Copy number Variation and Disease Resistance in Plants

Aria Dolatabadian, Dhwani Apurva Patel, David Edwards and Jacqueline Batley*

School of Plant Biology and Institute of Agriculture, University of Western Australia, Crawley, WA, 6009, Australia

***Corresponding author:** Jacqueline.batley@uwa.edu.au; +61 (0)8 6488 5929

Abstract

Plant genome diversity varies from single nucleotide polymorphisms (SNPs) to large-scale deletions, insertions, duplications, or re-arrangements. These re-arrangements of sequences resulting from duplication, gains or losses of DNA segments are termed copy number variations (CNVs). During the last decade, numerous studies have emphasized the importance of CNVs as a factor affecting human phenotype; in particular, CNVs have been associated with risks for several severe diseases. In plants, the exploration of the extent and role of CNVs in resistance against pathogens and pests is just beginning. Since CNVs are likely to be associated with disease resistance in plants, an understanding of the distribution of CNVs could assist in the identification of novel plant disease-resistance genes. In this paper, we review existing information about CNVs; their importance, role and function, as well as their association with disease resistance in plants.

Introduction

With the advent of next-generation sequencing (Bayer et al. 2015) and genotyping methods such as array or bead-based genotyping (Dalton-Morgan et al. 2014; Mason et al. 2015) that generate enormous quantities of data, more genetic associations are being uncovered than ever before (Abel and Duncavage 2013; Golicz et al. 2015). However, little is known about the functional impact of copy number variations (CNVs) at the cellular and organismal level.

The identification of structural variants (SVs), especially CNVs, among the genomes of individuals has provided the rationale to redefine genomes as dynamic entities (Muñoz-Amatriaín et al. 2013). Copy number variations were first considered to reflect human diversity (Zhang et al. 2009), but studies revealed that CNVs are also common in animal (Fadista et al. 2010; Liu et al. 2010; Gazave et al. 2011; Nicholas et al. 2011; Berglund et al. 2012) and plant (Springer et al. 2009; Lee et al., 2015; Trębicki et al. 2015) genomes. Genomic data obtained from different plant species as a part of large-scale sequencing projects, highlights CNV as one contributor to natural diversity on the genomic level, for example Bai et al. (2016) have validated 28 functional CNV genes including *OsMADS56*, *BPH14*, *OsDCL2b* and *OsMADS30*, implying that CNVs might be involved in control of flowering time, insect resistance, RNA interference, response to salt and dehydration stress. Most CNV genes were found to be located in non-co-linear positions by comparison to *O. glaberrima*. Several recent studies provide insights into the extent of this type of structural variation in plants, however fewer studies have been directed towards understanding the role of CNVs in plants (Saxena et al. 2014).

Structural Variation within the Genome

Genomic variation can be present in many forms, including SNPs (Mason et al. 2015), variable number of tandem repeats (VNTRs; e.g., mini- and microsatellites) (Katoh et al. 2015), presence/absence of transposable elements (e.g. Alu elements; Freeman et al. 2006), and insertions, deletions, duplications, inversions, and copy number variation of DNA segments ranging in size from a few base pairs to entire chromosomes (Sebat et al. 2004; Conrad et al. 2006; Redon et al. 2006; Muñoz-Amatriaín et al. 2015). It has recently become clear that much of the natural genetic variation that exists between individuals is due to alterations in the number of copies of genes rather than small differences in the nucleotide sequence (Girirajan et al. 2011; Veltman and Brunner 2012). Understanding how structural variation affects phenotype is therefore a major challenge of modern genetics.

Genomic variations that involve segments of DNA larger than 1 kb in length (Feuk et al. 2006) can be classified as a structural variation of which Copy Number Variation (CNVs) and Presence Absence Variation

(PAVs) are the most commonly known. These can be further categorised as microscopic (detectable with optical microscopes) or sub-microscopic ranging from ~1 kb to 3 Mb in size (Feuk et al. 2006) depending on the method of their detection (Saxena et al. 2014). CNVs are chromosomal deletions, insertions and/or duplications that are typically defined as DNA segments that are present in a different number of copies when compared to a reference genome (Feuk, et al. 2006; Scherer et al. 2007; Figure 1), whereas PAVs can be considered an extreme form of CNV, where the sequence is completely missing from one or more individual (Saxena et al. 2014).

There are two main mechanisms of structural variation formation. The first mechanism is known as non-homologous end joining (NHEJ) and requires very low level of sequence similarity at the breakpoints. It is the result of aberrant repair of uneven double-stranded breaks produced following DNA damage (Bignell et al. 2007; Campbell et al. 2008). A second mechanism proposed for repetitive sequences in the genome is termed non-allelic homologous recombination and this requires high sequence similarity at the breakpoints (Kolomietz et al. 2002; Kidd et al. 2008). Several other mechanisms for structural variation production have also been proposed, such as fork stalling and template switching (FoSTeS) (Stankiewicz and Lupski 2010).

CNV Detection Methods

Several methods have been developed to detect CNVs: quantitative and digital PCR, in situ fluorescent hybridization (Weaver et al. 2010), the paralogue ratio test (Armour et al. 2007), multiplex amplifiable probe hybridization (Armour et al. 2000) and multiplex ligation-dependent probe amplification (Marcinkowska-Swojak et al. 2013). PCR can also identify small translocations and inversions, as well as indel polymorphisms and CNVs (Wang et al. 2006). Microarray-based techniques were among the first used to detect genome-wide variation in human and plant genomes. Initial studies of CNVs in plants have used array-based approaches, for example studies in *Arabidopsis thaliana* (DeBolt 2010), maize (Swanson-Wagner et al. 2010), rice (Yu et al. 2011) and barley (Muñoz-Amatriaín et al. 2013).

Recent advances in massive parallel sequencing technology and development of novel analytical algorithms now allow for the detection of CNVs using short read sequencing data (Xi et al. 2012; Zhao et al. 2013; Wang et al. 2014). Four sequencing-based approaches are commonly used in CNV detection (Table 1): (1) the read depth (RD) approach, which relies on changes in normalized read depth to estimate gains and losses of copies; (2) the read pair (RP) approach, which is based on discordantly mapped read pairs; (3) the split read (SR) approach, which uses gapped read alignments (Alkan et al. 2011) and (4) the assembly (AS) approach (Pirooznia et al. 2015),

which generates contigs/scaffolds that are then compared with the reference genome to discover structural variation (Teo et al. 2012).

Compared with array-based approaches sequencing-based approaches, especially RP and SR, have clear advantages, including the ability to detect CNVs of relatively small sizes (< 1 kb), and to infer the breakpoints of CNVs at nucleotide level resolution (Li and Olivier 2013). This is important to assess their genomic impact and to infer their formation mechanisms (Lam et al. 2009). In plants, only a few studies have used sequencing-based approaches to date and most relied solely on the RD method (Turner et al. 2010; Cao et al. 2011; Zheng et al. 2011; Flagel et al. 2013), which may entail a size bias towards larger variants with low breakpoint resolution (Xi et al., 2012; Li and Olivier, 2013).

CNVs and Pangenome Studies

Pangenome construction is a powerful approach which has been developed to understand the extent to which genomic variation occurs within a species. The term pangenome refers to the complete and non-redundant set of genes in the entire species; it is composed of core genes, which are present in all individuals, and variable genes, which are present only in some individuals (Golicz et al. 2016; Hurgobin and Edwards, 2017). Due to the presence of structural variation in the form of CNVs and PAVs, a single reference genome is not sufficient to fully represent the entire genetic diversity of a certain species. Accordingly, to obtain the complete genomic content of any given species, it is necessary to construct its pangenome.

Golicz et al. (2016) found that a large number of genes with annotations related to major agronomic traits, such as disease resistance, in the *B. oleracea* pangenome were affected by PAV and CNV. Similarly, Yao et al. (2015) have reported that the variable genome of rice was enriched with genes related to defence to biotic stress, including NBS LRR genes and genes coding for protein kinases and abiotic stress tolerance. Gene copy number variation in *B. rapa* pangenome was studied by Lin et al. (2014), who found evidence for copy number differences in a peroxidase (EC 1.11.1.7), pointing to a role for the phenylpropanoid biosynthesis pathway in the generation of morphological variation. They have reported that lower copy number of genes in turnip coding for a glucosyltransferase (EC 2.4.1.111) may cause the reduction of 4-D-glucoside, coniferin, syringin and hence increase the production of different lignins. Furthermore, 49,000–169,000 copy number variants were identified in *Medicago* genomes (Zhou et al. 2017). Li et al. (2014) established and analysed the pangenome of *Glycine soja*. They reported that intergenomic comparisons identified lineage-specific genes and genes with CNV or large-effect mutations, some of which show evidence of positive selection and may contribute to variation of agronomic

traits such as biotic resistance, seed composition, flowering and maturity time, organ size and final biomass. A genome-wide analysis of structural variation in three inter-crossable poplar species: *Populus nigra, Populus deltoides*, and *Populus trichocarpa* was performed by Pinosio et al. (2016) to characterise the size and the composition of the poplar pan-genome. They detected a total of 7,889 deletions and 10,586 insertions relative to the *P. trichocarpa* reference genome, and 3,230 genes affected by CNV.

CNVs and their Importance

The American geneticist Calvin Bridges first discovered CNVs in 1936, when he noticed that flies that inherit a duplicate copy of the *Bar* gene developed very small eyes. Over the past several years, many new CNVs in different species have been identified, leading researchers to believe that CNVs are as important a component of genomic diversity as SNPs.

Structural variations, including CNVs, have been identified in several plant species, including *Arabidopsis* (DeBolt 2010), barley (*Hordeum vulgare*) (Muñoz-Amatriaín et al. 2013), foxtail millet (*Setaria italica*) (Bai et al. 2013), maize (*Zea mays*) (Swanson-Wagner et al. 2010), rice (*Oryza sativa*) (Xu et al. 2012), sorghum (*Sorghum bicolor*) (Zheng et al. 2011), soybean (*Glycine max*) (McHale et al. 2012) and wheat (*Triticum aestivum*) (Nishida et al. 2013). In several cases, these structural variations were found to be associated with phenotypic variations such as leaf size in *Arabidopsis thaliana* (Horiguchi et al. 2009), fruit shape in tomato (Xiao et al. 2008), aluminium tolerance in maize (Maron et al. 2013), stress and disease resistance in barley (Muñoz-Amatriaín et al. 2013) and grain size in rice (Wang et al. 2015).

In the genome, there are regions that seem to be more prone to CNV than others, due to their specific structural features that will locally induce the mechanisms leading to CNV formation, e.g., non-allelic recombination (Zmienko et al. 2016; Samelak-Czajka et al. 2017). Differences in the DNA sequence of species' genomes contribute to their uniqueness. These variations influence many traits, including organism's fitness, susceptibility to disease and contribute to the adaptation to environmental challenges, as well as to co-evolutionary interactions between host and pathogen or a symbiont (Kondrashov, 2012, Żmieńko et al., 2014).

The biological effect of CNVs is dependent on the affected sequences and their interactions with the rest of the genome. The importance of CNVs may be greater if they contain regulatory regions and/or genes and these CNVs may contribute to phenotype variation (Cong et al. 2008). Copy number variation may also have potential functional effects: they can cause changes in gene structure, gene dosage, or expression regulation, and expose recessive alleles to selection (Bickhart et al. 2012).

In the model plant species *Arabidopsis* (DeBolt 2010) and rice (Yu et al. 2011), CNVs were detected in 402 and 641 genes, respectively. Genome-wide patterns of CNVs have also been detected in sorghum by comparing two sweet and one grain inbred sorghum lines, identifying 3234 CNVs in 2600 genes (Zheng et al. 2011). Among the legumes, soybean was the first species to have its genome analysed for CNVs, and a total of 267 CNVs with an average size of 18–23 kb, were detected across the genomes assayed (McHale et al. 2012). In contrast to maize (Belo et al. 2010; Swanson-Wagner et al. 2010), higher levels of CNV were identified in high-recombination regions in soybean and barley (McHale et al. 2012; Muñoz-Amatriaín et al. 2013).

Despite the prevalence of CNVs in plant genomes and their frequent overlap with protein-coding regions, only a few have been associated with particular phenotypes (Żmieńko et al. 2014). With further studies, we expect to grow our understanding of CNVs impact on plant phenotype, both in the aspect of long-term evolution as well as a mechanism of rapid adaptation to environmental challenges.

CNV – Their Role in Plants

CNV variation has been implicated to play in role in several different processes associated with plants. The first plant species to be extensively assessed for CNVs was maize (Springer et al. 2009; Belo et al. 2010; Żmieńko et al. 2014), and many of the CNVs identified in 19 diverse inbred maize lines and 14 teosinte accessions were found to be associated with domestication (Swanson-Wagner et al. 2010; Chia et al. 2012). This identified 479 genes with higher copy number and 3410 genes with fewer copies following comparison with a reference genome (Swanson-Wagner et al. 2010).

Many CNVs have been observed in outcrossing and autogamous species (Żmieńko et al. 2014). Changes in gene copy number may provide a way to alter the effective dosage of a gene, which may directly change the phenotype. If the new variant is beneficial, the copy number in a particular region may accumulate, and the phenotypic effects may intensify. An example of rapid evolution in a plant is resistance to glyphosate in Palmer amaranth (*Amaranthus palmeri*) a major weed pest in the southern part of the United States (Żmieńko et al. 2014). It was shown that Palmer amaranth resistance to glyphosate is driven by an increase in 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene copy number, which is associated with increased EPSPS transcript and protein levels, as well as increased glyphosate dose survival rate (Gaines et al. 2010, 2011; Sammons and Gaines, 2014). In addition, Iwakami et al. (2017) have found differences in acetolactate synthase (ALS) gene copy numbers among thifensulfuron-methyl resistant short-awned foxtail (*Alopecurus aequalis*) accessions. They have reported that two copies, ALS1 and ALS2, were conserved in all accessions, while some carried two additional

copies, ALS3 and ALS4. A single-base deletion in ALS3 and ALS4 further indicated that they represented pseudogenes. Good examples of a CNV affecting phenotype is found in the diversity of flowering times and plant heights in wheat and canola (Diaz et al. 2012; Li et al. 2012; Würschum et al. 2015; Schiessl et al. 2017). Several confirmed examples of a CNV linked to phenotype also concern plant stress tolerance (Żmieńko et al. 2014; Sieber et al. 2016). The importance of CNV at the Fr-A2 locus was shown in durum wheat, in which the Fr-A2 locus explained approximately 90% of the genotypic variation of frost tolerance (Sieber et al. 2016). Furthermore, Würschum et al. (2017) have shown that CNV of C-repeat Binding Factor (CBF) genes at the Fr-A2 locus is the essential component for winter survival, with CBF-A14 CNV being the most likely causal polymorphism, accounting for 24.3% of the genotypic variance. Changes in gene copy number have been reported to be associated with tolerance to toxic soil chemicals in plants. Copy number expansion of the metal pump gene *HMA4*, for example, contributes to hyper-accumulation and hyper-tolerance to zinc and cadmium in *A. halleri* (Hanikenne et al. 2008; Hanikenne et al. 2013). Similarly, boron-tolerant genotypes of barley contain four times as many copies of the boron transporter gene (*Bot1*) than intolerant genotypes (Sutton et al. 2007), and aluminium tolerance in maize is associated with higher copy number of the multidrug and toxin extrusion gene *MATE1* (Maron et al. 2013). CNVs were identified for the MATE1 gene in aluminium-tolerant lines, but these were not common in teosinte. This study suggested that multiple copies of the MATE1 gene arose recently and probably after domestication.

CNVs and Disease Resistance in Plants

CNVs have been found to be associated with nucleotide-binding leucine-rich repeat (NB-LRR) genes and receptor-like kinase (RLK) genes, known to be involved in plant defence-related mechanisms (Saxena et al. 2014). Sequence variation within the central LRR domain and variation in LRR copy number play important roles in determining recognition specificity (Gururani et al. 2012). CNVs can also be linked to variation in gene expression (Orozco et al. 2009; Ortiz-Estevez et al. 2011). For example, Scots pine trees (*Pinus sylvestris*) were tested for CNV of a thaumatin-like protein gene involved in resistance against root rot by Šķipars et al. (2011) who identified variation in the gene copy number of the thaumatin-like protein gene.

Among the functionally annotated genes, those which are usually over-represented within CNV regions are genes encoding proteins with a nucleotide binding domain (NB) and one or more leucine-rich repeat (LRR) domains (known as NB-LRR genes), as well as genes encoding receptor-like kinases (RLK). Yu et al. (2013) identified several disease resistance genes within the CNV regions in rice (Table 2). These genes were

considerably enriched for specific biological functions involved in cell death, protein phosphorylation, and defence response. Furthermore, genetic mechanisms for copy number variation of resistance genes were investigated through phylogenetic comparison of resistance genes in the Cucurbitaceae family by Lin et al. (2013). Their analysis of *R* genes showed frequent loss of *R*-gene loci in different Cucurbitaceae species (Table 2). Chalhoub et al. (2014) identified 425 nucleotide binding site leucine-rich repeat (NBS-LRR) sequences encoding resistance gene homologs in *Brassica napus* using genome sequencing. They confirmed the absence of five NBS-LRR genes from the A_n sub-genome, and three from the C_n sub-genome. This variation may reflect differential selection for resistance to diseases. Hardigan et al. (2016) examined the breadth of genome-wide structural variation in a panel of monoploid/doubled monoploid clones generated from native populations of diploid potato (*Solanum tuberosum*) and found CNVs on chromosome 11 at 42.59 to 43.05 Mb. This location pertained to a cluster of 16 genes encoding nucleotide binding site leucine-rich repeat (NBS-LRR) disease resistance proteins of which,14 showed variation in copy number. Association analysis of traits involved in leaf development and disease resistance in 103 maize lines using both SNPs and CNVs revealed that CNVs contribute greatly to the variation of analysed phenotypes and provide complementary information to SNPs (Chia et al. 2012). Jamann et al. (2014) found that CNV polymorphism was significantly associated with resistance to northern leaf blight based on nested association mapping GWAS.

Gene ontology term enrichment analysis of the 672 genes located within CNV regions in soybean revealed that genes related to disease resistance response were significantly over-represented (McHale et al. 2012). In addition, it has been reported that resistance gene function is adapted to frequent re-arrangements and copy number variations (Leister et al. 1998). Copy number variation of a 31 kb repeat segment observed in different haplotypes of the *Rhg1* locus encode multiple gene products in soybean cyst nematode (SCN) resistant varieties (Lee et al. 2015). The cloning of *Rhg1* was the first observation that plant disease resistance loci can consist of a multi gene cluster CNV of non-canonical resistance genes in tandem formation (Cook et al. 2012). In SCN-susceptible varieties, one copy of the 31 kb segment per haploid genome was present. SCN resistance was found to be associated with increased expression of CNV-related genes (Cook et al. 2012). Copy number variations related to disease resistance have also been identified in several plant species (Table 2), where disease resistance genes represent a significant fraction of genes in CNV regions and were significantly enriched for resistance gene models (Xu et al. 2012; Lu et al. 2012). For instance, Boocock et al. (2015) have identified 876 CNV regions, which spanned 3.5% of the apple genome and were enriched for genes involved in disease resistance against apple

scab. Bertioli et al. (2003) showed that in peanut and legumes *R*-genes have undergone extensive copy number variation.

It is expected that high copy number of resistance genes in plants is advantageous because it will offer better resistance against pathogens (Lin et al. 2013). On the other hand, low copy number might be a result of less challenge from pathogens (Zhai et al. 2011). This supports the hypothesis that CNV and the genes encoded within these regions contribute to disease resistance in plants through natural genome variation. CNV could enable gene diversification and evolution of new resistance genes

Conclusions and future directions

Although several projects have been completed in order to detect CNVs and to understand how they are implicated in different species, the field still lacks sufficient results in the area such as association of CNVs with disease resistance in plants. There are still limitations in accurate detection of CNVs, which need to be improved in future. These are not only due to the difference in the quality of different commonly used short read NGS sequencing technologies, which can result in the detection of platform-specific variants,, but also due to variations when using different, commonly used bioinformatics tools. This should be taken into account when analysing CNV data as different CNV regions were detected by Lam et al. (2012) and O'Raw et al. (2013) when working on the same data. Nonetheless, in the future, in addition to studying the role of CNV in plant physiology, analysis and quantification of CNV in plants will likely be used in plant breeding as part of acquisition of desirable traits.

An integrated map of CNV in a plant will be helpful to understand the distribution of CNV, as well as its evolutionary mechanism. It is also useful for future mapping and cloning of *R*-genes, the most divergent gene family in plant genomes which has been shown to have considerable copy number variation, presence/absence polymorphism as well as sequence variation. Through constant improvement in genome sequencing and ever decreasing costs of this technology, more crop genomes are being sequenced. Multiple cultivars within a species having sequences available and this is being extended to the availability of pan-genomes for many species (Golicz et al., 2015; Golicz et al., 2016). These ever-increasing genomic resources will enable a higher accuracy of CNV detection and association with traits. In future, marker assisted selection can be used as a potential tool for genetic improvement using both CNV and SNP association with disease resistance. A greater understanding of selection pressures of various diseases on CNV will further our knowledge of plant-pathogen interactions. This information will suggest a way forward where information about CNVs can be applied to trait association and breeding.

Previous studies provide CNV estimates across the plant genome, enabling further research into the role of such variations in resistance genes.

The ability to use next-generation sequencing to identify CNV paves the road to make correlations between phenotypic and genotypic characteristics, therefore detecting these variations, using new DNA sequencing technologies, is the first step towards identification of CNV associated economic traits to integrate them into plant genomic selection programs.

Acknowledgements: *Australian Research Council Projects FT130100604, DP1601004497, LP140100537, LP160100030, LP130100925*

Author contribution statement

Concept and Structure of manuscript: JB

Figure: AD

Manuscript writing: AD, DP, DE, JB

All authors read and approved the final manuscript

Disclosure of potential conflicts of interest

The authors declare no conflicts of interest

References

Abel HJ, Duncavage EJ (2013) Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer Genet 206:432-440

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12:363-376

Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ (2007) Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 35:1-8

Armour JA, Sismani C, Patsalis PC, Cross G (2000) Measurement of locus copy number by hybridisation with amplifiable probes. Nucleic Acids Res 28:605-609

Bai H, Cao Y, Quan J et al (2013) Identifying the genome-wide sequence variations and developing new molecular markers for genetics research by re-sequencing a landrace cultivar of foxtail millet. PLoS One 8:e73514

Bai Z, Chen J, Liao Y, Wang M, Liu R, Ge S, et al. (2016). The impact and origin of copy number variations in the *Oryza* species. BMC Genomics 17:261.

Bayer PE, Ruperao P, Mason AS, Stiller J, Chan CKK, Hayashi S, Long Y, Meng J, Sutton T, Visendi P, Varshney RK, Batley J, Edwards D (2015) High resolution skim genotyping by sequencing reveals the distribution of crossovers and gene conversions in *Cicer arietinum* and *Brassica napus*. Theor Appl Genet128:1039-1047

Belo A, Beatty MK, Hondred D et al (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. Theor Appl Genet 120:355-67

Berglund J, Nevalainen EM, Molin A-M, Perloski M. The Lupa Consortium. Andre C, Zody MC, Sharpe T, Hitte C, Lindblad-Toh K, Lohi H, Webster MT (2012) Novel origins of copy number variation in the dog genome. Genome Biol13:R73

Bertioli DJ, Leal-Bertioli SCM, Lion MB, Santos VL, Pappas Jr G, Cannon SB, Guimaraes PM (2003) A large scale analysis of resistance gene homologues in *Arachis*. Mol Gen Genomics 270:34-45

Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF et al (2012) Copy number variation of individual cattle genomes using next-generation sequencing. Genome Res 22:778-790

Bignell GR, Santarius T, Pole JCM et al (2007) Architectures of somatic genomic rearrangement in human cancer amplicons at sequence-level resolution. Genome Res 17:1296-1303

Boocock J, Chagné D, Merriman TR, Black MA (2015) The distribution and impact of common copy-number variation in the genome of the domesticated apple, Malus x domestica Borkh. BMC Genomics 16:848

Bradeen JM, Iorizzo M, Mollov DS, Raasch J, Kramer LC, Millett BP, Austin-Phillips S, Jiang J, Carputo D (2009) Higher copy numbers of the potato RB transgene correspond to enhanced transcript and late blight resistance levels. Mol Plant Microbe Interact 22:437-446

Campbell PC, Stephens PJ, Pleasance ED et al (2008) Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 40:722-729

Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Müller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. Nat Genet 43:956-963

Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B et al (2014) Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345:950-953

Chen K, Wallis JW, Mclellan MD, Larson DE, Kalicki JM, Pohl CS et al (2009) Break Dancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6: 677-681

Chia JM, 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 44:803-807

Cong B, Barrero LS, Tanksley SD (2008) Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication, Nature Genetics 40:800-804

Conrad DF, Andrews, TD, Carter NP, Hurles ME, Pritchard JKA (2006) High-resolution survey of deletion polymorphism in the human genome. Nature Genet. 38:75-81

Cook DE, Lee TG, Guo X et al (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. Science 338:1206-1209

Dalton-Morgan J, Hayward A, Alamery S, Tollenaere R, Mason AS, Campbell E, Patel D, Lorenc MT, Yi B, Long Y, Meng J, Raman R, Raman H, Lawley C, Edwards D and Batley J (2014) A high-throughput SNP array in the amphidiploid species *Brassica napus* shows diversity in resistance genes. Func Integr Genomic 14:643-655

David P, Chen NWG, Pedrosa-Harand A, Thareau V, Sévignac M, Cannon SB, Debouck D, Langin T, Geffroy V (2009) A Nomadic Subtelomeric Disease Resistance Gene Cluster in Common Bean. Plant Physiol Nov 151: 1048-1065

DeBolt S (2010) Copy number variation shapes genome diversity in *Arabidopsis* over immediate family generational scales. Genome Biol Evol 2:441-453

Diaz 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:e33234

Dixon MS, Hatzixanthis K, Jones DA, Harrison K, Jones JDG (1998) The Tomato Cf-5 Disease Resistance Gene and Six Homologs Show Pronounced Allelic Variation in Leucine-Rich Repeat Copy Number. The Plant Cell 10:1915-1925

Douchkov D, Lück S, Johrde A, Nowara D, Himmelbach A, Rajaraman J, Stein N, Sharma R, Kilian B, Schweizer P (2014) Discovery of genes affecting resistance of barley to adapted and non-adapted powdery mildew fungi. Genome Biol 15:518

Fadista J, Thomsen B, Holm LE, Bendixen C (2010) Copy number variation in the bovine genome. BMC Genomics 11:284.

Feuk L, Marshall CR, Wintle RF et al (2006) Structural variants: changing the landscape of chromosomes and design of disease studies. HumMol Genet 15: R57-66

Flagel LE, Willis J, Vision TJ (2013) The standing pool of genomic structural variation in a natural population of *Mimulus guttatus*. Genome Biol Evol 6:53-64

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C (2006) Copy number variation: new insights in genome diversity. Genome Res 16:949-61

Gaines TA, Shaner DL, Ward SM, Leach JE, Preston C, Westra P (2011) Mechanism of resistance of evolved glyphosate-resistant *Amaranthus palmeri*. J Agric Food Chem 59:5886-5889

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang J, Preston C, Leach JE, Westra P (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proc Natl Acad Sci USA 107:1029-1034

Gazave E, Darré F, Morcillo-Suarez C, Petit-Marty N, Carreño A, Marigorta UM, Ryder OA, Blancher A, Rocchi M, Bosch E, Baker C, Marquès-Bonet T, Eichler EE, Navarro A (2011) Copy number variation analysis in the great apes reveals species-specific patterns of structural variation. Genome Res 21:1626-163

Gillet-Markowska A, Richard H, Fischer G, Lafontaine I (2014) Ulysses: accurate detection of low-frequency structural variations in large insert-size sequencing libraries. Bioinformatics 31:801-808

Girirajan S, Campbell CD, Eichler EE (2011) Human copy number variation and complex genetic disease. Annu Rev Genet 45:203-226

Golicz AA, Batley J, Edwards D (2016) Towards plant pangenomics. Plant Biotechnology Journal 14:1099-1105

Golicz AA, Bayer PE, Barker GC, Edger PP, Kim HR, Martinez PA, Chan CKC, Severn-Ellis A, McCombie WR, Parkin IAP, Paterson AH, Pires JC, Sharpe AG, Tang H, Teakle GR, Town CD, Batley J, Edwards D. (2016) The pangenome of an agronomically important crop plant *Brassica oleracea*. Nature Communications 7

Golicz AA, Schliep M, Lee HT, Larkum AWD, Dolferus R, Batley J, Chan CKK, Sablok G, Ralph PJ, Edwards D (2015) Genome-wide survey of the seagrass *Zostera muelleri* suggests modification of the ethylene signalling network. J Exp Bot 6:1-10

Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D et al (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. Plant Physiol 157:757-769

Gururani MA, Venkatesh J, Upadhyaya CP, Nookaraju A, Pandey SK, Park SW (2012) Plant disease resistance genes: current status and future directions. Physiol Mol Plant Pathol 78:51-65

Hanikenne M, Kroymann J, Trampczynska A, Bernal M, Motte P, Clemens S, Krämer U (2013) Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. PLoS Genet 9:e1003707

Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Krämer U (2008) Evolution of metal hyperaccumulation required *cis*-regulatory changes and triplication of *HMA4*. Nature 453:391-395

Hardigan MA, Crisovan E, Hamilton JP, Kim J, Laimbeer P, Leisner CP, Manrique-Carpintero NC, Newton L, Pham GM, Vaillancourt B, Xueming Y, Zeng Z, Douches DS, Jiang J, Veilleux RE, Buell CB (2016) Genome Reduction Uncovers a Large Dispensable Genome and Adaptive Role for Copy Number Variation in Asexually Propagated *Solanum tuberosum*. The Plant Cell 28:388-405

Horiguchi G, Gonzalez N, Beemster GT, Inzé D, Tsukaya H (2009) Impact of segmental chromosomal duplications on leaf size in the grandifolia-D mutants of *Arabidopsis thaliana*. Plant J 60:122-133

Hurgobin B, and Edwards D (2017) SNP Discovery Using a Pangenome: Has the Single Reference Approach Become Obsolete? Biology 6(1), 21

Iwakami S, Shimono Y, Manabe Y, Endo M, Shibaike H, Uchino A, Tominaga T (2017). Copy Number Variation in Acetolactate Synthase Genes of Thifensulfuron-Methyl Resistant *Alopecurus aequalis* (Shortawn Foxtail) Accessions in Japan. Front Plant Sci 8:254

Jamann TM, Poland JA, Kolkman JM, Smith LG, Nelson RJ (2014) Unraveling Genomic Complexity at a Quantitative Disease Resistance Locus in Maize. Genetics 198:333-344

Jiang Y, Wang Y, Brudno M (2012) PRISM: pair-read informed split- read mapping for base-pair level detection of insertion, deletion and structural variants. Bioinformatics 28:2576-2583

Katoh H, H Inoue H, Iwanami T (2015) Changes in Variable Number of Tandem Repeats in '*Candidatus* Liberibacter asiaticus' through Insect Transmission. PLoS One 10(9): e0138699

Kidd JM, Cooper GM, Donahue WF et al (2008) Mapping and sequencing of structural variation from eight human genomes. Nature 453:56-64

Klambauer G, Schwarzbauer K, Mayr A, Clevert, DA, Mitterecker A, Bodenhofer U et al (2012) cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res 40: e 69

Kolomietz E, Meyn MS, Pandita A et al (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumours. Genes Chromosomes Cancer 35:97-112

Kondrashov FA (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc Biol Sci 279:5048-5057

Lam HY, Mu XJ, Stütz AM, Tanzer A, Cayting PD, Snyder M, Kim PM, Korbel JO, Gerstein MB (2009) Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. Nat Biotechnol 28:47-55

Lam HYK, Clark MJ, Chen R, Chen R, Natsoulis G, O'Huallachain M, Dewey FE, Habegger L, Ashley EA, Gerstein MB, Butte AJ, Ji HP, Snyder M (2012) Performance comparison of whole-genome sequencing platforms. Nat Biotechnol 30:78-82

Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic frame work for structural variant discovery. Genome Biol 15: R84

Lee OG, Kumar I, Diers BW, Hudson ME (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. Mol Ecol 24:1774-1791

Leister D, Kurth J, Laurie DA, Yano M, Sasaki T, Devos K, Graner A, Schulze-Lefert P (1998) Rapid reorganization of resistance gene homologues in cereal genomes. Proc Natl Acad Sci U S A. 1998 95:370-375

Li W, Olivier M (2013) Current analysis platforms and methods for detecting copy number variation. Physiol Genomics 45:1-16

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 196:282-291

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, Hong 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, QiuL (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits Nature Biotechnology 32: 1045-1054

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 rapa* genomes: a turnip, a rapid cycling and a Chinese cabbage. BMC Genomics 15:250

Lin X, Zhang Y, Kuang H, Chen J (2013) Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. BMC Genomics 14:335

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LCA, Smith TPL, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDaneld TG, Keele JW (2010) Analysis of copy number variations among diverse cattle breeds. Genome Res 20:693-703

Lu P, Han X, Qi J et al (2012) Analysis of *Arabidopsis* genome-wide variations before and after meiosis and meiotic recombination by re-sequencing *Landsberg erecta* and all four products of a single meiosis. Genome Res 22: 508-518

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. Plant Biol 14:253

Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P (2013) An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example. Hum Mutat 34:763-773

Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova, TV, Kudrna D et al (2013) Aluminum tolerance in maize is associated with higher *MATE1* gene copy number. Proc Natl Acad Sci 110:5241-5246

Marschall T, Hajirasouliha I, Schonhuth A (2013) MATE-CLEVER: mendelian-inheritance-aware discovery and genotyping of midsize and long indels. Bioinformatics 29:3143-3150

Mason AM, Zhang J, Tollenaere R; Vasquez Teuber P, Dalton-Morgan J, Hu L, Yan G, Edwards D, Redden R and Batley J (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. Mol Ecol Resour 15:1091-1101

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 159:1295-1308

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 Biology 14:R58

Muñoz-Amatriaín M, Lonardi S, Luo MC et al (2015) Sequencing of 15 622 gene-bearing BACs clarifies the gene-dense regions of the barley genome. Plant J 84:216-227

Nguyen HT, Merriman TR, Black MA (2014) The CNVrd2 package: measurement of copy number at complex loci using high-throughput sequencing data. Front Genet 5:248

Nicholas TJ, Baker C, Eichler EE, Akey JM (2011) A high-resolution integrated map of copy number polymorphisms within and between breeds of the modern domesticated dog. BMC Genomics 12:414

Nijkamp, JF, Van Den Broek MA, Geertman JM, Reinders MJ, Daran JM, De Ridder D (2012) *De novo* detection of copy number variation by co-assembly. Bioinformatics 28, 3195-3202

Nishida H, Yoshida T, Kawakami K et al (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 Breed 31:27-37

O'Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, Bodily P, Tian L, Hakonarson H, Johnson WE, Wei Z, Kai Wang, Lyon GJ (2013) Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing Genome Med 5:28

Orozco LD, Cokus SJ, Ghazalpour A, Ingram-Drake L, Wang S, Van Nas A, Che N, Araujo JA, Pellegrini M, Lusis AJ (2009) Copy number variation influences gene expression and metabolic traits in mice. Hum Mol Genet 18: 4118-4129

Ortiz-Estevez M, De Las Rivas J, Fontanillo C, Rubio A (2011) Segmentation of genomic and transcriptomic microarrays data reveals major correlation between DNA copy number aberrations and gene-loci expression. Genomics 97: 86-93

Pinosio S, Giacomello S, Faivre-Rampant P, Taylor G, Jorge V, Le Paslier MC, Zaina G, Bastien C, Cattonaro F, Marroni F, Morgante M (2016) Characterization of the Poplar Pan-Genome by Genome-Wide Identification of Structural Variation Mol Biol Evol 33: 2706-2719

Pirooznia M, Goes FS, Zandi PP (2015) Whole-genome CNV analysis: advances in computational approaches. Front Genetics 6: 138

Redon R, Ishikawa S, Fitch KR et al (2006) Global variation in copy number in the human genome. Nature 444: 444-454

Samelak-Czajka A, Marszalek-Zenczak M,, Marcinkowska-Swojak M, Kozlowski P, Figlerowicz M, Zmienko A (2017) MLPA-Based Analysis of Copy Number Variation in Plant Populations. Front Plant Sci 8:222

Sammons RD, Gaines T A (2014) Glyphosate resistance: state of knowledge. Pest Manag Sci. 70:1367–1377

Saxena RK, Edwards D, Varshney RK (2014) Structural variations in plant genomes. Brief Funct Genomics 13:296-307

Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L (2007) Challenges and standards in integrating surveys of structural variation. Nat Genet 39:S7-S15

Schiessl S, Huettel B, Kuehn D, Reinhardt R, Snowdon R (2017) Post-polyploidisation morphotype diversification associates with gene copy number variation. Nature/Scientific Reports 7:41845

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M et al (2004) Large-scalecopy number polymorphism in the human genome. Science 305:525-528

Sieber AN, 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 129:1087–1097

Šķipars V, Krivmane B, Ruņģis D (2011) Thaumatin-like protein gene copy number variation in Scots pine (*Pinus sylvestris*). Environ Exp Biol 9:75-81

Slabaugh MB, Yu JK, Tang S, Heesacker A, Hu X, Lu G, Han F, Bidney D, Knapp SJ (2003) Haplotyping and mapping a large cluster of resistance gene candidates in sunflower using multilocus intron fragment length polymorphisms. Plant Biotech J 1:167-185

Springer NM, Ying K, Fu Y et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/ absence variation (PAV) in genome content. PLoS Genet 5:e1000734

Stankiewicz P, Lupski JR (2010) Structural variation in the human genome and its role in disease. Annu Rev Med 61:437-455

Sutton T, Baumann U, Hayes J, Collins NC, Shi BJ, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M et al (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science 318:1446-1449

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 20:1689-1699

Teo SM, Pawitan Y, Ku CS, Chia KS, Salim A (2012) Statisticalchal lenges associated with detecting copy number variations with next-generation sequencing. Bioinformatics 28:2711-2718

Trappe K, Emde AK, Ehrlich HC, Reinert K (2014) Gustaf: detecting and correctly classifying SVs in the NGS twilight zone. Bioinformatics 30:3484-3490.

Trębicki P, Nancarrow N, Cole E, Bosque-Pérez NA, Constable FE, Freeman AJ, Rodoni B, Yen AL, Luck JE, Fitzgerald GJ (2015) Virus disease in wheat predicted to increase with a changing climate. Global Change Biology 21:3511-3519

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population re-sequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat Genet 42:260-263

Vallejos CE, Astua-Monge G, Jones V, Plyler TR, Sakiyama NS, Mackenzie SA (2006) Genetic and Molecular Characterization of the I Locus of *Phaseolus vulgaris*. Genetics 172: 1229-1242

Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. Nat Rev Genet 18;13:565-75.

Wang D, Amornsiripanitch N, Dong X (2006) A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants. PLoS Pathogens 2:e123

Wang H, Nettleton D, Ying K (2014) Copy number variation detection using next generation sequencing read counts. BMC Bioinformatics 15:109

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. Nature Genetics 47: 944-949

Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones RC, Li-vak KJ (2010) Taking qPCR to a higher level: analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. Methods 50:271-276

Wei C, Chen J, Kuang H (2016) Dramatic Number Variation of R Genes in Solanaceae Species Accounted for by a Few R Gene Subfamilies. PLoS One 11:e0148708

Würschum T, Boeven1 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 Genetics 16:96

Würschum T, Longin CFH, Hahn V, Tucker MR, Leiser WL (2017) Copy number variations of CBF genes at the Fr-A2 locus are essential components of winter hardiness in wheat. The Plant Journal 89:764-773

Xi R, Lee S, Park PJ (2012) A survey of copy-number variation detection tools based on high-throughput sequencing data. Curr Protoc Hum Genet Chapter 7: Unit7-19.

Xiao H, Jiang N, Schaffner E, Stockinger EJ, Vander Knaap E (2008). Retro-transposon-mediated gene duplication under lies morphological variation of tomato fruit. Science 319:1527-1530

Xu X, Liu X, Ge S et al (2012) Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat Bio Technol 30:105-111

Yang S, Lia J, Zhanga X, Zhangb Q, Huanga J, Chena JQ, Hartlc DL, Tian D (2013) Rapidly evolving R genes in diverse grass species confer resistance to rice blast disease. Proc Natl Acad Sci 110:18572-18577

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 16:1–20.

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 12:372

Yu P, Wang CH, Xu Q, Feng Y, Yuan XP, Yu HY, Wang YP, Tang SX, Wei XH (2013) Genome-wide copy number variations in *Oryza sativa* L. BMC Genomics 14:649

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, González AJ, Yan Z, Kitto SL, Grusak MA (2011) MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. Genes Dev 25:2540-2553

Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10:451-481

Zhang R, Murat F, Pont C, Langin T, Salse J (2014) Paleo-evolutionary plasticity of plant disease resistance genes. BMC Genomics 15:187

Zhao M, Wang Q, Wang Q, Jia P, Zhao Z (2013) Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 14 (Suppl 11): S1

Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, Liu TF, Jiang S, Ramachandran S, Liu CM, Jing HC (2011) Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). Genome Biol 12:1-14

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 (2017) 18:261

Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q et al (2012) Using ERDS to infer copy number variants in high-coverage genomes. Am J Hum Genet 91:408-421

Żmieńko A, Samelak A, Kozłowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theor Appl Genet 127:1-18

Zmienko A, Samelak-Czajka A, Kozlowski P, Szymanska M, Figlerowicz M (2016) . *Arabidopsis thaliana* population analysis reveals high plasticity of the genomic region spanning MSH2 AT3G18530 and AT3G18535 genes and provides evidence for NAHR-driven recurrent CNV events occurring in this location. BMC Genomics 17:893

Table and Figure Legends

Table 1: Copy number variation detection analysis tools for whole genome sequencing data

Table 2: CNVs identified in relation to disease resistance in different plant species

Figure 1: Some of the types of CNVs. Deletion in segment C, insertion in segment E, duplication in segment C, relocation in segments C and D, inversion in segment A

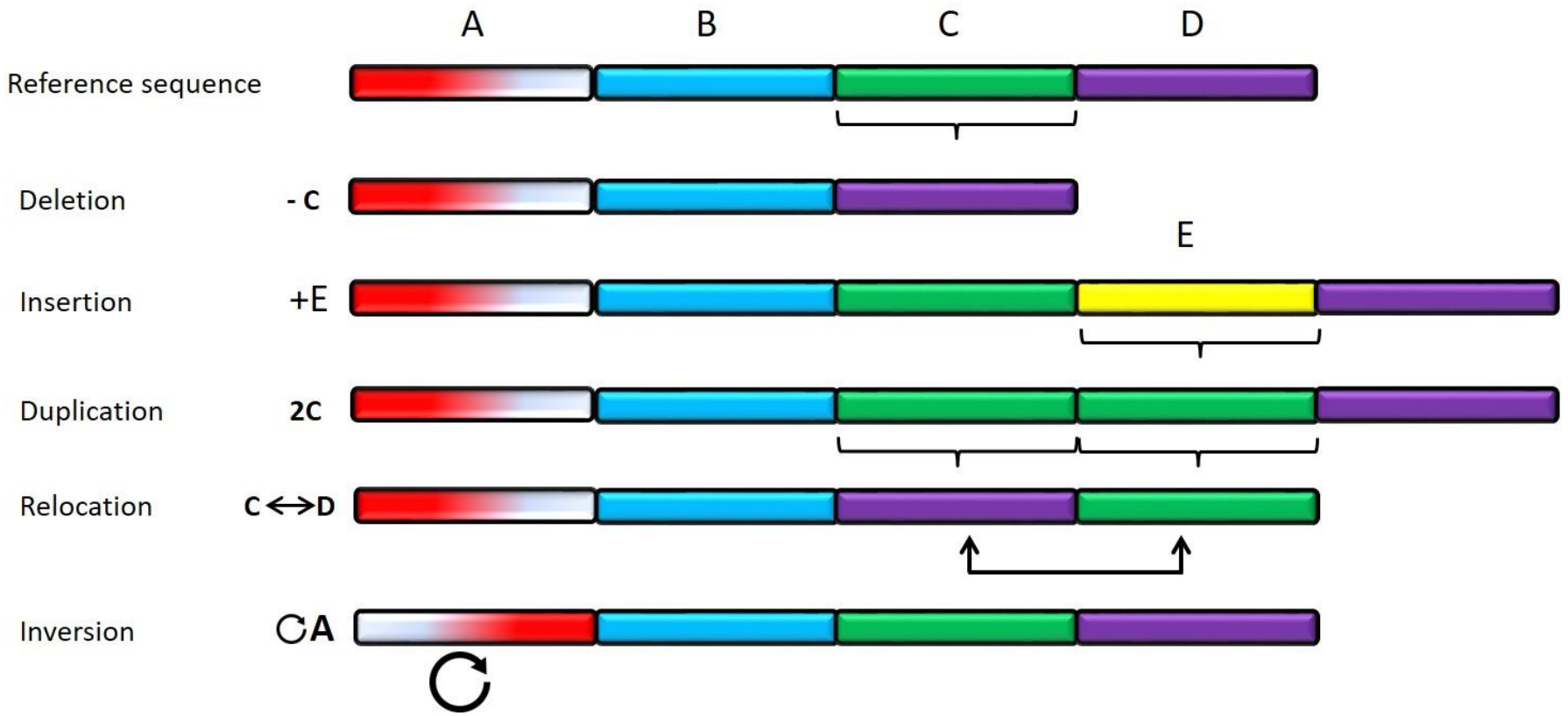

Figure 1 Some of the types of CNVs. Deletion in segment C, insertion in segment E, duplication in segment C, relocation in segments C and D, inversion in segment A

Table 1 Copy number variation detection analysis tools for whole genome sequencing data.

Name	Language	Reference	Availability
RP: Read Paired			
BreakDancer	Perl/C++	Chen et al., (2009)	http://gmt.genome.wustl.edu/packages/breakdancer/
Ulysses	Python/R	Gillet-Markowska et al., (2014)	https://github.com/gillet/ulysses
SR: Spit Read			
PRISM	C	Jiang et al., (2012)	http://compbio.cs.toronto.edu/prism/
Gustaf	C++	Trappe et al., (2014)	http://www.seqan.de/projects/gustaf/
RD: Read Depth			
cm.MOPS	R	Klambauer et al., (2012)	http://www.bioinf.jku.at/software/cnmops/
CNVrd2	R	Nguyen et al. (2014)	http://www.bioconductor.org/packages/release/bioc/html/CNVrd2.html
ERDS	C	Zhu et al., (2012)	http://www.utahresearch.org/mingfuzhu/erds/
AS: Assembly			
Magnolya	Python	Nijkamp et al., (2012)	http://bioinformatics.tudelft.nl/dbl/software
CA: Combined Approach			
Clever-sv	C++	Marschall et al., (2013)	https://code.google.com/p/clever-sv
LUMPY	C++	Layer et al., (2014)	https://github.com/arq5x/lumpy-sv

Table 2 CNVs identified in relation to disease resistance in different plant species

Species	CNVs number	CNVs size	Key findings	Reference
Apple (*Malus domestica*)	876	16.4 kb	Putative CNV regions overlapped 845 gene models and were enriched for *R*-gene models	Boocock et al. 2015
Arabidopsis thaliana	NR	NR	53 genes of NB-LRR genes in the reference genome appeared to be deleted. only four RLP genes appear to be deleted in at least one of the 80 accessions	Guo et al. 2011
Peanut (*Arachis hypogaea*)	NR	NR	Most *Arachis* NBS sequences fall within legume-specific clades, some of which appear to have undergone extensive CNV expansions	Bertioli et al. 2003
Rice (*Oryza sativa*)	1,676	81.2 kb	In relation to disease resistance 1,676 CNVs were identified having more copies than the reference genome.	Xu et al. 2012
	4-8.7	NR	Many rapidly evolving plant *R*-genes in maize, sorghum, *brachypodium*, and rice confer resistance to one or more strains of rice blast disease when present in a rice cultivar genome.	Yang et al. 2013
	2,886	10.28 Mb	The chromosome 11 is enriched with CNV and disease resistance genes	Yu et al. 2013
Barley (*Hordeum vulgare*)	115,003	4.4 kb	The majority of the 'cell death' genes were R-genes encoding nucleotide-binding site leucine-rich repeat (NBS-LRR) protein and affected by CNVs	Muñoz-Amatriaín et al. 2013
Bean (*Phaseolus vulgaris*)	NR	NR	CNV data resulted in a meta-dataset of 51 strong candidate genes with convergent evidence for a role in QR.	Douchkov et al. 2014
	NR	NR	Comparisons among a few TIR–NBS–LRR paralogs within the *I* locus showed variation among them. Increases in CNV of a given sequence lead to increased sequence diversity	Vallejos et al. 2006
	NR	NR	The copy number of *khipu* tandem repeats in relation to disease resistance varies from one *Phaseolus* species to another	David et al. 2009
Cucurbitaceae family	NR	NR	There is low CNV of *R*-genes in Cucurbitaceae. The CNV of LRR-LRK encoding genes is correlated with the number of NBS-LRR encoding genes in different species. The Cucurbitaceae species have not only low copy number but also low diversity of *R*-genes.	Lin et al. 2013
Potato (*Solanum tuberosum*)	NR	NR	Late blight resistance was enhanced as copy numbers and transcript levels of *RB* transgene increased.	Bradeen et al. 2009
Solanaceae family	6,013	NR	The *R*-gene copy number is inconsistent with the number of predicted genes or genome sizes among Solanaceae species. For example, the tetraploid tobacco has the largest genome and the largest number of predicted genes, but has low *R* gene number. CNV in the family Accounted for by a Few *R* Subfamilies.	Wei et al. 2016
Soybean (*Glycine max*)	1-10	NR	The sequence of the individual repeat units, as well as copy number, plays a role in the type specificity of Rhg1- mediated nematode resistance.	Lee et al., 2015
Tomato (*Lycopersicon esculentum*)	NR	NR	Soybean cyst nematode resistance mediated by the soybean quantitative trait locus *Rhg1* is conferred by copy number variation that increases the expression of a set of dissimilar genes in a repeated multigene segment.	Cook et al. 2012
	NR	NR	In striking contrast to the *Cf-9* gene family, six of seven homologs in the *Cf-2*/*Cf-5* gene family vary in LRR copy number, ranging from 25 to 38 LRRs. *Cf-5* and one adjacent homolog differ by only two LRRs.	Dixon et al. 1998
Sunflower (*Helianthus annuus*)	NR	NR	There is a correlation between high HaRGC1 paralogue copy number and functional disease resistance.	Slabaugh et al. 2003
Sorghum (*Sorghum bicolor*)	NR	NR	A key mechanism driving the rapid variation in NBS-encoding genes is the highly dynamic clustering, through lineage specific rearrangements via PAVs and CNVs.	Mace et al. 2014
Rice, Sorghum, Maize, *Brachypodium distachyon*, *Populus trichocarpa*, *Carica papaya*, Soybean, *Lotus japonicas*, *Fragaria vesca*, *Theobroma cacao*	NR	NR	The particular evolution of *R*-genes *via* clusterization was highly dynamic through lineage-specific rearrangements leading to the observed conservation/erosion of *R*-genes collinearity between grasses, referenced as CNV and PAV.	Zhang et al. 2014

NR: Not reported. The number of CNVs detected in different species varies due to difference between genome assemblies technologies