

How high-throughput sequencing empowers the research of polyploidy in vegetable crops

Weixuan Du¹, Xiaoning Wang¹, Xiaokun Zhao¹, Yun Pei^{1,2}, Lei Xia¹, Qinzhen Zhao¹, Chunyan Cheng¹, Yuhui Wang¹, Ji Li¹, Chuntao Qian¹, Qunfeng Lou¹, Rong Zhou³, Carl-Otto Ottosen³, Jinfeng Chen¹ and Xiaqing Yu^{1*}

¹ State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, No.1 Weigang, Nanjing, Jiangsu 210095, China

² College of Agriculture, Guizhou University, Guiyang 550025, China

³ Department of Food Science, Plant, Food & Climate, Aarhus University, Agro Food Park 48, Aarhus N DK-8200, Denmark

* Corresponding author, E-mail: xqyu@njau.edu.cn

Abstract

Vegetables are not only economically important, but also essential for a healthy human diet providing fiber, minerals and essential nutrients. All flowering plants, including many vegetable crops, are polyploids, pinpointing the significance of polyploidy in plant evolution and crop breeding. In the last two decades, the fast development of sequencing has facilitated genome wide investigation of genetic and epigenetic changes that has occurred during the polyploidization process. With the achievement of more and more high-quality plant genomes, ancient polyploidization (also known as whole genome duplication, WGD) events have frequently been seen, which is vital for the understanding of domestication and differentiation history of vegetables. Moreover, advanced joint analysis of multi-omics data has been applied for efficient elucidation of underlying molecular mechanisms of complex traits in vegetables. This paper summarizes the status for the research on vegetable polyploids facilitated by high-throughput sequencing, that improve the understanding of plant evolution and support effective vegetable breeding via polyploidy.

Citation: Du W, Wang X, Zhao X, Pei Y, Xia L, et al. 2024. How high-throughput sequencing empowers the research of polyploidy in vegetable crops. *Vegetable Research* 4: e006 <https://doi.org/10.48130/vegres-0024-0005>

Introduction

Vegetables are one of the basic food sources for human beings, providing essential vitamins, dietary fiber and minerals for human health. Meanwhile, the economic importance of vegetables is increasingly recognized for meeting the diverse dietary needs of a growing population. Polyploids, defined as having three or more sets of chromosomes, often show larger organs and enhanced adaptation to adverse environments. Therefore, polyploidy has been used as an approach for crop breeding, including vegetables. Indeed, polyploids are prevalent in nature, many important crops are polyploids, such as wheat (*Triticum aestivum* L.), cotton (*Gossypium barbadense* L.), rapeseed (*Brassica napus* L.), etc. Polyploid vegetables often possess unique flavors, novel traits, and wider ecological adaptation than their diploid ancestors, providing breeders with greater phenotypic diversity. Comparative transcriptomics analysis demonstrated the importance of polyploidy for the domestication and species, genetic, and physiological diversity in the *Brassicaceae* family^[1,2]. Depending on their genomic composition, polyploids can be classified into autopolyploids, produced by whole genome duplication (WGD) within the same species, and allopolyploids, formed by combining different species. Compared to autopolyploids, allopolyploids are known to be more common and an effective route to hybrid speciation^[3].

Sequencing technology facilitated the study of genomics, epigenomics, transcriptomics and other omics. Sanger dideoxynucleotide sequencing pioneered the research into the new world of parsing genetic code. The publication of the first vegetable genome sequence of cucumber (*Cucumis sativus* L.) in 2009 represents the start of the post-genomic era in vegetable research^[4]. After that, integrative analysis using multi-omics data on model plants and crops enabled the investigation of various biological characteristics in a systematic way. To date, the genomes of many vegetable species have been sequenced, a platform (TVIR, <http://tvir.bio2db.com>) containing 59 vegetable genomes has also been constructed for comparative and functional genomic studies^[5].

The development of plant genomics facilitated by sequencing technology opens a new window for studying the evolutionary history of species, which have revealed the deep roots of polyploidy throughout plant evolution^[6–8]. Research has shown that all green plants have undergone one or more WGD events during the evolution, pinpointing the significance of polyploidy in plant evolution^[9–11]. According to the formation time, polyploids can also be divided into neopolyploids and paleopolyploids^[12]. Ploidy is one of the main drivers of diversity in angiosperms^[13]. WGD also provides adaptive advantages to cells and organisms, especially in unstable, stressful environments, for example, polyploidy enhances plant salt tolerance^[14–16], which improves the natural competitiveness of

polyploid plants, making them more likely to survive. In all, the prevalence of natural polyploids documented their evolutionary success^[17].

Although there have been examples of successful assembling of different copies of genes within the genome of the same organism from short-reading data^[18,19], genomic research in polyploids is still lagging behind due to the short-read length limitations of the first two generations of sequencing and complexity of polyploid genomes. In recent years, the development of long and ultra-long read sequencing provided an unprecedented opportunity for precise genomic analysis in polyploids. For instance, the employment of long read sequencing with Next Generation Sequencing (NGS) as a complement could identify genomic structural variants (SVs) in polyploid crop species like *B. napus*, which used to be challenging to detect using only NGS data^[20].

So far, the genome sequencing and assembly of many polyploid vegetable species have been accomplished, providing vital resources for the research and breeding of vegetables (Fig. 1). For example, the difficulty of assembling the genomes of potatoes is recognized, and recent studies have not only resolved the genome sequences of diploid potatoes^[21], but also completed a high-quality chromosome-scale reference genome sequence of a tetraploid variety^[22]. Whole-genomic analysis of diploid and polyploid *Brassica* species revealed highly conserved genomes since the formation of hexaploid ancestor of *Brassica*^[23]. The effects of whole genome triplication (WGT) on intraspecific diversification of *B. rapa* were reported through pan-genomic analysis using 16 *de novo* assembled genomes and two reported genomes^[24]. This information provides prerequisites for the evolutionary understanding and genetic improvement of vegetables.

The highly complex genome structures of neo-polyploids pose great challenges to genome assembly. Not only are advances in sequencing technologies important for the acquisition of high-quality and long-continuity of polyploid genomes, but also the development of assembly algorithms significantly enhance the precision and reliability of assembly and analysis. The phenotypic advantages of polyploids, especially allopolyploids that merge two divergents, but similar genomes, also add difficulties for polyploid genome sequencing, which further hinder the study of its molecular mechanism^[25–27]. Hi-C, a new technology created by combining chromosome conformation capture with high-throughput sequencing, can locate genome sequences to chromosomes, which is compatible with NGS sequencing and third-generation sequencing (TGS), is increasingly applied for chromosome-level genome assembly^[28,29]. However, in polyploids, the sequence-similar allele fragments can be mistakenly linked together, forming many chimeric assemblies, limiting the assembly accuracy of polyploid genomes^[30]. To solve this problem, many algorithms have been developed for the complex sequence structure of polyploid genomes^[31], such as canu^[32], hifiasm^[33], ALLHiC^[30,34] and PolyGembler^[35]. Of these, ALLHiC is specifically designed to solve the Hi-C assembly puzzle of polyploid species and highly heterozygous genomes by correcting this mismatch of similar fragments in polyploid assembly. For example, the first autopolyploid sugarcane (*Saccharum spontaneum* L.) genome was deciphered by the ALLHiC algorithm. With an integrative strategy of ALLHiC and canu using HiFi and Hi-C sequencing data, accurate and chromosome-level assembly of polyploid genomes is becoming a routine procedure^[36,37]. In this review, the analysis tools used for vegetable genome assemblies were summarized in Table 1.

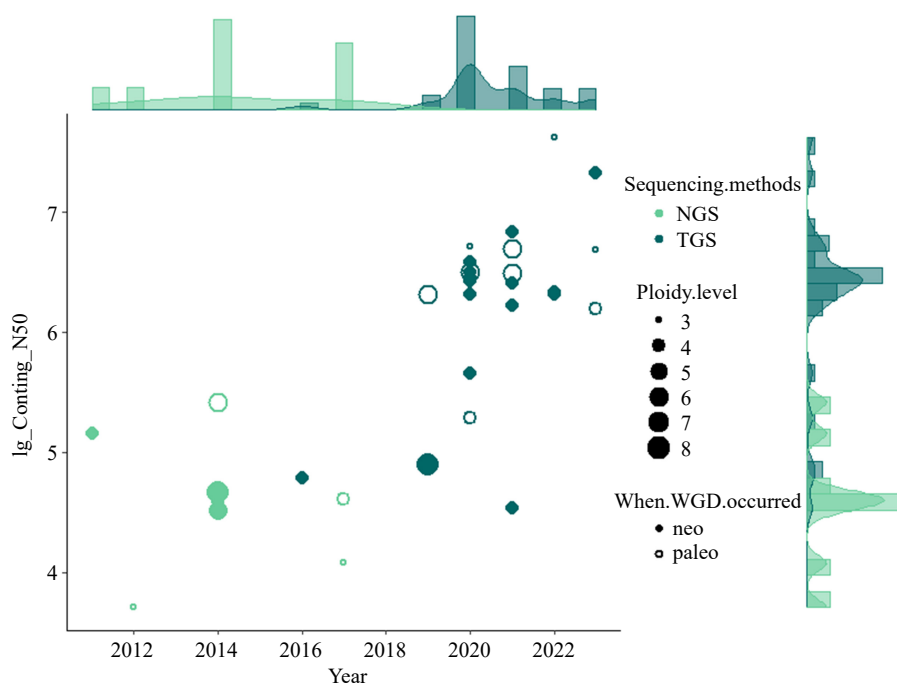


Fig. 1 Overview of currently sequenced polyploid vegetables with the contig N50 (\log_{10}) of assemblies and publication year as Y and X axis, respectively. The type of sequencing technology (Next-generation sequencing, NGS and Third-generation sequencing, TGS) are indicated with different colors. The size of data points is scaled by the ploidy level. Neo- and paleo- polyploids are differentiated with closed and open data points.

Table 1. Analysis tools used for vegetable polyploid genome assembly.

Species	Ref	Analysis tools	Year of publication
<i>Solanum tuberosum</i>	[38]	SOAPdenovo; EMBOSS; SOAPdenovoalign	2011
<i>Solanum lycopersicum</i>	[39]	Newbler; CABOG; BAC FISH; EuGene; ABYSS	2012
<i>Brassica napus</i>	[40]	SOAPdenovo; Newbler; GapCloser;	2014
<i>Fragaria × ananassa</i>	[25]	SOAPdenovo; Newbler; GapCloser;	2014
<i>Camelina sativa</i>	[41]	SOAPdenovo; Bambus; GapCloser; NUCmer	2014
<i>Brassica oleracea</i>	[42]	SOAPdenovo; Bambus; GapCloser;	2014
<i>Brassica juncea var. tumida</i>	[43]	ALLPATHS-LG; PBjelly; IrysView	2016
<i>Cucurbita maxima Duch.</i>	[44]	ShortRead; QuorUM; SOAPdenovo; Newbler; GapCloser; Pilon	2017
<i>Lactuca sativa</i>	[45]	SOAPdenovo; HiRis	2017
<i>Fragaria × ananassa</i>	[46]	NRGene; HiRise; PBjelly; Pilon; CoGe	2019
<i>Brassica oleracea</i>	[47]	FALCON; Quiver; BWA-MEM	2019
<i>Brassica napus</i>	[48]	Falcon; CANU; palign; pilon; 3D-DNA; Juicebox; Juicer	2020
<i>Medicago sativa</i>	[49]	CANU; MECAT; HERA; BWA-MEM; Redundans; Purge Haplotigs; BWA-MEM; Pilon	2020
<i>Medicago sativa</i>	[36]	CANU; ALLHiC; Jcvi; Juicebox	2020
<i>Brassica oleracea</i>	[50]	PBcR wgs8.3rc1; FALCON; Quiver; Pilon; BWA-MEM; SALSA; Mummer	2020
<i>Solanum melongena</i>	[51]	wtdbg2; Racon; Pilon; fragScaff; LACHESIS	2020
<i>Allium sativum</i>	[52]	FALCON; Quiver; Pilon; Purge Haplotigs; FragScaff; Burrows-Wheeler Aligner; ALLHiC	2020
<i>Brassica napus</i>	[53]	CANU; Pilon; 3D-DNA; ALLMAPS; ALLHiC	2021
<i>Brassica juncea</i>	[54]	CANU; BWA mem; MaSuRcA; RefAligner; IrysSolve; Pilon	2021
<i>Brassica juncea</i>	[55]	Jellyfish; FALCON; Sspace-longread; Quiver; BioNano Solve; BioNano Access; Juicebox; HiCPlotter	2021
<i>Cucumis × hytivus</i>	[26]	CANU; Pilon; SOAPdenovo; SSPACE; RefAligner; cutadapt; LACHESIS; HiC-Pro	2021
<i>Brassica oleracea</i>	[56]	Falcon; Arrow; PBjelly	2021
<i>Solanum tuberosum</i>	[22]	pbccs; hifiasm; BWA-MEM; ALLHiC; jcvi	2022
<i>Solanum tuberosum</i>	[57]	hifiasm; pilon;	2022
<i>Capsicum annum</i>	[58]	MECAT2; Pilon; CANU; Juicer; Juicerbox; 3D-DNA; Phred; CA59	2022
<i>Raphanobrassica</i>	[59]	JELLYFISH; CANU; Arrow; Pilon; JUICER;	2023
<i>Vicia faba</i>	[60]	hifiasm; Sniffles; findGSE; purge_haplotigs; Merqury; RagTag; GSAIalign; Liftoff	2023
<i>Lactuca sativa</i>	[61]	CANU; WTDBG2; Arrow; Pilon; 3D-DNA; Juicer; Falcon	2023

Driven by the abundance of data in the genomics era, researchers have an increasing number of research options (e.g., gene editing and multiple omics), and the study of polyploid evolution has recently made significant progress. This review summarizes the research progress in vegetable polyploids driven by sequencing technology and the subsequent studies underpinning important traits and genes, which will further promote germplasm innovation and breeding utilization via polyploidy in vegetables (Fig. 2).

Genomic sequencing facilitates the identification of ancient evolutionary events of vegetables

Whole genome sequencing methods, including survey sequence analysis of RNA or DNA whole genome assembly and so on, provide data for the inference of ancient WGD events in plant evolutionary history^[62]. High-quality chromosome-scale genome assembly enables orthologous genes identification and revealed conserved homologous regions in the homology studies of *Cucumis* diploid and polyploid^[26]. These are essential evidence for understanding the evolution and domestication of crops.

Through the analysis of high-quality genomes that have been sequenced and assembled, genome-wide comparative genomic analysis of *Cucurbitaceae* species identified an overlooked ancient *Cucurbit*-Common Tetraploidization (CCT) event shortly after the *Core-Eudicotium*-Ordinary Hexaploid (ECH), which might contribute to the origination and divergence of *Cucurbitaceae* plants. But by analysis of their sequences, *Cucurbitaceae* crops also largely retain the genetic information of

their diploid progenitor cells, a strong genetic stability after paleopolyploidy that is not common in studied cases of polyploid evolution.^[44,63] This feature provides advantages for studying the evolution of *Cucurbitaceae*. Genomic analysis of Chinese cabbage (*B. rapa* ssp. *pekinensis*) revealed whole-genome triplication (WGT) events during evolution and domestication. Compared with *Arabidopsis thaliana*, this WGT and subgenome dominance increased the total number of genes of Chinese cabbage by about three times, and promoted the expansion of multiple morphotypes^[64]. Ancestors of *Allium sativum* have lived through two WGD events except for the ancient WGD shared by monocotyledons, suggesting that WGD promotes genome expansion and the proliferation of transposable elements (TEs) in *A. sativum*^[52]. However, WGD is certainly not the only factor that affects genome size. No evidence of recent WGD were found in faba bean (*Vicia faba*), many genes were duplicated in tandem, which may contribute to its giant genome^[60]. The study of stem lettuce (*Lactuca sativa* var. *angustana*) also showed that the expansion of gene families of *ADS*, *TAD* and *KAS* was due to both WGD and tandem duplication^[61].

Extensive research has shown that allopolyploidy triggers genome-wide gene structural and expressional changes in the first few generations due to the merger of more than two divergent genomes within a single nucleus^[65], followed by gradually reduced genetic changes with the process of diploidization. This eventually contributed to the successfully establishment as a new species by returning to stable reproductive process and adapted to new environmental niches^[66]. It is found that diploidization and fertility recovery are faster in synthetic allopolyploids than naturally occurring allopolyploids^[26],

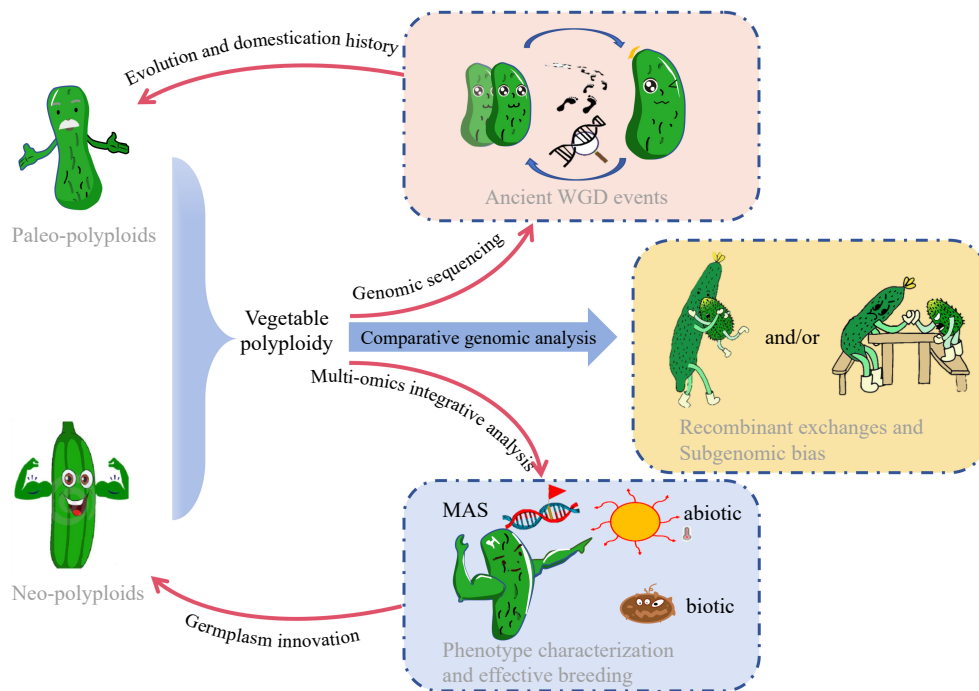


Fig. 2 Schematic diagram illustrating the application of high-throughput sequencing for evolutionary genomics, promoting germplasm innovation and effective breeding in neo- and paleo- polyploid vegetables with cucumber as an example.

suggesting that breeding *via* polyploidy is an efficient and feasible way to improve crops.

During the process of introduction, the divergent environments may drive the changes of polyploids to adapt local environment, which could increase the genetic diversity. For example, the genome of rapeseed (*B. napus*) differentiated rapidly through regional hybridization and intensive selection after the introduction from Europe to Asia, largely expanded the gene pool^[67].

Genomic analysis identifies recombinant exchanges shaping polyploid genome structure and gene expression

DNA sequence recombinant exchanges often occurs shortly after polyploidy in plants, which has been studied in many polyploids^[68–71]. The rearrangement may have led to gene loss such as the loss of *Pll* in the ancestors of *Compositae*^[61]. Neo-allopolyploids are challenged to maintain meiotic stability for being able to reproduce after merging two or more divergent genomes. However, there is a possibility of crossover between chromosomes from different ancestors, known as inter-homoeologue crossover, which will lead to aneuploidy gamete production, hampering normal meiotic pairing. Therefore, the key to maintain stable allopolyploid meiosis is to prevent the formation of inter-homoeologue crossovers^[72]. Studies have shown that changes in *MSH4* expression may have specific effects on homoeologous crossover, which suggests that *MSH4* proteins may be 'major players' in allopolyploid meiotic adaptation^[69].

Homoeologs are a pair of genes or chromosomes from one ancestor which separated until the allopolyploidy event brought it together again^[73]. Homoeologous rearrangements between homoeologous chromosomes have different levels

and degrees of influence on genotype, genetic stability, and seed yield of offspring after meiosis according to parental inheritance and positive and reverse crossing^[74]. For example, the tendency of homoeologous recombination in *B. napus* is mainly associated with TEs in the diploid ancestors^[75]. This phenomenon produces new transcripts that can further result in new protein variants and thus contribute to the evolution of allopolyploid plant species^[68], which occurred frequently in the process of polyploidy. This exchange played an important role in the evolution and diversity of *B. napus*^[70]. Exchanges of large chromosomal segments among homoeologous chromosomes may affect the expression of many genes, but many homoeologous exchanges are tolerated with hardly any changes. However, for those dose-sensitive genes, biased expression changes may occur to favor the re-establishment under certain selection pressures^[71]. The identification and manipulation of related genes to control changes between homoeologous chromosomes is receiving increased attention, which is crucial for polyploidy research and utilization. The development of sequencing technology provides a feasible approach to detect the changes and its regulating genes in polyploids. It has been demonstrated that recombination-related genes like *Fanconi Anemia Complementation Group M* (FANCM) is related to meiotic stability, so that homoeologous chromosome recombination, chromosomal deletion and secondary meiosis can be regulated *via* manipulation of FANCM^[59]. By quantification of homoeologous recombination levels in *B. napus*, three quantitative trait loci (QTLs) were identified for the contribution of the control of homoeologous exchanges, with a major QTL on chromosome A9, *BnaPh1*, contributing between 32% and 58% to the observed variation^[76]. Integrative strategy including sequencing technology can accelerate plant breeding *via* allopolyploidization with extensive genetic exchanges and creation of new genes.

Comparative genomic analyses reveal different fates of subgenomes in polyploids

During the process of hybridization and whole genome replication, drastic genome-wide changes happened in the newly formed allopolyploids and their offspring^[77]. These changes can be asymmetric between subgenomes. Subgenomes with less sequence loss and higher gene expression levels are termed dominant, while the other subgenomes are termed submissive, this phenomenon is collectively known as subgenome dominance^[78,79]. Nevertheless, from the perspective of individual genes, the expression pattern can vary, with genes from submissive subgenomes expressing higher than those homoeologous from dominant subgenomes^[80].

By studying genomic data from strawberry and epigenomics data from rapeseeds. Previous studies have shown that the bias in subgenomic gene expression levels is closely related to the genetic background of the subgenome, usually immediately after the first hybridization event^[46,81]. Although the number of genes with this expression bias increases with generation, the difference between biased genes towards parents is narrowing and continues to have profound effects on plant evolution. In *Brassica* polyploids, the subgenomic expression bias of ancestral genomes has a persistent impact on the intraspecific diversity of existing *Brassica* crops^[24,82]. However, in *Cucurbita* species, no significant subgenome dominance were detected^[44], which may be due to the similar genetic background between the ancestral species. Despite such few cases, subgenome dominance were more often seen and played essential roles in contributing to their agricultural value, such as *B. napus*^[83,84], *Fragaria* × *ananassa*^[46] and *Cucumis* × *hytivus*^[26].

The combination of multi-omics provides an effective way to interpretate the nature of subgenome dominance from different angles. In addition to changes in genomic structure and expression, subgenome dominance is associated with many epigenetic changes, including DNA methylation, small RNA (sRNA) regulation and chromatin modifications. Changes of sRNA ensures the initial genomic stability of neo-allopolyploid formation by controlling for factors such as various noncoding elements and stress-related gene expression^[85]. Previous studies of synthetic allopolyploid wheat have found that small interfering RNAs and chromatin modifications have different effects on different subgenomes, which may be closely related to differences in expression between subgenomic genes^[86].

It has been hypothesized that the combination of different epigenetic states (such as histone modifications and DNA methylation levels) and the distribution of TEs resulted in the imbalance in subgenome populations^[46,80,87,88]. For example, the combination of different degrees of DNA methylation, the different histone modifications, and different TE content between subgenomes in rapeseed could cause subgenomic imbalances^[88]. Alger & Edger hypothesized that difference in gene expression levels between subgenomes may be because of methylation of the TEs rather than the TEs themselves^[79]. However, this hypothesis is challenged by recent research in synthetic oilseed rape, they found that the difference in TE dose between allopolyploid subgenomes is not sufficient to cause subgenome bias expression^[84]. Histone modifications could also contribute to the formation of subgenome dominance^[82,88]. In addition, different tissues and cells may also have different subgenome bias. For example, one subgenome

of allopolyploid blueberries has higher gene expression at almost all studied tissues and developmental stages, but the other subgenome has higher gene expression during fruit development^[89]. While the reasons for the formation of subgenome dominance are complex, it is well acknowledged that this phenomenon plays a very important role in polyploid evolution and requires further study.

Integrative analysis of multi-omics data enables efficient crop breeding

Polyploidy is a classical and efficient crop breeding approach, which can produce novel characters within a very short time. Indeed, many crops have been domesticated from these polyploid ancestors, suggesting the significant role of polyploidy. Polyploids often show better quality and adaptative characteristics than diploids, but their genomes are also larger and more complex, with a larger proportion of redundant genes and regulators, making the association of mutant alleles in polyploid crops difficult to identify. Most polyploid crops have recessive mutations, which seriously hinders the use of forward genetics, therefore, reverse genetics are more promising in the study of polyploid species^[90,91]. In any case, an accurate and contiguous reference assembly provided by the cutting-age sequencing and assembling techniques is important for the study of polyploid crops.

Based on the improvement of throughput and read length of sequencing technology, breakthroughs have been made in marker assisted selection (MAS) of polyploid crops such as *Brassicaceae*^[92] and *Solanaceae*^[21]. Through single-cell transcriptome sequencing, the responses of different cell types of cabbage to high temperature were uncovered, and the cell-specific photosynthetic response and differential genes were revealed^[64,93]. Along with the high-quality chromosome-scale assembly of octaploid strawberries (*F. × ananassa*), metabolomic analysis showed that dominant subgenomes largely control several biological pathways associated with important agronomic traits^[46]. Genome-wide association study (GWAS) using single nucleotide polymorphisms (SNPs) generated from resequencing data of a specific population has been successfully applied to detect quantitative trait loci in many crop species, including vegetables^[94]. Causal variants were also identified in polyploid species by using modified strategy of GWAS methods like single-dose SNPs^[95]. Polyploid tailored software like GWASpoly were also introduced and used for GWAS in polyploid sugarcane^[96] and potato^[97]. Moreover, multi-omics datasets were also integrated with GWAS to uncover trait-determining genes in allopolyploid oilseed rape (*B. napus* L.)^[98] and octoploid strawberry^[99]. The newly developed tools or integrative strategy of multi-omics pave the way for effectively elucidating the underlying mechanism of various phenotypes in polyploid vegetables, and thus can be extremely useful for crop germplasm enhancement.

Conclusions and prospects

Polyploid vegetable crops are widely grown due to their unique traits for higher yields and better qualities. With the utilization of advanced sequencing technology, sequences of many complex vegetable genomes, especially polyploids, have been successfully characterized. It shows that the contribution

of polyploidy to the rapid expansion of angiosperms, known as 'Darwin's abominable mystery'. More importantly, as polyploidy is one of the main approaches for germplasm innovation, these outputs provide fundamentally important resources for crop enhancement. The use of sequencing and related methodologies to reveal its related molecular mechanisms is of specific significance for the better use of polyploidy for vegetable crop improvement. This is especially the case for vegetable crops where the vegetative parts are consumed, thus, organ magnification due to polyploidy bring direct economic benefits.

Despite those progresses achieved by sequencing technology, the availability of multi-omics data and more efficient analysis tools, there are still many problems remaining. Examples are a comprehensive understanding of polyploidization, the significance of subgenome dominance, the genetic determinants of larger organs or not in polyploids, and so on. Polyploid vegetables belong to different families, however, in-depth basic research has been carried out in only few species and results are often contradictory. Another possible reason is that many vegetables are regional, which explain why tomato, cucumber and *Brassica* crops have in focus. Thus, the research on minor vegetables are way behind major polyploid crops like wheat and cotton, despite the continuous lower costs of sequencing. Research work on 'famous' vegetable crops species are often easier to publish in more influential journals that prefer a wider potential audience, while the work on minor vegetable crops receive less attention and less funding for doing fundamental research, which may affect the motivation and confidence of the corresponding researchers.

Against the background of continuous global population growth, climate change, and deteriorating resource environment, using polyploidy to increase crop diversity and enhance environmental adaptability is essential in crop breeding to secure global food security and food quality. Furthermore, health problems caused by high calorie and high sugar diets in the past makes people pay more attention to the need for healthy dietary fiber, pinpointing the importance for vegetable breeding. The complex genomes and uncertainty of phenotypic changes of polyploids are the main limitations that hamper the development of polyploid breeding. Advanced technologies like ultra-long read and single-cell sequencing, which are essential for polyploid research, are still unaffordable for wide application in vegetable research. We call for more attention and support for research in non-model vegetable crops, which is fundamentally important for diversification in vegetable germplasm and a healthy diet.

Previous evidence has shown the successful application of multi-omics in polyploid crop research. Lower and lower sequencing costs will providing more cost-effective and valuable resources that will provide us with deeper understanding of polyploidy, such as the molecular mechanism of enhanced adaptation to adverse environments. The growing amount of sequencing data puts forward higher requirements for data processing and efficient utilization. How to screen out those key factors from vast amounts of data will be the next question that needs to be addressed. Recent progress on variant prediction using Artificial Intelligence may be a promising direction for polyploidy-based molecular design crop breeding^[100]. Meanwhile, even if candidate genes were sifted out, how to verify their function remains a challenge for most non-model

vegetable crops. The establishment and optimization of genetic transformation systems in a broader range of vegetables, especially polyploids species, deserves attention with significant resources to secure future proofed crops.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Yu X; data collection: Du W; draft manuscript preparation: Du W, Wang X, Zhao X, Pei Y, Xia L, Zhao Q, Cheng C, Wang Y, Li J, Qian C, Lou Q, Zhou R, Chen J, Yu X; manuscript revision: Du W, Yu X, Ottosen CO. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (2023YFF1002000), the National Natural Science Foundation of China (32372697) and the Natural Science Foundation of Jiangsu Province (BK2022148).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 29 August 2023; Revised 6 December 2023; Accepted 9 January 2024; Published online 5 February 2024

References

1. Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, et al. 2014. Polyploid evolution of the *Brassicaceae* during the Cenozoic era. *The Plant Cell* 26:2777–91
2. Qi X, An H, Hall TE, Di C, Blischak PD, et al. 2021. Genes derived from ancient polyploidy have higher genetic diversity and are associated with domestication in *Brassica rapa*. *New Phytologist* 230:372–86
3. Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. 2016. On the relative abundance of autopolyploids and allopolyploids. *New Phytologist* 210:391–98
4. Huang S, Li R, Zhang Z, Li L, Gu X, et al. 2009. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* 41:1275–81
5. Yu T, Ma X, Liu Z, Feng X, Wang Z, et al. 2022. TVIR: a comprehensive vegetable information resource database for comparative and functional genomic studies. *Horticulture Research* 9:uhac213
6. Edger PP, Heide-Fischer HM, Bekaert M, Rota J, Glöckner G, et al. 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences of the United States of America* 112:8362–66
7. Jiao Y, Li J, Tang H, Paterson AH. 2014. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *The Plant Cell* 26:2792–802
8. Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, et al. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biology* 13:R3

Polyploid vegetables sequencing

9. Chanderbali AS, Jin L, Xu Q, Zhang Y, Zhang J, et al. 2022. *Buxus* and *Tetracentron* genomes help resolve *eudicot* genome history. *Nature Communications* 13:643
10. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100
11. Soltis DE, Burleigh JG. 2009. Surviving the K-T mass extinction: new perspectives of polyploidization in angiosperms. *Proceedings of the National Academy of Sciences of the United States of America* 106:5455–56
12. Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ. 2005. Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. *Systematic Biology* 54:441–54
13. Zenil-Ferguson R, Burleigh JG, Freyman WA, Igić B, Mayrose I, et al. 2019. Interaction among ploidy, breeding system and lineage diversification. *New Phytologist* 224:1252–65
14. Liu Y, Yu Y, Sun J, Cao Q, Tang Z, et al. 2019. Root-zone-specific sensitivity of K⁺- and Ca²⁺-permeable channels to H₂O₂ determines ion homeostasis in salinized diploid and hexaploid *Ipomoea trifida*. *Journal of Experimental Botany* 70:1389–405
15. Chao D, Dilkes B, Luo H, Douglas A, Yakubova E, et al. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* 341:658–59
16. Ruiz M, Quiñones A, Martínez-Cuenca MR, Aleza P, Morillon R, et al. 2016. Tetraploidy enhances the ability to exclude chloride from leaves in carrizo citrange seedlings. *Journal of Plant Physiology* 205:1–10
17. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, et al. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America* 106:13875–79
18. Kates HR, Johnson MG, Gardner EM, Zerega NJC, Wickett NJ. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany* 105:404–16
19. Kamneva OK, Syring J, Liston A, Rosenberg NA. 2017. Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. *BMC Evolutionary Biology* 17:180
20. Chawla HS, Lee H, Gabur I, Vollrath P, Tamilselvan-Nattar-Amutha S, et al. 2021. Long-read sequencing reveals widespread intragenic structural variants in a recent allopolyploid crop plant. *Plant Biotechnology Journal* 19:240–50
21. Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, et al. 2020. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nature Genetics* 52:1018–23
22. Wang F, Xia Z, Zou M, Zhao L, Jiang S, et al. 2022. The autotetraploid potato genome provides insights into highly heterozygous species. *Plant Biotechnology Journal* 20:1996–2005
23. He Z, Ji R, Havlickova L, Wang L, Li Y, et al. 2021. Genome structural evolution in *Brassica* crops. *Nature Plants* 7:757–65
24. Cai X, Chang L, Zhang T, Chen H, Zhang L, et al. 2021. Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biology* 22:166
25. Hirakawa H, Shirasawa K, Kosugi S, Tashiro K, Nakayama S, et al. 2014. Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Research* 21:169–81
26. Yu X, Wang P, Li J, Zhao Q, Ji C, et al. 2021. Whole-genome sequence of synthesized allopolyploids in *Cucumis* reveals insights into the genome evolution of allopolyploidization. *Advanced Science* 8:2004222
27. Wang M, Tu L, Yuan D, Zhu D, Shen C, et al. 2019. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nature Genetics* 51:224–29
28. Wang K, Wang J, Zhu C, Yang L, Ren Y, et al. 2021. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* 184:1362–1376.e18
29. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, et al. 2018. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. *Nature* 563:501–07
30. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* 5:833–45
31. Schrunner SD, Mari RS, Ebler J, Rautiainen M, Seillier L, et al. 2020. Haplotype threading: accurate polyploid phasing from long reads. *Genome Biology* 21:252
32. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27:722–36
33. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170–75
34. Zhang J, Zhang X, Tang H, Zhang Q, Hua X, et al. 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nature Genetics* 50:1565–73
35. Zhou C, Olukolu B, Gemenet DC, Wu S, Gruneberg W, et al. 2020. Assembly of whole-chromosome pseudomolecules for polyploid plant genomes using outbred mapping populations. *Nature Genetics* 52:1256–64
36. Chen H, Zeng Y, Yang Y, Huang L, Tang B, et al. 2020. Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications* 11:2494
37. Wang Y, Yu J, Jiang M, Lei W, Zhang X, et al. 2023. Sequencing and assembly of polyploid genomes. In *Polyploidy*, ed. Van de Peer Y, volume 2545. New York, NY: Humana. pp. 429–58 https://doi.org/10.1007/978-1-0716-2561-3_23
38. Xu X, Pan S, Cheng S, Zhang B, Mu D, et al. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475:189–95
39. Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–41
40. Chalhoub B, Denoeud F, Liu S, Parkin IAP, Tang H, et al. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345:950–53
41. Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, et al. 2014. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nature Communications* 5:3706
42. Parkin IAP, Koh C, Tang H, Robinson SJ, Kagale S, et al. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid *Brassica oleracea*. *Genome Biology* 15:R77
43. Yang J, Liu D, Wang X, Ji C, Cheng F, et al. 2016. The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nature Genetics* 48:1225–32
44. Sun H, Wu S, Zhang G, Jiao C, Guo S, et al. 2017. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes. *Molecular Plant* 10:1293–306
45. Reyes-Chin-Wo S, Wang Z, Yang X, Kozik A, Arikait S, et al. 2017. Genome assembly with *in vitro* proximity ligation data and whole-genome triplication in lettuce. *Nature Communications* 8:14953
46. Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, et al. 2019. Origin and evolution of the octoploid strawberry genome. *Nature Genetics* 51:541–47
47. Sun D, Wang C, Zhang X, Zhang W, Jiang H, et al. 2019. Draft genome sequence of cauliflower (*Brassica oleracea* L. var. *botrytis*)

- provides new insights into the C genome in *Brassica* species. *Horticulture Research* 6:82
48. Song J, Guan Z, Hu J, Guo C, Yang Z, et al. 2020. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nature Plants* 6:34–45
 49. Shen C, Du H, Chen Z, Lu H, Zhu F, et al. 2020. The chromosome-level genome sequence of the autotetraploid alfalfa and resequencing of core germplasms provide genomic resources for alfalfa research. *Molecular Plant* 13:1250–61
 50. Lv H, Wang Y, Han F, Ji J, Fang Z, et al. 2020. A high-quality reference genome for cabbage obtained with SMRT reveals novel genomic features and evolutionary characteristics. *Scientific Reports* 10:12394
 51. Wei Q, Wang J, Wang W, Hu T, Hu H, et al. 2020. A high-quality chromosome-level genome assembly reveals genetics for important traits in eggplant. *Horticulture Research* 7:153
 52. Sun X, Zhu S, Li N, Cheng Y, Zhao J, et al. 2020. A chromosome-level genome assembly of garlic (*Allium sativum*) provides insights into genome evolution and alliin biosynthesis. *Molecular Plant* 13:1328–39
 53. Chen X, Tong C, Zhang X, Song A, Hu M, et al. 2021. A high-quality *Brassica napus* genome reveals expansion of transposable elements, subgenome evolution and disease resistance. *Plant Biotechnology Journal* 19:615–30
 54. Paritosh K, Yadava SK, Singh P, Bhayana L, Mukhopadhyay A, et al. 2021. A chromosome-scale assembly of allotetraploid *Brassica juncea* (AABB) elucidates comparative architecture of the A and B genomes. *Plant Biotechnology Journal* 19:602–14
 55. Kang L, Qian L, Zheng M, Chen L, Chen H, et al. 2021. Genomic insights into the origin, domestication and diversification of *Brassica juncea*. *Nature Genetics* 53:1392–402
 56. Guo N, Wang S, Gao L, Liu Y, Wang X, et al. 2021. Genome sequencing sheds light on the contribution of structural variants to *Brassica oleracea* diversification. *BMC Biology* 19:93
 57. Sun H, Jiao W, Krause K, Campoy JA, Goel M, et al. 2022. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics* 54:342–48
 58. Liao Y, Wang J, Zhu Z, Liu Y, Chen J, et al. 2022. The 3D architecture of the pepper genome and its relationship to function and evolution. *Nature Communications* 13:3479
 59. Zhang X, Zhang S, Liu Z, Zhao W, Zhang X, et al. 2023. Characterization and acceleration of genome shuffling and ploidy reduction in synthetic allopolyploids by genome sequencing and editing. *Nucleic Acids Research* 51:198–217
 60. Jayakodi M, Golicz AA, Kreplak J, Fehete LI, Angra D, et al. 2023. The giant diploid faba genome unlocks variation in a global protein crop. *Nature* 615:652–59
 61. Shen F, Qin Y, Wang R, Huang X, Wang Y, et al. 2023. Comparative genomics reveals a unique nitrogen-carbon balance system in Asteraceae. *Nature Communications* 14:4334
 62. Wong GKS, Soltis DE, Leebens-Mack J, Wickett NJ, Barker MS, et al. 2020. Sequencing and analyzing the transcriptomes of a thousand species across the tree of life for green plants. *Annual Review of Plant Biology* 71:741–65
 63. Wang J, Sun P, Li Y, Liu Y, Yang N, et al. 2018. An overlooked paleotetraploidization in Cucurbitaceae. *Molecular Biology and Evolution* 35:16–26
 64. Sun X, Feng D, Liu M, Qin R, Li Y, et al. 2022. Single-cell transcriptome reveals dominant subgenome expression and transcriptional response to heat stress in Chinese cabbage. *Genome Biology* 23:262
 65. McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801
 66. Li Z, McKibben MTW, Finch GS, Blischak PD, Sutherland BL, et al. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72:387–410
 67. Zou J, Mao L, Qiu J, Wang M, Jia L, et al. 2019. Genome-wide selection footprints and deleterious variations in young Asian allotetraploid rapeseed. *Plant Biotechnology Journal* 17:1998–2010
 68. Zhang Z, Gou X, Xun H, Bian Y, Ma X, et al. 2020. Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proceedings of the National Academy of Sciences of the United States of America* 117:14561–71
 69. Gonzalo A, Lucas MO, Charpentier C, Sandmann G, Lloyd A, et al. 2019. Reducing *MSH4* copy number prevents meiotic crossovers between non-homologous chromosomes in *Brassica napus*. *Nature Communications* 10:2354
 70. Hurgobin B, Golicz AA, Bayer PE, Chan CKK, Tirnaz S, et al. 2018. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnology Journal* 16:1265–74
 71. Lloyd A, Blary A, Charif D, Charpentier C, Tran J, et al. 2018. Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytologist* 217:367–77
 72. Zhang H, Bian Y, Gou X, Zhu B, Xu C, et al. 2013. Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. *Proceedings of the National Academy of Sciences of the United States of America* 110:3447–52
 73. Glover NM, Redestig H, Dessimoz C. 2016. Homoeologs: what are they and how do we infer them? *Trends in Plant Science* 21:609–21
 74. Ferreira de Carvalho J, Stoeckel S, Eber F, Lodé-Taburel M, Gilet MM, et al. 2021. Untangling structural factors driving genome stabilization in nascent *Brassica napus* allopolyploids. *New Phytologist* 230:2072–84
 75. Bayer PE, Scheben A, Golicz AA, Yuan Y, Faure S, et al. 2021. Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids. *Plant Biotechnology Journal* 19:2488–500
 76. Higgins EE, Howell EC, Armstrong SJ, Parkin IAP. 2021. A major quantitative trait locus on chromosome A9, *BnaPh1*, controls homoeologous recombination in *Brassica napus*. *New Phytologist* 229:3281–93
 77. Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, et al. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proceedings of the National Academy of Sciences of the United States of America* 103:5224–29
 78. Bird KA, VanBuren R, Puzey JR, Edger PP. 2018. The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytologist* 220:87–93
 79. Alger EI, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology* 54:108–13
 80. Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, et al. 2017. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell* 29:2150–67
 81. Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, et al. 2021. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist* 230:354–71
 82. Li M, Sun W, Wang F, Wu X, Wang J. 2021. Asymmetric epigenetic modification and homoeolog expression bias in the establishment and evolution of allopolyploid *Brassica napus*. *New Phytologist* 232:898–913
 83. Gao P, Quilichini TD, Yang H, Li Q, Nilsen KT, et al. 2022. Evolutionary divergence in embryo and seed coat development of U's Triangle *Brassica* species illustrated by a spatiotemporal transcriptome atlas. *New Phytologist* 233:30–51

Polyploid vegetables sequencing

84. Zhang K, Zhang L, Cui Y, Yang Y, Wu J, et al. 2023. The lack of negative association between TE load and subgenome dominance in synthesized *Brassica* allotetraploids. *Proceedings of the National Academy of Sciences of the United States of America* 120:e2305208120
85. Martinez Palacios P, Jacquemot MP, Tapie M, Rousselet A, Diop M, et al. 2019. Assessing the response of small RNA populations to allopolyploidy using resynthesized *Brassica napus* allotetraploids. *Molecular Biology and Evolution* 36:709–26
86. Jiao W, Yuan J, Jiang S, Liu Y, Wang L, et al. 2018. Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. *The Plant Journal* 93:828–42
87. Wendel JF, Lisch D, Hu G, Mason AS. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Current Opinion in Genetics & Development* 49:1–7
88. Zhang Q, Guan P, Zhao L, Ma M, Xie L, et al. 2021. Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus*. *Molecular Plant* 14:604–19
89. Colle M, Leisner CP, Wai CM, Ou S, Bird KA, et al. 2019. Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *GigaScience* 8:giz012
90. Sashidhar N, Harloff HJ, Jung C. 2020. Identification of phytic acid mutants in oilseed rape (*Brassica napus*) by large-scale screening of mutant populations through amplicon sequencing. *New Phytologist* 225:2022–34
91. Sashidhar N, Harloff HJ, Potgieter L, Jung C. 2020. Gene editing of three *BnITPK* genes in tetraploid oilseed rape leads to significant reduction of phytic acid in seeds. *Plant Biotechnology Journal* 18:2241–50
92. Yang J, Wang J, Li Z, Li X, He Z, et al. 2021. Genomic signatures of vegetable and oilseed allopolyploid *Brassica juncea* and genetic loci controlling the accumulation of glucosinolates. *Plant Biotechnology Journal* 19:2619–28
93. Guo X, Liang J, Lin R, Zhang L, Zhang Z, et al. 2022. Single-cell transcriptome reveals differentiation between adaxial and abaxial mesophyll cells in *Brassica rapa*. *Plant Biotechnology Journal* 20:2233–35
94. McLeod L, Barchi L, Tumino G, Tripodi P, Salinier J, et al. 2023. Multi-environment association study highlights candidate genes for robust agronomic quantitative trait loci in a novel worldwide *Capsicum* core collection. *The Plant Journal* 116:1508–28
95. Sumitomo K, Shirasawa K, Isobe S, Hirakawa H, Hisamatsu T, et al. 2019. Genome-wide association study overcomes the genome complexity in autohexaploid chrysanthemum and tags SNP markers onto the flower color genes. *Scientific Reports* 9:13947
96. Yang X, Luo Z, Todd J, Sood S, Wang J. 2020. Genome-wide association study of multiple yield traits in a diversity panel of polyploid sugarcane (*Saccharum* spp.). *The Plant Genome* 13:e20006
97. Zhang F, Qu L, Gu Y, Xu Z, Xue H. 2022. Resequencing and genome-wide association studies of autotetraploid potato. *Molecular Horticulture* 2:6
98. Xuan L, Yan T, Lu L, Zhao X, Wu D, et al. 2020. Genome-wide association study reveals new genes involved in leaf trichome formation in polyploid oilseed rape (*Brassica napus* L.). *Plant, Cell & Environment* 43:675–91
99. Fan Z, Tieman DM, Knapp SJ, Zerbe P, Famula R, et al. 2022. A multi-omics framework reveals strawberry flavor genes and their regulatory elements. *New Phytologist* 236:1089–107
100. Wu Y, Li D, Hu Y, Li H, Ramstein GP, et al. 2023. Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell* 186:2313–2328.e15



Copyright: © 2024 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.