Knox R, Wiebe K, Xiang D, Rohde A, Golds T, Čížková J, Akpinar BA, Biyiklioglu S, Gao L, N'Daiye A, Kubaláková M, Šafář J, Alfama F, Adam-Blondon A-F, Flores R, Guerche C, Loaec M, Quesneville H, Condie J, Ens J, Maclachlan R, Tan Y, Alberti A, Aury J-M, Barbe V, Couloux A, Cruaud C, Labadie K, Mangenot S, Wincker P, Kaur G, Luo M, Sehgal S, Chhuneja P, Gupta OP, Jindal S, Kaur P, Malik P, Sharma P, Yadav B, Singh NK, Khurana J, Chaudhary C, Khurana P, Kumar V, Mahato A, Mathur S, Sevanthi A, Sharma N, Tomar RS, Holušová K, Plíhal O, Clark MD, Heavens D, Kettleborough G, Wright J, Balcárková B, Hu Y, Salina E, Ravin N, Skryabin K, Beletsky A, Kadnikov V, Mardanov A, Nesterov M, Rakitin A, Sergeeva E, Handa H, Kanamori H, Katagiri S, Kobayashi F, Nasuda S, Tanaka T, Wu J, Cattonaro F, Jiumeng M, Kugler K, Pfeifer M, Sandve S, Xun X, Zhan B, Batley J, Bayer PE, Edwards D, Hayashi S, Tulpová Z, Visendi P, Cui L, Du X, Feng K, Nie X, Tong W, Wang L (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science. https://doi.org/10.1126/science.aar7191

Bai Z, Chen J, Liao Y, Wang M, Liu R, Ge S, Wing RA, Chen M (2016) The impact and origin of copy number variations in the *Oryza* species. BMC Genomics 17:261. https://doi.org/10.1186/s12864-016-2589-2

Belay G, Merker A (2004) Cytogenetic studies in Ethiopian landraces of tetraploid wheat (*Tviticum Turgidum* L.). II. Spontaneous chromosome translocations and fertility. Hereditas 126(1):35–43. https://doi.org/10.1111/j.1601-5223.1997.00035.x

Belay G, Merker A (2006) Cytogenetic analysis of a spontaneous 5B/6B translocation in tetraploid wheat landraces from Ethiopia, and implications for breeding. Plant Breed 117(6):537–542. https://doi.org/10.1111/j.1439-0523.1998.tb02203.x

Beló A, Beatty MK, Hondred D, Fengler KA, Li B, Rafalski A (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120(2):355–367. https://doi.org/10.1007/s00122-009-1128-9

Bridges CB (1936) The BAR "gene" a duplication. Science 83(2148):210–211. https://doi.org/10.1126/science.83.2148.210

Chaisson MJP, Huddleston J, Dennis MY, Sudmant PH, Malig M, Hormozdiari F, Antonacci F, Surti U, Sandstrom R, Boitano M, Landolin JM, Stamatoyannopoulos JA, Hunkapiller MW, Korlach J, Eichler EE (2014) Resolving the complexity of the human genome using single-molecule sequencing. Nature 517(7536):608–611. https://doi.org/10.1038/nature13907

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, Correa M, Da Silva C, Just J, Falentin C, Koh CS, Le Clainche I, Bernard M, Bento P, Noel B, Labadie K, Alberti A, Charles M, Arnaud D, Guo H, Daviaud C, Alamery S, Jabbari K, Zhao M, Edger PP, Chelaifa H, Tack D, Lassalle G, Mestiri I, Schnel N, Le Paslier M-C, Fan G, Renault V, Bayer PE, Golicz AA, Manoli S, Lee T-H, Thi VHD, Chalabi S, Hu Q, Fan C, Tollenaere R, Lu Y, Battail C, Shen J, Sidebottom CHD, Canaguier A, Chauveau A, Berard A, Deniot G, Guan M, Liu Z, Sun F, Lim YP, Lyons E, Town CD, Bancroft I, Meng J, Ma J, Pires JC, King GJ, Brunel D, Delourme R, Renard M, Aury J-M, Adams KL, Batley J, Snowdon RJ, Tost J, Edwards D, Zhou Y, Hua W, Sharpe AG, Paterson AH, Guan C, Wincker P (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345(6199):950–953. https://doi.org/10.1126/science.1253435

Chang C, Lu J, Zhang H-P, Ma C-X, Sun G (2016) Copy number variation of cytokinin oxidase gene *Tackx4* associated with grain weight and chlorophyll content of flag leaf in common wheat. PLoS ONE 10(12):e0145970. https://doi.org/10.1371/journal.pone.0145970

Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares IL, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser B, Vingron M, Ropers HH (2008) Mapping translocation breakpoints by next-generation sequencing. Genome Res 18(7):1143–1149. https://doi.org/10.1101/gr.076166.108

Chester M, Gallagher JP, Symonds VV, Cruz da Silva AV, Mavrodiev EV, Leitch AR, Soltis PS, Soltis DE (2012) Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). Proc Natl Acad Sci USA 109(4):1176–1181. https://doi.org/10.1073/pnas.1112041109

Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, Gore M, Guill KE, Holland J, Hufford MB, Lai J, Li M, Liu X, Lu Y, McCombie R, Nelson R, Poland J, Prasanna BM, Pyhäjärvi T, Rong T, Sekhon RS, Sun Q, Tenaillon MI, Tian F, Wang J, Xu X, Zhang Z, Kaeppler SM, Ross-Ibarra J, McMullen MD, Buckler ES, Zhang G, Xu Y, Ware D (2012) Maize HapMap2 identifies extant variation from a genome in flux. Nat Genet 44:803. https://doi.org/10.1038/ng.2313

Clarke WE, Parkin IA, Gajardo HA, Gerhardt DJ, Higgins E, Sidebottom C, Sharpe AG, Snowdon RJ, Federico ML, Iniguez-Luy FL (2013) Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. PLoS ONE 8(12):e81992. https://doi.org/10.1371/journal.pone.0081992

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, Bassett AS, Seller A, Holmes CC, Ragoussis J (2007) QuantiSNP: an objective Bayes hidden-Markov model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 35(6):2013–2025. https://doi.org/10.1093/nar/gkm076

Cook DE, Lee TG, Guo X, Melito S, Wang K, Bayless AM, Wang J, Hughes TJ, Willis DK, Clemente TE, Diers BW, Jiang J, Hudson ME, Bent AF (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. Science 338(6111):1206–1209. https://doi.org/10.1126/science.1228746

Darracq A, Vitte C, Nicolas S, Duarte J, Pichon J-P, Mary-Huard T, Chevalier C, Bérard A, Le Paslier M-C, Rogowsky P, Charcosset A, Joets J (2018) Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. BMC Genomics 19(1):119. https://doi.org/10.1186/s12864-018-4490-7

Díaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the *Photoperiod-B1* and *Vernalization-A1* genes Is associated with altered flowering time in wheat (*Triticum aestivum*). PLoS ONE 7(3):e33234. https://doi.org/10.1371/journal.pone.0033234

Edwards D, Batley J, Snowdon RJ (2013) Accessing complex crop genomes with next-generation sequencing. Theor Appl Genet 126(1):1–11. https://doi.org/10.1007/s00122-012-1964-x

English AC, Salerno WJ, Hampton OA, Gonzaga-Jauregui C, Ambreth S, Ritter DI, Beck CR, Davis CF, Dahdouli M, Ma S, Carroll A, Veeraraghavan N, Bruestle J, Drees B, Hastie A, Lam ET, White S, Mishra P, Wang M, Han Y, Zhang F, Stankiewicz P, Wheeler DA, Reid JG, Muzny DM, Rogers J, Sabo A, Worley KC, Lupski JR, Boerwinkle E, Gibbs RA (2015) Assessing structural variation in a personal genome—towards a human reference diploid genome. BMC Genomics 16(1):332. https://doi.org/10.1186/s12864-015-1479-3

Escaramís G, Docampo E, Rabionet R (2015) A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics 14(5):305–314. https://doi.org/10.1093/bfgp/elv014

Esteve-Codina A, Paudel Y, Ferretti L, Raineri E, Megens H-J, Silió L, Rodríguez MC, Am Groenen M, Ramos-Onsins SE, Pérez-Enciso M (2013) Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genomics 14(1):148. https://doi.org/10.1186/1471-2164-14-148

Fadista J, Thomsen B, Holm L-E, Bendixen C (2010) Copy number variation in the bovine genome. BMC Genomics 11(1):284. https://doi.org/10.1186/1471-2164-11-284

Francia E, Morcia C, Pasquariello M, Mazzamurro V, Milc JA, Rizza F, Terzi V, Pecchioni N (2016) Copy number variation at the *HvCBF4–HvCBF2* genomic segment is a major component of frost resistance in barley. Plant Mol Biol 92(1):161–175. https://doi.org/10.1007/s11103-016-0505-4

Friebe B, Jiang J, Raupp WJ, McIntosh RA, Gill BS (1996) Characterization of wheat-alien translocations conferring resistance to diseases and pests: current status. Euphytica 91(1):59–87. https://doi.org/10.1007/BF00035277

Gabur I, Chawla HS, Liu X, Kumar V, Faure S, von Tiedemann A, Jestin C, Dryzska E, Volkmann S, Breuer F, Delourme R, Snowdon R, Obermeier C (2018) Finding invisible quantitative trait loci with missing data. Plant Biotechnol J. https://doi.org/10.1111/pbi.12942

Gaeta RT, Chris Pires J (2010) Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytol 186(1):18–28. https://doi.org/10.1111/j.1469-8137.2009.03089.x

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95(5):1971. https://doi.org/10.1073/pnas.95.5.1971

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S (2012) The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. Nature 488(7412):535–539. https://doi.org/10.1038/nature11346

Ghosh S, Qu Z, Das PJ, Fang E, Juras R, Cothran EG, McDonell S, Kenney DG, Lear TL, Adelson DL, Chowdhary BP, Raudsepp T (2014) Copy number variation in the horse genome. PLoS Genet 10(10):e1004712. https://doi.org/10.1371/journal.pgen.1004712

Golicz AA, Batley J, Edwards D (2016a) Towards plant pangenomics. Plant Biotechnol J 14(4):1099–1105. https://doi.org/10.1111/pbi.12499

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, Martinez PA, Chan CKK, Severn-Ellis A, McCombie WR, Parkin IAP, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D (2016b) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nat Commun 7:13390. https://doi.org/10.1038/ncomms13390

Gonzalez E (2005) The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307(5714):1434–1440. https://doi.org/10.1126/science.1101160

Grandke F, Snowdon R, Samans B (2016) gsrc: an R package for genome structure rearrangement calling. Bioinformatics 3:545–546. https://doi.org/10.1093/bioinformatics/btw648

Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four *Triticeae* Genomes. Plant Physiol 135(1):459. https://doi.org/10.1104/pp.103.038083

Hajirasouliha I, Hormozdiari F, Alkan C, Kidd JM, Birol I, Eichler EE, Sahinalp SC (2010) Detection and characterization of novel sequence insertions using paired-end next-generation sequencing. Bioinformatics 26(10):1277–1283. https://doi.org/10.1093/bioinformatics/btq152

Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Yang X, Zeng Z, Douches DS, Jiang J, Veilleux RE, Buell CR (2016) Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum tuberosum*. Plant Cell 28(2):388–405. https://doi.org/10.1105/tpc.15.00538

Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10(8):551–564. https://doi.org/10.1038/nrg2593

Hattori Y, Nagai K, Furukawa S, Song X-J, Kawano R, Sakakibara H, Wu J, Matsumoto T, Yoshimura A, Kitano H, Matsuoka M, Mori H, Ashikari M (2009) The ethylene response factors *SNORKEL1* and *SNORKEL2* allow rice to adapt to deep water. Nature 460(7258):1026–1030. https://doi.org/10.1038/nature08258

Haun WJ, Hyten DL, Xu WW, Gerhardt DJ, Albert TJ, Richmond T, Jeddeloh JA, Jia G, Springer NM, Vance CP, Stupar RM (2010) The composition and origins of genomic variation among individuals of the soybean reference cultivar Williams 82. Plant Physiol 155(2):645–655. https://doi.org/10.1104/pp.110.166736

He Z, Cheng F, Li Y, Wang X, Parkin IAP, Chalhoub B, Liu S, Bancroft I (2015) Construction of *Brassica* A and C genome-based ordered pan-transcriptomes for use in rapeseed genomic research. Data Brief 4:357–362. https://doi.org/10.1016/j.dib.2015.06.016

He Z, Wang L, Harper AL, Havlickova L, Pradhan AK, Parkin IAP, Bancroft I (2016) Extensive homoeologous genome exchanges in allopolyploid crops revealed by mRNAseq-based visualization. Plant Biotech J 15:594–604. https://doi.org/10.1111/pbi.12657

Higgins EE, Clarke WE, Howell EC, Armstrong SJ, Parkin IAP (2018) Detecting *de Novo* homoeologous recombination events in cultivated *Brassica napus* using a genome-wide SNP array. G3 (Bethesda) 8(8):2673–2683. https://doi.org/10.1534/g3.118.200118

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Penagaricano F, Lindquist E, Pedraza MA, Barry K, de Leon N, Kaeppler SM, Buell CR (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell 26(1):121–135. https://doi.org/10.1105/tpc.113.119982

Horn R, Snowdon R, Kusterer B (2002) Structural genome analysis using molecular cytogenetic techniques. In: Esser K, Lüttge U, Beyschlag W, Hellwig F (eds) Progress in botany: genetics, physiology, ecology. Springer, Berlin, pp 55–79

Hurgobin B, Golicz AA, Bayer PE, Chan C-KK, Tirnaz S, Dolatabadian A, Schiessl SV, Samans B, Montenegro JD, Parkin IAP, Pires JC, Chalhoub B, King GJ, Snowdon R, Batley J, Edwards D (2017) Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. Plant Biotechnol J 16(7):1265–1274. https://doi.org/10.1111/pbi.12867

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA 103(45):16666. https://doi.org/10.1073/pnas.0604379103

Iovene M, Zhang T, Lou Q, Buell CR, Jiang J (2013) Copy number variation in potato—an asexually propagated autotetraploid species. Plant J 75(1):80–89. https://doi.org/10.1111/tpj.12200

Jiao W-B, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. Curr Opin Plant Biol 36:64–70. https://doi.org/10.1016/j.pbi.2017.02.002

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, Zhang M, Xie S, Lai J (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44(7):812–815. https://doi.org/10.1038/ng.2312

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524. https://doi.org/10.1038/nature22971

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258(5083):818. https://doi.org/10.1126/science.1359641

Keane TM, Wong K, Adams DJ, Flint J, Reymond A, Yalcin B (2014) Identification of structural variation in mouse genomes. Front Genet 5:1061. https://doi.org/10.3389/fgene.2014.00192

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N, Stockinger EJ (2010) CBF gene copy number variation at *Frost Resistance-2* is associated with levels of freezing tolerance in temperate-climate cereals. Theor Appl Genet 121(1):21–35. https://doi.org/10.1007/s00122-010-1288-7

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M (2007) Paired-end mapping reveals extensive structural variation in the human genome. Science 318(5849):420–426. https://doi.org/10.1126/science.1149504

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D (2008) Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 40(10):1253–1260. https://doi.org/10.1038/ng.237

Lagercrantz U, Putterill J, Coupland G, Lydiate D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9(1):13–20

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, Jiao Y, Ni P, Zhang J, Li D, Guo X, Ye K, Jian M, Wang B, Zheng H, Liang H, Zhang X, Wang S, Chen S, Li J, Fu Y, Springer NM, Yang H, Wang J, Dai J, Schnable PS, Wang J (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet 42(11):1027–1030. https://doi.org/10.1038/ng.684

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok P-Y (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30(8):771–776. https://doi.org/10.1038/nbt.2303

Law CN, Worland AJ (2006) The control of adult-plant resistance to yellow rust by the translocated chromosome 5BS–7BS of bread wheat. Plant Breed 116(1):59–63. https://doi.org/10.1111/j.1439-0523.1997.tb00975.x

Lee JA, Carvalho CMB, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131(7):1235–1247. https://doi.org/10.1016/j.cell.2007.11.037

Lee TG, Kumar I, Diers BW, Hudson ME (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. Mol Ecol 24(8):1774–1791. https://doi.org/10.1111/mec.13138

Li W, Olivier M (2013) Current analysis platforms and methods for detecting copy number variation. Physiol Genom 45(1):1–16. https://doi.org/10.1152/physiolgenomics.00082.2012

Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, Zhang X, Guo X, Gu Y, Zhang L, Jia J, Kong X (2012) A tandem segmental duplication (TSD) in green revolution gene *Rht-D1b* region underlies plant height variation. New Phytol 196(1):282–291. https://doi.org/10.1111/j.1469-8137.2012.04243.x

Li Y, Zhou G, Ma J, Jiang W, Jin L, Zhang Z, Guo Y, Zhang J, Sui Y, Zheng L, Zhang S, Zuo Q, Shi X, Li Y, Zhang W, Hu Y, Kong G, H-l H, Tan B, Song J, Liu Z, Wang Y, Ruan H, Yeung CKL, Liu J, Wang H, Zhang L, Guan R, Wang K, Li W, Chen S, Chang R, Jiang Z, Jackson SA, Li R, Qiu L (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and

agronomic traits. Nat Biotechnol 32(10):1045–1052. https://doi.org/10.1038/nbt.2979

Lin K, Zhang N, Severing EI, Nijveen H, Cheng F, Visser RGF, Wang X, de Ridder D, Bonnema G (2014) Beyond genomic variation—comparison and functional annotation of three *Brassica* rapagenomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics 15(1):250. https://doi.org/10.1186/1471-2164-15-250

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50 K SNP BeadChip array. BMC Genomics 14(1):229. https://doi.org/10.1186/1471-2164-14-229

Liu M, Stiller J, Holušová K, Vrána J, Liu D, Doležel J, Liu C (2016) Chromosome-specific sequencing reveals an extensive dispensable genome component in wheat. Sci Rep. https://doi.org/10.1038/srep36398

Liu S, Kandoth PK, Lakhssassi N, Kang J, Colantonio V, Heinz R, Yeckel G, Zhou Z, Bekal S, Dapprich J, Rotter B, Cianzio S, Mitchum MG, Meksem K (2017) The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. Nat Commun 8:14822. https://doi.org/10.1038/ncomms14822

Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, Li Y, Li Y, Semagn K, Zhang X, Hernandez AG, Mikel MA, Soifer I, Barad O, Buckler ES (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. Nat Commun. https://doi.org/10.1038/ncomms7914

Lupski JR (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14(10):417–422. https://doi.org/10.1016/S0168-9525(98)01555-8

Ma L, Chung WK (2001) Quantitative analysis of copy number variants based on real-time lightcycler PCR. In: Haines JL, Korf BR, Morton CC, Seidman CE, Seidman JG, Smith DR (eds) Current protocols in human genetics, vol 107. Wiley, Hoboken, pp 7.21.1–7.21.8

Ma J, Stiller J, Zheng Z, Wei Y, Zheng Y-L, Yan G, Doležel J, Liu C (2015) Putative interchromosomal rearrangements in the hexaploid wheat (*Triticum aestivum* L.) genotype 'Chinese Spring' revealed by gene locations on homoeologous chromosomes. BMC Evol Biol 15:37. https://doi.org/10.1186/s12862-015-0313-5

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR, Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. Nat Commun 4:2320. https://doi.org/10.1038/ncomms3320

Mace E, Tai S, Innes D, Godwin I, Hu W, Campbell B, Gilding E, Cruickshank A, Prentis P, Wang J, Jordan D (2014) The plasticity of NBS resistance genes in sorghum is driven by multiple evolutionary processes. BMC Plant Biol. https://doi.org/10.1186/s12870-014-0253-z

Mamtani M, Anaya J-M, He W, Ahuja SK (2010) Association of copy number variation in the *FCGR3B* gene with risk of autoimmune diseases. Genes Immun 11(2):155–160. https://doi.org/10.1038/gene.2009.71

Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Pineros MA, Schatz MC, Wing RA, Kochian LV (2013) Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. Proc Natl Acad Sci USA 110(13):5241–5246. https://doi.org/10.1073/pnas.1220766110

Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, Akhunov ED, Hedley PE, Gonzales AM, Morrell PL, Kilian B, Blattner FR, Scholz U, Mayer KFX, Flavell AJ, Muehlbauer GJ, Waugh R, Jeddeloh JA, Stein N (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. Plant J 76(3):494–505. https://doi.org/10.1111/tpj.12294

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J, Bayer M, Ramsay L, Liu H, Haberer G, Zhang X-Q, Zhang Q, Barrero RA, Li L, Taudien S, Groth M, Felder M, Hastie A, Šimková H, Staňková H, Vrána J, Chan S, Muñoz-Amatriaín M, Ounit R, Wanamaker S, Bolser D, Colmsee C, Schmutzer T, Aliyeva-Schnorr L, Grasso S, Tanskanen J, Chailyan A, Sampath D, Heavens D, Clissold L, Cao S, Chapman B, Dai F, Han Y, Li H, Li X, Lin C, McCooke JK, Tan C, Wang P, Wang S, Yin S, Zhou G, Poland JA, Bellgard MI, Borisjuk L, Houben A, Doležel J, Ayling S, Lonardi S, Kersey P, Langridge P, Muehlbauer GJ, Clark MD, Caccamo M, Schulman AH, Mayer KFX, Platzer M, Close TJ, Scholz U, Hansson M, Zhang G, Braumann I, Spannagl M, Li C, Waugh R, Stein N (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544(7651):427–433. https://doi.org/10.1038/nature22043

Mason AS, Snowdon RJ (2016) Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. Plant Biol (Stuttg) 18(6):883–892. https://doi.org/10.1111/plb.12462

Mason AS, Higgins EE, Snowdon RJ, Batley J, Stein A, Werner C, Parkin IAP (2017) A user guide to the Brassica 60 K Illumina Infinium™ SNP genotyping array. Theor Appl Genet 130(4):621–633. https://doi.org/10.1007/s00122-016-2849-1

McHale LK, Haun WJ, Xu WW, Bhaskar PB, Anderson JE, Hyten DL, Gerhardt DJ, Jeddeloh JA, Stupar RM (2012) Structural variants in the soybean genome localize to clusters of biotic stress-response genes. Plant Physiol 159(4):1295–1308. https://doi.org/10.1104/pp.112.194605

McVey M, Lee SE (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet 24(11):529–538. https://doi.org/10.1016/j.tig.2008.08.007

Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, Chan C-KK, Visendi P, Lai K, Doležel J, Batley J, Edwards D (2017) The pangenome of hexaploid bread wheat. Plant J 90(5):1007–1013. https://doi.org/10.1111/tpj.13515

Moore JK, Haber JE (1996) Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. Mol Cell Biol 16(5):2164–2173. https://doi.org/10.1128/MCB.16.5.2164

Morgante M, de Paoli E, Radovic S (2007) Transposable elements and the plant pan-genomes. Curr Opin Plant Biol 10(2):149–155. https://doi.org/10.1016/j.pbi.2007.02.001

Mostovoy Y, Levy-Sakin M, Lam J, Lam ET, Hastie AR, Marks P, Lee J, Chu C, Lin C, Džakula Ž, Cao H, Schlebusch SA, Giorda K, Schnall-Levin M, Wall JD, Kwok P-Y (2016) A hybrid approach for *de novo* human genome sequence assembly and phasing. Nat Methods 13(7):587–590. https://doi.org/10.1038/nmeth.3865

Muñoz-Amatriaín M, Eichten SR, Wicker T, Richmond TA, Mascher M, Steuernagel B, Scholz U, Ariyadasa R, Spannagl M, Nussbaumer T, Mayer KFX, Taudien S, Platzer M, Jeddeloh JA, Springer NM, Muehlbauer GJ, Stein N (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. Genome Biol 14(6):203. https://doi.org/10.1186/gb-2013-14-6-r58

Neik TX, Barbetti MJ, Batley J (2017) Current status and challenges in identifying disease resistance genes in *Brassica napus*. Front Plant Sci 8:1788. https://doi.org/10.3389/fpls.2017.01788

Nelson JC, Sorrells ME, Van-Deynze AE, Lu YH, Atkinson M, Bernard M, Leroy P, Faris JD, Anderson JA (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5, and 7. Genetics 141(2):721–731

Nishida H, Yoshida T, Kawakami K, Fujita M, Long B, Akashi Y, Laurie DA, Kato K (2013) Structural variation in the 5′ upstream region of photoperiod-insensitive alleles *Ppd-A1a* and *Ppd-B1a* identified in hexaploid wheat (*Triticum aestivum* L.), and their effect on heading time. Mol Breeding 31(1):27–37. https://doi.org/10.1007/s11032-012-9765-0

Osborn TC, Butrulle DV, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiate DJ (2003) Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. Genetics 165(3):1569–1577

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). Genome 38(6):1122–1131. https://doi.org/10.1139/g95-149

Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, Denoeud F, Belcram H, Links MG, Just J, Clarke C, Bender T, Huebert T, Mason AS, Pires JC, Barker G, Moore J, Walley PG, Manoli S, Batley J, Edwards D, Nelson MN, Wang X, Paterson AH, King G, Bancroft I, Chalhoub B, Sharpe AG (2014) Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. Genome Biol 15(6):R77. https://doi.org/10.1186/gb-2014-15-6-r77

Pearce S, Saville R, Vaughan SP, Chandler PM, Wilhelm EP, Sparks CA, Al-Kaff N, Korolev A, Boulton MI, Phillips AL, Hedden P, Nicholson P, Thomas SG (2011) Molecular characterization of *Rht-1 d*warfing genes in hexaploid wheat. Plant Physiol 157(4):1820. https://doi.org/10.1104/pp.111.183657

Pires JC, Zhao JW, Schranz ME, Leon EJ, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (*Brassicaceae*). Biol J Linn Soc Lond 82(4):675–688. https://doi.org/10.1111/j.1095-8312.2004.00350.x

Qian L, Voss-Fels K, Cui Y, Jan HU, Samans B, Obermeier C, Qian W, Snowdon RJ (2016) Deletion of a stay-green gene associates with adaptive selection in *Brassica napus*. Mol Plant 9(12):1559–1569. https://doi.org/10.1016/j.molp.2016.10.017

Reeves TG, Rajaram S, van Ginkel M, Trethowan R, Braun H, Cassaday K (1999) New wheats for a secure, sustainable future. CIMMYT, Mexico

Richter TE, Pryor TJ, Bennetzen JL, Hulbert SH (1995) New rust resistance specificities associated with recombination in the *Rp1* complex in maize. Genetics 141(1):373–381

Riley R, Coucoli H, Chapman V (1967) Chromosomal interchanges and the phylogeny of wheat. Heredity 22:233. https://doi.org/10.1038/hdy.1967.29

Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D (2006) *APP* locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38(1):24–26. https://doi.org/10.1038/ng1718

Samans B, Chalhoub B, Snowdon RJ (2017) Surviving a genome collision: genomic signatures of allopolyploidization in the recent crop species. Plant Genome. https://doi.org/10.3835/plantgenome2017.02.0013

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct Genomics 13(4):296–307. https://doi.org/10.1093/bfgp/elu016

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon R (2017a) Post-polyploidisation morphotype diversification associates with gene copy number variation. Sci Rep 7:41845. https://doi.org/10.1038/srep41845

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon RJ (2017b) Targeted deep sequencing of flowering regulators in *Brassica napus* reveals extensive copy number variation. Sci Data. https://doi.org/10.1038/sdata.2017.13

Schiessl S-V, Katche E, Ihien E, Chawla HS, Mason AS (2018) The role of genomic structural variation in the genetic improvement of polyploid crops. Crop J. https://doi.org/10.1016/j.cj.2018.07.006

Schmidt MH-W, Vogel A, Denton AK, Istace B, Wormit A, van de Geest H, Bolger ME, Alseekh S, Maß J, Pfaff C, Schurr U, Chetelat R, Maumus F, Aury J-M, Koren S, Fernie AR, Zamir D, Bolger AM, Usadel B (2017) De novo Assembly of a new *Solanum pennellii* accession using nanopore sequencing. Plant Cell 29(10):2336–2348. https://doi.org/10.1105/tpc.17.00521

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, van Buren P, Vaughn MW, Ying K, Yeh C-T, Emrich SJ, Jia Y, Kalyanaraman A, Hsia A-P, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia J-M, Deragon J-M, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115. https://doi.org/10.1126/science.1178534

Schröder J, Hsu A, Boyle SE, Macintyre G, Cmero M, Tothill RW, Johnstone RW, Shackleton M, Papenfuss AT (2014) Socrates: identification of genomic rearrangements in tumour genomes by re-aligning soft clipped reads. Bioinformatics 30(8):1064–1072. https://doi.org/10.1093/bioinformatics/btt767

Sears ER (1939) Cytogenetic studies with polyploid species of wheat. I. chromosomal aberrations in the progeny of a haploid of *Triticum Vulgare*. Genetics 24(4):509–523

Sharp AJ, Cheng Z, Eichler EE (2006) Structural variation of the human genome. Annu Rev Genomics Hum Genet 7:407–442. https://doi.org/10.1146/annurev.genom.7.080505.115618

Sharpe AG, Parkin IA, Keith DJ, Lydiate DJ (1995) Frequent nonreciprocal translocations in the amphidiploid genome of oilseed rape (*Brassica napus*). Genome 38(6):1112–1121

Sieber A-N, Longin CFH, Leiser WL, Würschum T (2016) Copy number variation of *CBF-A14* at the *Fr-A2* locus determines frost tolerance in winter durum wheat. Theor Appl Genet 129(6):1087–1097. https://doi.org/10.1007/s00122-016-2685-3

Singleton AB (2003) Synuclein locus triplication causes Parkinson's disease. Science 302(5646):841. https://doi.org/10.1126/science.1090278

Snowdon RJ (2007) Cytogenetics and genome analysis in *Brassica* crops. Chromosome Res 15(1):85–95. https://doi.org/10.1007/s10577-006-1105-y

Springer NM, Ying K, Fu Y, Ji T, Yeh C-T, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, Iniguez AL, Barbazuk WB, Jeddeloh JA, Nettleton D, Schnable PS, Ecker JR (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet 5(11):e1000734. https://doi.org/10.1371/journal.pgen.1000734

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61(1):437–455. https://doi.org/10.1146/annurev-med-100708-204735

Stein N, Mascher M (2019) Capturing pangenome diversity for breeding: a case study in barley. Theor Appl Genet **(to be published in same special issue)**

Stein A, Coriton O, Rousseau-Gueutin M, Samans B, Schiessl SV, Obermeier C, Parkin IAP, Chèvre A-M, Snowdon RJ (2017) Mapping of homoeologous chromosome exchanges influencing quantitative trait variation in *Brassica napus*. Plant Biotechnol J 15(11):1478–1489. https://doi.org/10.1111/pbi.12732

Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, Liu H, Ma X, Jiao Y, Wang B, Wei X, Stein JC, Glaubitz JC, Lu F, Yu G, Liang C, Fengler K, Li B, Rafalski A, Schnable PS, Ware DH, Buckler ES, Lai J (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat Genet 50(9):1289–1295. https://doi.org/10.1038/s41588-018-0182-0

Sutton T, Baumann U, Hayes J, Collins NC, Shi B-J, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science 318(5855):1446–1449. https://doi.org/10.1126/science.1146853

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20(12):1689–1699. https://doi.org/10.1101/gr.109165.110

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics 190(4):1563–1574. https://doi.org/10.1534/genetics.111.137349

Taylor CM, Kamphuis LG, Zhang W, Garg G, Berger JD, Mousavi-Derazmahalleh M, Bayer PE, Edwards D, Singh KB, Cowling WA, Nelson MN (2018) INDEL variation in the regulatory region of the major flowering time gene *LanFTc1* is associated with vernalization response and flowering time in narrow-leafed lupin (*Lupinus angustifolius* L.). Plant Cell Environ. https://doi.org/10.1111/pce.13320

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc Natl Acad Sci USA 102(39):13950. https://doi.org/10.1073/pnas.0506758102

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796. https://doi.org/10.1038/35048692

The International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345(6194):1251788. https://doi.org/10.1126/science.1251788

Tranchida-Lombardo V, Aiese Cigliano R, Anzar I, Landi S, Palombieri S, Colantuono C, Bostan H, Termolino P, Aversano R, Batelli G, Cammareri M, Carputo D, Chiusano ML, Conicella C, Consiglio F, D'Agostino N, de Palma M, Di Matteo A, Grandillo S, Sanseverino W, Tucci M, Grillo S (2018) Whole-genome re-sequencing of two Italian tomato landraces reveals sequence variations in genes associated with stress tolerance, fruit quality and long shelf-life traits. DNA Res 25(2):149–160. https://doi.org/10.1093/dnares/dsx045

Tseng S-H, Peng S-F, Cheng Y-M (2017) Analysis of B chromosome nondisjunction induced by the *r-X1* deficiency in maize. Chromosome Res 42(2):223. https://doi.org/10.1007/s10577-017-9567-7

Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus* L. Genetics 169(2):967–979. https://doi.org/10.1534/genetics.104.033209

Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, Hara N, Kitomi Y, Inukai Y, Ono K, Kanno N, Inoue H, Takehisa H, Motoyama R, Nagamura Y, Wu J, Matsumoto T, Takai T, Okuno K, Yano M (2013) Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. Nat Genet 45(9):1097–1102. https://doi.org/10.1038/ng.2725

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009) The flowering world: a tale of duplications. Trends Plant Sci 14(12):680–688. https://doi.org/10.1016/j.tplants.2009.09.001

Vialette-Guiraud ACM, Adam H, Finet C, Jasinski S, Jouannic S, Scutt CP (2011) Insights from ANA-grade angiosperms into the early evolution of *CUP-SHAPED COTYLEDON* genes. Ann Bot 107(9):1511–1519. https://doi.org/10.1093/aob/mcr024

Villareal RL, Toro E, Mujeeb-Kazi A, Rajaram S (1995) The 1BL/1RS chromosome translocation effect on yield characteristics in a *Triticum aestivum* L. cross. Plant Breed 114(6):497–500. https://doi.org/10.1111/j.1439-0523.1995.tb00843.x

Voss-Fels K, Snowdon RJ (2016) Understanding and utilizing crop genome diversity via high-resolution genotyping. Plant Biotechnol J 14(4):1086–1094. https://doi.org/10.1111/pbi.12456

Wang W, Wang S, Hou C, Xing Y, Cao J, Wu K, Liu C, Zhang D, Zhang L, Zhang Y, Zhou H (2014) Genome-wide detection of copy number variations among diverse horse breeds by array CGH. PLoS ONE 9(1):e86860. https://doi.org/10.1371/journal.pone.0086860

Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W, Meng X, Liu R, Chen H, Jing Y, Wang Y, Zhu X, Li J, Qian Q (2015) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. Nat Genet 47:944. https://doi.org/10.1038/ng.3346

Wang X, Wang H, Liu S, Ferjani A, Li J, Yan J, Yang X, Qin F (2016) Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. Nat Genet 48(10):1233–1241. https://doi.org/10.1038/ng.3636

Wulff BBH, Moscou MJ (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. Front Plant Sci 5:692. https://doi.org/10.3389/fpls.2014.00692

Würschum T, Boeven PHG, Langer SM, Longin CFH, Leiser WL (2015) Multiply to conquer: copy number variations at *Ppd-B1* and *Vrn-A1* facilitate global adaptation in wheat. BMC Genet 16(1):949. https://doi.org/10.1186/s12863-015-0258-0

Würschum T, Longin CFH, Hahn V, Tucker MR, Leiser WL (2016) Copy number variations of CBF genes at the *Fr-A2* locus are essential components of winter hardiness in wheat. Plant J 89(4):764–773. https://doi.org/10.1111/tpj.13424

Xiong Z, Pires JC (2011) Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. Genetics 187(1):37–49. https://doi.org/10.1534/genetics.110.122473

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc Natl Acad Sci USA 108(19):7908–7913. https://doi.org/10.1073/pnas.1014138108

Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, Heuer S, Ismail AM, Bailey-Serres J, Ronald PC, Mackill DJ (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. Nature 442(7103):705–708. https://doi.org/10.1038/nature04920

Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L, Li J, He W, Zhang G, Zheng X, Zhang F, Li Y, Yu C, Kristiansen K, Zhang X, Wang J, Wright M, McCouch S, Nielsen R, Wang J, Wang W (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Biotechnol 30(1):105–111. https://doi.org/10.1038/nbt.2050

Yao W, Li G, Zhao H, Wang G, Lian X, Xie W (2015) Exploring the rice dispensable genome using a metagenome-like assembly strategy. Genome Biol 16:187. https://doi.org/10.1186/s13059-015-0757-3

Yang S, Li J, Zhang X, Zhang Q, Huang J, Chen J-Q, Hartl DL, Tian D (2013) Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Natl Acad Sci USA 110(46):18572–18577. https://doi.org/10.1073/pnas.1318211110

Ye K, Schulz MH, Long Q, Apweiler R, 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(21):2865–2871. https://doi.org/10.1093/bioinformatics/btp394

Yu P, Wang C, Xu Q, Feng Y, Yuan X, Yu H, Wang Y, Tang S, Wei X (2011) Detection of copy number variations in rice using array-based comparative genomic hybridization. BMC Genomics 12(1):372. https://doi.org/10.1186/1471-2164-12-372

Yu P, Wang C-H, Xu Q, Feng Y, Yuan X-P, Yu H-Y, Wang Y-P, Tang S-X, Wei X-H (2013) Genome-wide copy number variations in *Oryza sativa* L. BMC Genomics 14(1):649. https://doi.org/10.1186/1471-2164-14-649

Zhang Z, Mao L, Chen H, Bu F, Li G, Sun J, Li S, Sun H, Jiao C, Blakely R, Pan J, Cai R, Luo R, Van de Peer Y, Jacobsen E, Fei Z, Huang S (2015) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. Plant Cell 27(6):1595–1604. https://doi.org/10.1105/tpc.114.135848

Zhao Q, Feng Q, Lu H, Li Y, Wang A, Tian Q, Zhan Q, Lu Y, Zhang L, Huang T, Wang Y, Fan D, Zhao Y, Wang Z, Zhou C, Chen J, Zhu C, Li W, Weng Q, Xu Q, Wang Z-X, Wei X, Han B, Huang X (2018) Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat Genet 50(2):278–284. https://doi.org/10.1038/s41588-018-0041-z

Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, Liu T-F, Jiang S, Ramachandran S, Liu C-M, Jing H-C (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol 12(11):R114. https://doi.org/10.1186/gb-2011-12-11-r114

Zhou P, Silverstein KAT, Ramaraj T, Guhlin J, Denny R, Liu J, Farmer AD, Steele KP, Stupar RM, Miller JR, Tiffin P, Mudge J, Young ND (2017) Exploring structural variation and gene family architecture with *de novo* assemblies of 15 *Medicago* genomes. BMC Genomics 18(1):261. https://doi.org/10.1186/s12864-017-3654-1

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theor Appl Genet 127(1):1–18. https://doi.org/10.1007/s00122-013-2177-7

Zou J, Hu D, Mason AS, Shen X, Wang X, Wang N, Grandke F, Wang M, Chang S, Snowdon RJ, Meng J (2018) Genetic changes in a novel breeding population of *Brassica napus* synthesized from hundreds of crosses between *B. rapa* and *B. carinata*. Plant Biotechnol J 16(2):507–519. https://doi.org/10.1111/pbi.12791