

Perspective

Subgenome dominance and its evolutionary implications in crop domestication and breeding

Zheng Wang^{1,2,3,4,†}, Jinghua Yang^{5,†}, Feng Cheng^{6,†}, Peirong Li^{1,2,3,4}, Xiaoyun Xin^{1,2,3,4}, Weihong Wang^{1,2,3,4}, Yangjun Yu^{1,2,3,4}, Deshuang Zhang^{1,2,3,4}, Xiuyun Zhao^{1,2,3,4}, Shuancang Yu^{1,2,3,4}, Fenglan Zhang^{1,2,3,4}, Yang Dong^{7,†,*} and Tongbing Su^{1,2,3,4,†,*}

¹Beijing Vegetable Research Center (BVRC), Beijing Academy of Agriculture and Forestry Science (BAAFS), Beijing 100097, China

²National Engineering Research Center for Vegetables, Beijing 100097, China

³Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (North China), Ministry of Agriculture, Beijing 100097, China

⁴Beijing Key Laboratory of Vegetable Germplasm Improvement, Beijing 100097, China

⁵Laboratory of Germplasm Innovation and Molecular Breeding, Institute of Vegetable Science, Zhejiang University, Hangzhou 310058, China

⁶Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

⁷State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, the Chinese Academy of Sciences, Beijing 100093, China

*Corresponding authors. E-mail: sutongbing@nercv.org; yang.dong@ibcas.ac.cn

†Equal contribution.

Abstract

Polyploidization or whole-genome duplication (WGD) is a well-known speciation and adaptation mechanism in angiosperms, while subgenome dominance is a crucial phenomenon in allopolyploids, established following polyploidization. The dominant subgenomes contribute more to genome evolution and homoeolog expression bias, both of which confer advantages for short-term phenotypic adaptation and long-term domestication. In this review, we firstly summarize the probable mechanistic basis for subgenome dominance, including the effects of genetic [transposon, genetic incompatibility, and homoeologous exchange (HE)], epigenetic (DNA methylation and histone modification), and developmental and environmental factors on this evolutionary process. We then move to *Brassica rapa*, a typical allopolyploid with subgenome dominance. Polyploidization provides the *B. rapa* genome not only with the genomic plasticity for adapting to changeable environments, but also an abundant genetic basis for morphological variation, making it a representative species for subgenome dominance studies. According to the ‘two-step theory’, *B. rapa* experienced genome fractionation twice during WGD, in which most of the genes responding to the environmental cues and phytohormones were over-retained, enhancing subgenome dominance and consequent adaption. More than this, the pangenome of 18 *B. rapa* accessions with different morphotypes recently constructed provides further evidence to reveal the impacts of polyploidization and subgenome dominance on intraspecific diversification in *B. rapa*. Above and beyond the fundamental understanding of WGD and subgenome dominance in *B. rapa* and other plants, however, it remains elusive why subgenome dominance has tissue- and spatiotemporal-specific features and could shuffle between homoeologous regions of different subgenomes by environments in allopolyploids. We lastly propose acceleration of the combined application of resynthesized allopolyploids, omics technology, and genome editing tools to deepen mechanistic investigations of subgenome dominance, both genetic and epigenetic, in a variety of species and environments. We believe that the implications of genomic and genetic basis of a variety of ecologically, evolutionarily, and agriculturally interesting traits coupled with subgenome dominance will be uncovered and aid in making new discoveries and crop breeding.

Polyploidization and genome evolution

The apparent incongruence between haploid nuclear DNA contents (C-value) and organismal complexity, known as the C-value enigma, is prevalent across the eukaryotic tree of life [1, 2]. This paradox is particularly conspicuous in angiosperms, which exhibit a great diversity in genome size, with a 2400-fold difference between the smallest genome (63 Mb; *Genlisea margaretae*) and the largest genome (149 Gb; *Paris japonica*) [3, 4]. One factor responsible for this remarkable feature of genome complexity among angiosperms is recurrent lineage-specific whole-genome duplication (WGD, also referred to as polyploidization) and small-scale genome

duplication events [5–7]. Polyploids commonly arise from accidental merging of unreduced gametes, in which cells or organisms acquire more than two sets of chromosomes [8]. Based on the parental genome status after polyploidization, polyploids are classified as neo- and palaeopolyploids. It is now widely accepted that the extant angiosperms evolved from palaeopolyploid ancestors with the genomic remnants of at least two ancient and independent WGDs [9–12]. In the process of WGD, selective expansion of transposable elements (TEs) contributed to the enormous differentiation of plant genome size [13]. Plant genomes tended to reduce in size as the result of TE loss and diploidization following

Received: 22 February 2022; Accepted: 30 March 2022; Published: 22 April 2022; Corrected and Typeset: 15 July 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Nanjing Agricultural University. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

WGDs due to adaptation to specific ecological niches of the plants [14].

Studies have revealed that all current plant species have evolved from one or more palaeopolyploidizations, which might be associated with dramatic environmental changes. With the increasing availability of plant genome sequences, it is becoming clearer that a wave of polyploidization events apparently took place around the Cretaceous–Palaeogene (K–Pg) boundary, which marks an extinction event probably caused by a meteor strike that occurred 60–70 million years ago (Mya) [7, 15, 16] (Fig. 1). Polyploid plants exhibit increased adaptation to extreme environmental conditions compared with those of their diploid parents, allowing them to better survive these disastrous climates than their diploid progenitors. Alternatively, Freeling hypothesized that polyploids were merely by-products resulting from adaptive selection that occurred during long-term asexual reproduction underground or under water [17]. Therefore, it has been suggested that in plants polyploidization is important because of its close relationship with the diversification of plant species, novel gene functions [18, 19], the domestication of crops, and the formation of vital agronomic traits [18, 20–23].

Polyploidization and evolutionary innovations in angiosperms

The evolutionary origin of flowers and fleshy fruits is considered a key morphological innovation that promoted the explosive radiation and propagation of angiosperms [24]. The developmental specificity of flower organs is determined by the combined action of MADS-box transcription factors, known as the ABC(DE)-model of floral organ identity [25, 26]. An ancient WGD that happened in the common ancestor of angiosperms was proposed to produce an array of genetic components that account for the origin of flowers (Fig. 1) [27]. Notably, lineage-specific duplication in the *AGAMOUS* gene clade resulted in the production of the C- and D-function lineages, which in turn specified the carpel and ovule identities, respectively [28]. It is likely that an ancient WGD, coupled with lineage-specific gene expansion, is tightly associated with floral organ elaboration in angiosperms. The evolutionary success of the sunflower family (Asteraceae) depends mainly on the development of their unique head-like inflorescence, termed the capitulum, which is linked to multiple documented WGDs [29, 30]. Fleshy fruits evolved independently to facilitate long-distance seed dispersal by attracting animals, and hence are considered to be an adaptive character that increased plant success and boosted adaptive radiation [31, 32]. For example, the origin of the key genetic regulators involved in fruit development and ripening in tomato (*Solanum lycopersicum*) was attributable to a WGD event dating from the K–Pg boundary [33]. Other evolutionary innovations, such as the origin of glucosinolates and subsequent structural elaborations in Brassicales, and

the origin of rhizobial nodulations in Papilionoids, are all associated with lineage-specific WGDs [34–36]. These findings revealed that the polyploidizations contributed to gene function innovation, crop domestication, and establishment of important agronomic traits [34–36].

Genome differentiation, subgenome dominance, and developmental mechanisms

Polyploidization is also categorized into allopolyploidization, in which a single nucleus is formed when the genomes of two different species hybridize, and autopolyploidization, representing genome duplication in the same species [11, 12]. In most allopolyploids, the sudden increase in duplicated genes is offset by gene fractionation, in which homoeologous subgenomes reciprocally lose genes or cis-regulatory elements, and in the due course the duplicated genes acquire single-copy status [37]. Dominance is a ubiquitous characteristic of great evolutionary importance [38], while in allopolyploids the phenomenon of subgenome dominance is crucial. Duplicated genes residing in the subgenome showing less fractionation tend to have higher expression levels and contribute more to morphological determination than those in the more fractionated subgenome [39]. Various taxa, such as maize [37], *Brassica* [39], cotton [40, 46], *Arabidopsis* [41, 45], *Tragopogon* [42], grasses [43], and wheat [44], were reported to possess subgenome dominance. In most of the above studies, subgenome dominance was discussed in terms of plant evolution and adaptation.

Researchers believe that mutation or deletion of upregulated gene copies tends to induce reduced fitness, which is likely to promote their retention by polyploids [39]. However, the genetic mechanism that generates and retains subgenome dominance remains obscure. Among these possible explanations, firstly, low TE abundance is believed to be associated with the biased gene expression in many allopolyploid species [47]. Genome stability is affected negatively by TE activation, whereas inactivation of TEs via methylation decreases the chance of transposon blooms but suppresses neighbouring gene expression. Consequently, genes in the TE-dense subgenome are less expressed than those in the TE-sparse subgenome, but not in all cases, such as when homoeologous genes from the recessive subgenome are expressed at a lower level than those from the dominant subgenome [39, 48]. Secondly, it is now generally believed that genetic incompatibility allows the evolution of dominance. Hybridization is the major origin of allopolyploidization; however, hybridization could disrupt complex regulatory networks [49]. Indeed, fitness may be reduced when highly dosage-sensitive constituents of complexes and pathways are disrupted [50]. Subgenome dominance has been reported to be affected by genetic incompatibilities in cells, such as the coordination of various metabolic, signalling, and regulatory networks, and as a result of merging

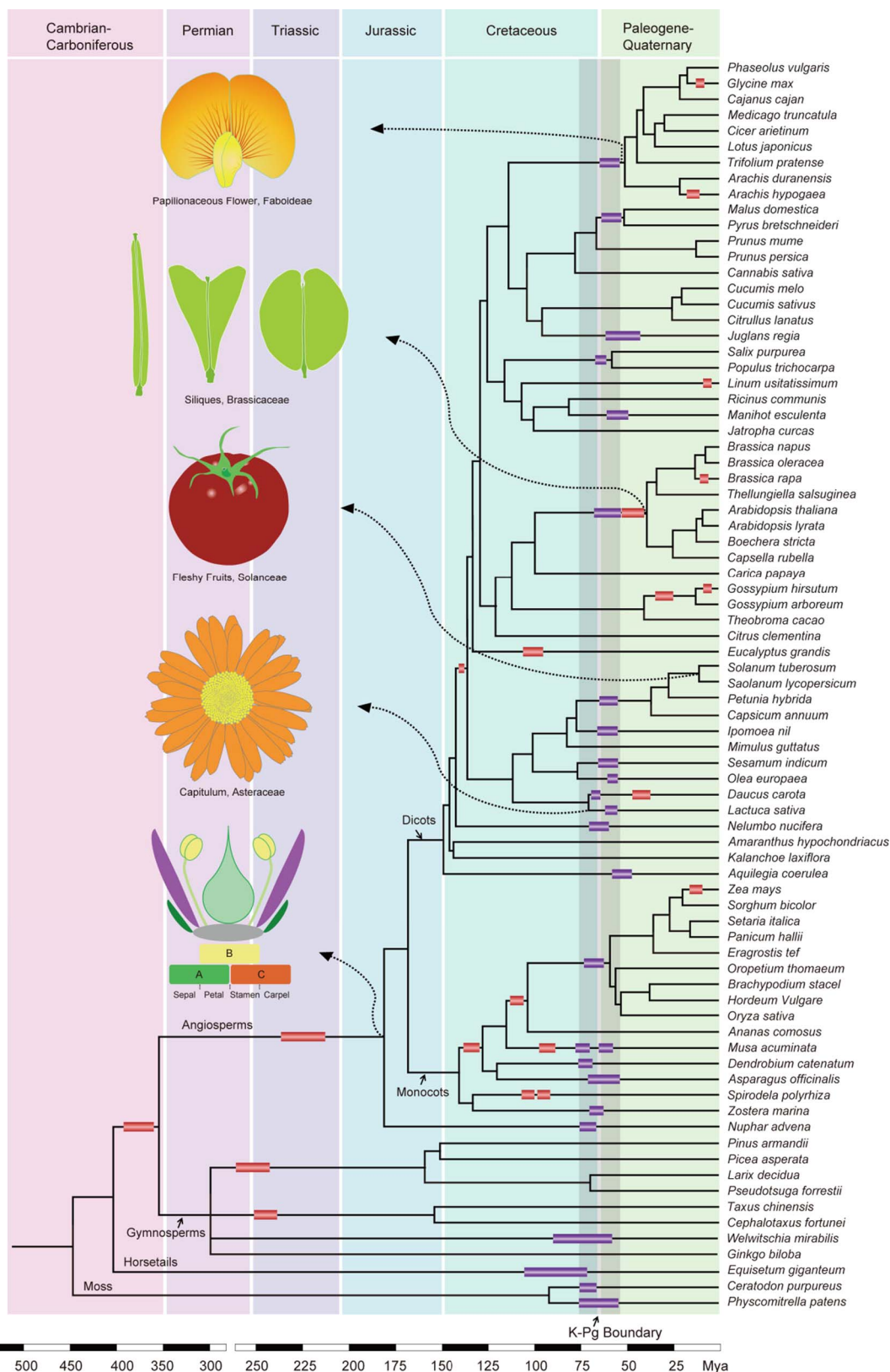


Figure 1. A phylogenetic tree of plants showing the association of WGDs with morphological innovations. A simplified phylogeny displaying the evolutionary relationship between representative plant species. Mapping of WGDs and key morphological innovations on the phylogeny is based on information from published data (Van de Peer et al., 2017; Cheng et al., 2018; [24, 36]). WGDs estimated to be between 55 and 75 million years old are shown in purple rectangular boxes, while others are shown in red rectangular boxes. The K-Pg boundary is indicated by a grey shaded area.

contrasting diploid progenitor species' genomes into a single nucleus. This might result in some pathways being controlled by one subgenome, while other pathways are controlled by the remaining subgenome [51]. This would result in phenotypic traits being partitioned to different subgenomes, such as in blueberry, cotton, and wheat [52, 53]. Thirdly, comparatively little is known about the extent to which chromatin dynamics affect subgenome dominance. In contrast to the stability and heritability of DNA methylation, more dynamism and plasticity are shown by higher-level chromatin modifications, including histone phosphorylation or acetylation. Recent studies of progenitor diploids and allopolyploid cotton revealed that extensive reorganization of domains associated with topology was correlated with alterations to methylation and chromatin status [54]. Gene expression and regulation are affected markedly by changes in DNA methylation and chromatin accessibility, which might be vital to establish subgenome dominance. Additionally, young allopolyploids are reported to undergo homoeologous exchanges (HEs), which can result in alterations to downstream phenotypes, genome-wide methylation patterns, and allele dosage, which might lead to genome stabilization and speciation events [55]. These findings were further supported by a report that the recently developed *Brassica napus* pangenome has highly variable levels of gene presence or absence among various cultivars, resulting from HEs involving genes associated with vital agronomic traits, such as chemical defence, disease resistance, and flowering time. Moreover, HEs display marked subgenome bias [56]. Consequently, more regions from one subgenome are substituted by regions from the other subgenome than the reverse situation, including in synthetic AADD wheat tetraploids, octoploid strawberry, and allopolyploid cotton [56–58]. In addition, HEs affect the expression level of a homoeolog according to the gene copy number [59].

Roles of subgenome dominance in domestication and intraspecific diversification of *Brassica rapa* crops

Kagale et al. identified the polyploidization events and their corresponding times of occurrence in cruciferous species [60]. The α and β WGDs happened about 47 and 124 Mya, respectively, which was before the Brassicaceae family diversified, while another WGT occurred more recently (<23 Mya). Comparisons of the genomes of *B. rapa* and those of other cruciferous plants allowed the reconstruction of the three *B. rapa* subgenomes and the deduction of the diploid ancestral genome ($2n=14$) that was present before the *Brassica* WGT event [61]. The different characteristics of the three *B. rapa* subgenomes prompted the authors to speculate that a tetraploid was formed by the merger of two ancestral genomes, and another hybridization with a third genome happened later. The 21 (3×7) ancestral chromosomes became reshuffled as a result of this two-step hexaploidization

process, which, via re-diploidization, subsequently evolved into the present day *B. rapa* genome comprising 10 chromosomes [62]. The two-step polyploidization provided the *B. rapa* genomes with not only the genomic plasticity to adapt to changing environments, but also an abundant genetic basis for further morphological innovations and variations, which has enabled *B. rapa* to become one of the most diverse species among the angiosperms.

According to the 'two-step theory', the *B. rapa* genome experienced genome fractionation twice after the WGD event, subsequently forming three subgenomes [39]. The two-round genome fractionation resulted in extensive gene loss in the three subgenomes; however, most of genes responding to the environmental cues and phytohormones were over-retained during this process, which further enhanced the adaptation and varied morphotypes of *B. rapa* [62]. *B. rapa*'s close relationship to *Arabidopsis thaliana* offers a good opportunity to study the diversification of the three subgenomes. Further analysis revealed the least fractionated subgenome (LF) retained a higher gene density than the moderately fractionated subgenome (MF1) and the most fractionated subgenome (MF2) [62]. Paralogous genes from the LF subgenome showed a dominant pattern over those from the MF1 and MF2 subgenomes, indicating that the subgenome dominance was caused by the biased fractionation after the WGDs [39]. In addition, the dominant expression status of pairwise syntenic paralogue genes was stable in different organs and different *B. rapa* varieties, and this dominant pattern correlated negatively with the biased distribution of TEs, which is consistent with the aforementioned pattern. However, TEs are not the only factors involved in the emergence of the dominant subgenome of *B. rapa*. Interestingly, 24-nucleotide small RNAs preferentially targeted to the TEs of the MF subgenomes subsequently resulted in TE methylation, which inhibited the expression of downstream genes [63].

Quantitative trait locus (QTL) mapping and *de novo* assemblies of different subspecies provide evidence and a genetic basis for the evolution and diversification of *B. rapa* crops [62, 64–68]. However, unique populations and few reference genomes have limited the identification of the structure variation and copy number variation of *B. rapa* and cannot resolve the genetic diversity of such a diverse species. Recently, a high-quality graph-based pangenome was constructed, providing important genetic information and shedding new light on *B. rapa* evolution and domestication [69]. The 18 genome accessions used in the *de novo* assembly represent most of the morphotypes of *B. rapa*, including the turnip, heading Chinese cabbage, non-heading pak choi, oilseed, sarsons, broccolieto (keto), and mizuna. Compared with the reference Chiifu (a heading Chinese cabbage) genome, the other 17 genomes appeared to show that 15.14–37.39% of the sequences of each genome had no synteny with the reference genome, indicating extensive genomic variation among different subspecies, which

is in line with the fact that different *B. rapa* subspecies have distinct morphological characteristics [70]. More importantly, Cai *et al.* [69] offered a novel insight into the evolution and intraspecific diversification of *B. rapa*, which goes further than the two-step theory. In the process of diversification, *B. rapa* developed into highly diverse morphotypes, which provide a powerful reference to investigate the effects of subgenome dominance on intraspecific diversification (Fig. 2). Cai *et al.* provided further evidence to support subgenome dominance at the intraspecific level by defining the conserved syntenic genes (CSGs) and flexible syntenic genes (FSGs) in *B. rapa* crops. There was a significantly lower proportion of FSGs in the LF subgenome than in the MF1 and MF2 subgenomes, indicating that the expansion of LF subgenome dominance might, at least partially, be explain by intraspecific diversification-related gene flexibility. Interestingly, FSGs tend to accumulate more non-synonymous mutations, structure variations, large-effect mutations, and long terminal repeat retrotransposons, and thus could be tightly associated with the morphological diversification and domestication of *B. rapa*. Furthermore, an inferred ancestral genome was constructed to study the fractionation and subgenome dominance during intraspecific diversification. Consistent with the lower gene loss rate observed in the LF subgenome in different *Brassica* crops [62], the LF subgenome retained more genes from the inferred *B. rapa* ancestral pangenome and showed a lower fractionation rate compared with the MF subgenomes. However, the ratio of FSGs was significantly higher in the MF subgenomes; this further increased LF subgenome dominance during intraspecific diversification.

Future perspectives and application of subgenome dominance knowledge in crop breeding

Subgenome dominance does not mean absolute dominance and expression bias. First, studies showing that local regions favouring homoeologs of one subgenome dominate the others are usually observed in many plants, such as wheat, *B. napus* [71] and cotton [72], but the global expression was not biased towards one specific subgenome. In other words, the submissive subgenome still encompasses certain genes that are more highly expressed than the homoeologs in the dominant subgenome. Next, in some cases the homoeologs in the submissive subgenome are more highly expressed in certain spatial and temporal contexts. Take the tetraploid blueberry, for example: while one subgenome has higher expression in nearly all tested tissues and developmental phases, the other subgenome is highly expressed during fruit development [52]. Similar results were observed in allotetraploid cotton [53]. We thus deduced that the submissive subgenome may fit better in a particular developmental stage, and thus selection may aid it to be dominantly expressed in certain spatiotemporal

contexts. Additionally, many polyploids were derived from hybridization, whereas it is reported that ecological environments during hybridization affect the arising of subgenome dominance and lead to polyploid adaptability. We thus assume that the environmental context could influence which subgenome becomes dominant in certain polyploids. On the other hand, it is reported that abiotic stress can adjust expression patterns in synthetic polyploid cotton, while natural polyploids remain largely unchanged [73]. All the above findings revealed that pre-existing differences between parental genomes could influence subgenome bias and further show that, if subgenome dominance can be predicted in certain hybrids and/or polyploid species, it will be possible to decide which will be of great use for breeding.

Further to the above discussion, it is intriguing that the bias in TE density and DNA methylation, as well as chromatin modification and HEs, do not always result in biased subgenome-wide gene expression; therefore, it is still an open question as to how these processes, working individually or together, affect subgenome dominance. Thus, resynthesized and corresponding natural allopolyploids, and their polyploid progenitors, represent a helpful system to investigate subgenome dominance establishment and escalation. Pre-existing differences between parental genomes, in terms of TE density, DNA methylation, and consequent gene expression, affect the dynamics of allopolyploid subgenomes [74, 75]. The tendency in some species for parental differences to be mimicked in subgenome dominance patterns has been named the 'parental legacy' [74]. However, it remains unclear to what extent the development of subgenome expression bias occurs because of pre-existing diploid progenitor features or results from non-recurrent and independent events during the formation of polyploids. On the other hand, we should bear in mind the clear differences between hybrids and polyploids. For example, hybrids have only one set of chromosomes from each parental progenitor, and this heterozygosity makes hybrids unstable; in contrast, polyploid subgenomes comprise two sets of homologous chromosomes from each parent, which are inherited stably in each generation. In addition, Han *et al.* demonstrated that the chromosomes of synthetic cotton and natural cotton tetraploids associate together dependent on their parent of origin [76], which prompted us to hypothesize that chromosomes do not rearrange themselves randomly. Therefore, there is an urgent need to take advantage of sophisticated techniques, such as the Hi-C chromosome conformation capture application for DNA-DNA interaction and single cell RNA-sequencing as backup for cell-specific gene expression, to stimulate post-polyploidy chromosomal interaction and neo-functional annotation analyses. Last but not least, we believe that histone modification and chromatin accessibility profiling could be used in polyploidy lineages to gain a better understanding of their effect on subgenome dominance establishment and evolution.

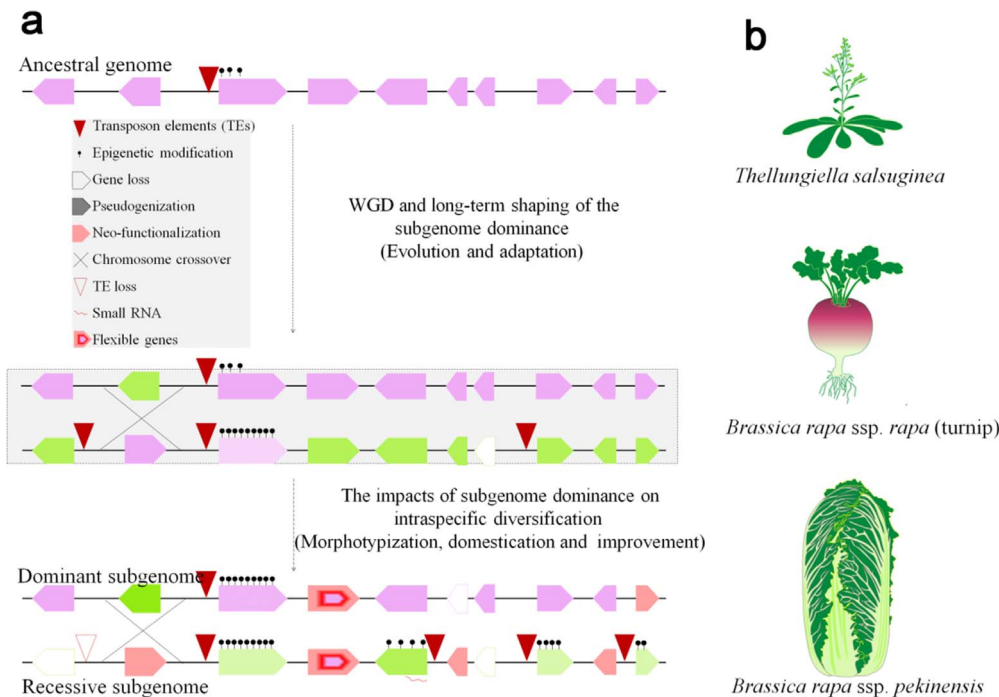


Figure 2. Subgenome dominance and its evolutionary implications in plants. **a** Establishment of subgenome dominance and its evolutionary implications in crop domestication and intraspecific diversification. Subgenome dominance was initially shaped and influenced by TE density, epigenetic modification, HE, and other unexplored factors when two distinct subgenomes merge. (Dotted box) To better understand the effects of subgenome dominance on intraspecific diversification, we assumed an intermediate status during the shaping of subgenome dominance, which actually does not exist during evolution. **b** Polyploidization and subgenome dominance constitute the fundamental driving force for evolution and morphotypization of *B. rapa*.

Studies on the functional divergence of homoeologs upon polyploidization revealed that the biased expression of homoeologs is associated with genome-wide selection, thus implying that transcriptional subgenome dominance enables trait selection [77, 78]. As seen in *B. rapa*, genetic studies and transcription analyses have identified certain QTLs and candidate regulators involved in the morphological variations of different *B. rapa* species [79–84]. As summarized in Table 1, we observed significant impacts of subgenome dominance on *B. rapa* diversification, the formation of morphotypes, domestication, and crop improvement, as indicated by the fact that genes of the LF subgenome were more likely to be selected, either naturally or artificially, during all the above evolutionary events. Besides, in allopolyploid *Brassica juncea*, homoeolog expression dominance has also aided the selection of genes related to lipid and glucosinolate metabolism in subvarieties used for oil production or as vegetables [78]. In addition, in allopolyploid wheat, directional selection had different effects on duplicated homoeologs, which resulted in contrasting variation patterns and inter-variant associations among wheat genomes [77]. Thus, when considering the domestication and improvement of crops, the identification of bias-expressed functional homoeologs is important in order to determine or engineer the genetic basis or checkpoint for interesting traits [85, 86]. These observations on the differential expression of homoeologs suggested that polyploid crop breeding programmes could be improved by focusing on that subset of genes

showing subgenome dominance, to enhance both the response to selection and the acquisition of mechanistic insights. Furthermore, it was revealed that parental legacy, depending on the differences in the expression or epigenetic modifications of gene pairs in the parents or progenitors, resulted in dominance in the remodelling of homoeolog expression bias and asymmetrical epigenetic modifications in *Brassica* and other crops [66]. Therefore, we believe that determining the genetic and epigenetic regulatory mechanisms of the differential expression of homoeologs, as well as the fact that the advantages of heterosis are encompassed by polyploidy-based breeding, will facilitate the *de novo* domestication or improvement of newly synthesized allopolyploids, which could be developed into new crops to strengthen food security. Actually, with recent advances in genetic engineering technology, *de novo* domestication of newly synthesized allopolyploids has attracted the attention of many cutting-edge scientists [87, 88].

In the present perspective, we discuss recent findings regarding the mechanisms underlying subgenome dominance in hybrids and allopolyploids, which have profound implications for evolutionary, ecological, and agricultural research. Therefore, there is a requirement to harness advances in various omics technologies to develop a research pipeline to provide a deeper understanding of polyploidization and subgenome dominance and their mechanistic interdependencies. Moreover, it is now possible to use genome editing tools directly to induce mutation by targeting the key genes involved in

Table 1. Summary of candidate genes related to important agricultural traits in *Brassica rapa*

Trait	Gene	Name	Chromosome	Position	Subgenome	Identification method	Reference	
Heading	BrPIN3.3	BraA07g030650.3C	A07	21870249	LF	Selection sweep	69	
	BrMYB3.3	BraA07g029180.3C	A07	21145989	/	Selection sweep	69	
	BrFL5.1	BraA01g019170.3C	A01	10320303	LF	Selection sweep	69	
	BrSAL4.2	BraA01g025930.3C	A01	15373141	MF1	Selection sweep	69	
	BrARF3.1	BraA04g024390.3C	A04	17723081	LF	Selection sweep	63	
	BrARF4.1	BraA10g018230.3C	A10	13535905	LF	Selection sweep	63	
	BrKAN2.1	BraA09g032840.3C	A09	25471282	LF	Selection sweep	63	
	BrKAN2.3	BraA05g023490.3C	A05	17380156	MF2	Selection sweep	63	
	BrBRX.1	BraA09g033250.3C	A09	25879277	LF	Selection sweep	63	
	BrBRX.2	BraA08g009040.3C	A08	7993035	MF1	Selection sweep	63	
	BrSPL9	BraA05g002720.3C	A05	1497805	LF	Homologous cloning	Wang et al., 2013	
	BrKS1	BraA07g042410.3C	A07	28393881	LF	MutMap analysis	82	
	Tuber for-ma-tion	BrSTP1.1	BraA06g007950.3C	A06	4363461	LF	Selection sweep	63
		BrSTP1.3	BraA09g061400.3C	A09	42842596	MF2	Selection sweep	63
BrEXPB3.2		BraA03g054290.3C	A03	28172489	MF1	Selection sweep	63	
Flowering	BrFR7.1	BraA07g026700.3C	A07	20256386	MF2	QTL; MutMap analysis	84	
	BrFLC1	BraA10g027720.3C	A10	18122666	LF	QTL; domestication	Yuan et al., 2009	
	BrFLC2	BraA02g003340.3C	A02	1616321	MF2	QTL; domestication	Xiao et al., 2016	
	BrFLC5	BraA03g015950.3C	A03	7336775	MF2	QTL	Xi et al., 2018	
	BrVIN3	BraA06g040160.3C	A06	26686911	MF2	QTL; selection sweep	Su et al., 2018	
	BrFT1	BraA02g016700.3C	A02	8897950	MF1	QTL; selection sweep	Su et al., 2018	
	BrCLF1	BraA04g017190.3C	A04	13126173	MF1	EMS; Map-based cloning	Huang et al., 2020	
	BrSDG8	BraA07g040740.3C	A07	27592362	MF1	EMS; Map-based Cloning	Fu et al., 2020	
	Trichome formation	BrpHL1a	BraA06g037290.3C	A06	24917368	LF	GWAS;QTL	Zhang et al., 2018
		BrChlH	BraA03g005840.3C	A03	2557478	MF1	EMS; MutMap analysis	Fu et al., 2019
Leaf colour	Brnym1	BraA03g050600.3C	A03	25985593	MF1	EMS; MutMap analysis	Wang et al., 2020	
	BrCRTISO	BraA09g063710.3C	A09	43923664	MF2	Map-based cloning	Su et al., 2015	
	Brhisn2	BraA05g023920.3C	A05	17771508	MF2	Map-based cloning	Su et al., 2021	
	Brmyb2	BraA07g032100.3C	A07	23201361	LF	Map-based cloning	He et al., 2020	
Disease resistance	BrCRT2	BraA06g006120.3C	A06	3535037	LF	GWAS; map-based cloning	Su et al., 2019	

TE insertion, DNA methylation, histone modification, and HEs associated with subgenome dominance and biased expressions, which will facilitate mechanistic investigations and advanced crop breeding.

Acknowledgements

This work was supported by grants from the Natural Science Foundation of China (32002057, 32172557, 32170227), the Innovation and Capacity-Building Project of BAAFS (KJCX20210427), “Young Talent Award” of Beijing Agricultural and forestry Science and Science Innovation Program (KYCX202001-07).

Author contributions

T.B.S., Y.D., Z.W., J.H.Y., and F.C. wrote the paper. T.B.S. and Y.D. carried out the analyses and drew the figures. P.R.L., X.Y.X., and W.H.W. revised the paper. All authors discussed the results and commented on the manuscript.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

References

- Swift H. The constancy of desoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci USA*. 1950;**36**:643–54.
- Choi I-Y, Kwon E-C, Kim N-S. The C- and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy. *Genes Genomics*. 2020;**42**:699–714.
- Greilhuber J, Borsch T, Muller K et al. Smallest angiosperm genomes found in Lentibulariaceae, with chromosomes of bacterial size. *Plant Biol*. 2006;**8**:770–7.
- Pellicer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? *Bot J Linn Soc*. 2010;**164**:10–5.
- Pellicer J, Hidalgo O, Dodsworth S et al. Genome size diversity and its impact on the evolution of land plants. *Genes*. 2018;**9**:88.
- Barrett CF, McKain MR, Sinn BT et al. Ancient polyploidy and genome evolution in palms. *Genome Biol Evol*. 2019;**11**:1501–11.
- Levin DA, Soltis DE. Factors promoting polyploid persistence and diversification and limiting diploid speciation during the K–Pg interlude. *Curr Opin Plant Biol*. 2018;**42**:1–7.
- Soltis DE, Segovia-Salcedo MC, Jordan-Thaden I et al. Are polyploids really evolutionary dead-ends (again)? A critical reappraisal of Mayrose et al. (2011). *New Phytol*. 2014;**202**:1105–17.

9. Cui L, Wall PK, Leebens-Mack JH *et al.* Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 2006;**16**:738–49.
10. Jiao Y, Wickett NJ, Ayyampalayam S *et al.* Ancestral polyploidy in seed plants and angiosperms. *Nature.* 2011;**473**:97–100.
11. Leitch A, Leitch I. Genomic plasticity and the diversity of polyploid plants. *Science.* 2008;**320**:481–3.
12. Jackson S, Chen ZJ. Genomic and expression plasticity of polyploidy. *Curr Opin Plant Biol.* 2010;**13**:153–9.
13. Lee SI, Kim NS. Transposable elements and genome size variations in plants. *Genomics Inform.* 2014;**12**:87–97.
14. Qiao X, Li Q, Yin H *et al.* Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 2019;**20**:1–23.
15. Fawcett JA, Maere S, Van De Peer Y. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc Natl Acad Sci USA* 2009;**106**:5737–42.
16. Vanneste K, Baele G, Maere S *et al.* Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* 2014;**24**:1334–47.
17. Freeling M. Picking up the ball at the K/Pg boundary: the distribution of ancient polyploidies in the plant phylogenetic tree as a spandrel of asexuality with occasional sex. *Plant Cell.* 2017;**29**:202–6.
18. Renny-Byfield S, Wendel JF. Doubling down on genomes: polyploidy and crop plants. *Am J Bot.* 2014;**101**:1711–25.
19. Otto SP, Whitton J. Polyploid incidence and evolution. *Annu Rev Genet.* 2000;**34**:401–37.
20. Hancock JF. Contributions of domesticated plant studies to our understanding of plant evolution. *Ann Bot.* 2005;**96**:953–63.
21. Renny-Byfield S, Rodgers-Melnick E, Ross-Ibarra J. Gene fractionation and function in the ancient subgenomes of maize. *Mol Biol Evol.* 2017;**34**:1825–32.
22. Zhang K, Wang X, Cheng F. Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic Plant J.* 2019;**5**:231–9.
23. Salman-Minkov A, Sabath N, Mayrose I. Whole-genome duplication as a key factor in crop domestication. *Nature Plants.* 2016;**2**:16115.
24. Dilcher D. Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. *Proc Natl Acad Sci USA* 2000;**97**:7030–6.
25. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. *Nature.* 1991;**353**:31–7.
26. Theissen G, Saedler H. Floral quartets. *Nature.* 2001;**409**:469–71.
27. De Bodt S, Maere S, Van de Peer Y. Genome duplication and the origin of angiosperms. *Trends Ecol Evol.* 2005;**20**:591–7.
28. Kramer EM, Jaramillo MA, Di Stilio VS. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics.* 2004;**166**:1011–23.
29. Barker MS, Kane NC, Matvienko M *et al.* Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol.* 2008;**25**:2445–55.
30. Huang C-H, Zhang C, Liu M *et al.* Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. *Mol Biol Evol.* 2016;**33**:2820–35.
31. Howe HF, Smallwood J. Ecology of seed dispersal. *Annu Rev Ecol Syst.* 1982;**13**:201–28.
32. Seymour GB, Chapman NH, Chew BL *et al.* Regulation of ripening and opportunities for control in tomato and other fruits. *Plant Biotechnol J.* 2013;**11**:269–78.
33. Tomato Genome Consortium. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature.* 2012;**485**:635–41.
34. Young ND, Debelle F, Guerts R *et al.* The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature.* 2011;**480**:520–4.
35. Edger PP, Heide-Fischer HM, Bakaert M *et al.* The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci USA.* 2015;**112**:8362–6.
36. Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol.* 2016;**30**:159–65.
37. Woodhouse MR, Schnable JC, Pedersen BS *et al.* Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homeologs. *PLoS Biol.* 2010;**8**:e1000409.
38. Kacser H, Burns JA. The molecular basis of dominance. *Genetics.* 1981;**97**:639–66.
39. Cheng F *et al.* Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. *PLoS One.* 2012;**7**:e36442.
40. Senchina DS, Alvarez I, Cronn RC *et al.* Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. *Mol Biol Evol.* 2003;**20**:633–43.
41. Wang J, Tian L, Lee HS *et al.* Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics.* 2006;**172**:507–17.
42. Buggs RJ, Chamala S, Wu W *et al.* Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. *Mol Ecol.* 2010;**19**:132–46.
43. Murat F, Zhang R, Guizard S *et al.* Shared subgenome dominance following polyploidization explains grass genome evolutionary plasticity from a seven protochromosome ancestor with 16K protogenes. *Genome Biol Evol.* 2014;**6**:12–33.
44. Pont C, Murat F, Guizard S *et al.* Wheat syntenome unveils new evidences of contrasted evolutionary plasticity between paleo- and neoduplicated subgenomes. *Plant J.* 2013;**76**:1030–44.
45. Akama S, Shimizu-Inatsugi R, Shimizu KK *et al.* Genome-wide quantification of homeolog expression ratio revealed non-stochastic gene regulation in synthetic allopolyploid *Arabidopsis*. *Nucleic Acids Res.* 2014;**42**:e46–6.
46. Renny-Byfield S, Gong L, Gallagher JP *et al.* Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol Biol Evol.* 2015;**32**:1063–71.
47. Alger EI, Edger PP. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr Opin Plant Biol.* 2020;**54**:108–13.
48. Schnable JC, Springer NM, Freeling M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci USA* 2011;**108**:4069–74.
49. Osborn TC, Pires JC, Birchler J *et al.* Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* 2003;**19**:141–7.
50. Birchler JA, Veitia RA. Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc Natl Acad Sci USA* 2012;**109**:14746–53.
51. Bird KA, Van Buren R, Puzey JR *et al.* The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* 2018;**220**:87–93.

52. Colle M, Leisner CP, Wai CM *et al.* Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *GigaScience*. 2019;**8**:giz012.
53. Flagel L, Udall J, Nettleton D *et al.* Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. *BMC Biol*. 2008;**6**:16.
54. Wang M, Wang PC, Lin M *et al.* Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nature Plants*. 2018;**4**:90–7.
55. Chester M, Gallagher JP, Symonds VV *et al.* Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc Natl Acad Sci USA*. 2012;**109**:1176–81.
56. Hurgobin B, Golicz AA, Bayer PE *et al.* Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus*. *Plant Biotechnol J*. 2018;**16**:1265–74.
57. Tennessen JA, Govindarajulu R, Ashman T-L *et al.* Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol Evol*. 2014;**6**:3295–313.
58. Guo H, Wang X, Gundlach H *et al.* Extensive and biased intergenomic nonreciprocal DNA exchanges shaped a nascent polyploid genome, *Gossypium* (cotton). *Genetics*. 2014;**197**:1153–63.
59. Lloyd A, Blary A, Chariff D *et al.* Homoeologous exchanges cause extensive dosage-dependent gene expression changes in an allopolyploid crop. *New Phytol*. 2018;**217**:367–77.
60. Kagale S, Koh C, Nixon J *et al.* The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. *Nat Commun*. 2014;**5**:3706.
61. Cheng F, Mandakova T, Wu J *et al.* Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *Plant Cell*. 2013;**25**:1541–54.
62. Wang X, Wang H, Wang J *et al.* The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet*. 2011;**43**:1035–9.
63. Cheng F, Sun C, Wu J *et al.* Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytol*. 2016;**211**:288–99.
64. Zhao J, Paulo MJ, Jamar D *et al.* Association mapping of leaf traits, flowering time, and phytate content in *Brassica rapa*. *Genome*. 2007;**50**:963–73.
65. Belser C, Istace B, Denis E *et al.* Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*. 2018;**4**:879–87.
66. Li P, Su T, Zhao X *et al.* Assembly of the non-heading pak choi genome and comparison with the genomes of heading Chinese cabbage and the oilseed yellow sarson. *Plant Biotechnol J*. 2021;**19**:966–76.
67. Li Y, Liu GF, Ma LM *et al.* A chromosome-level reference genome of non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*]. *Hortic Res*. 2020;**7**:212.
68. Su T, Wang W, Li P *et al.* A genomic variation map provides insights into the genetic basis of spring Chinese cabbage *Brassica rapa* ssp. *Mol Plant*. 2018;**11**:1360–76.
69. Cai X, Chang L, Zhang T *et al.* Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biol*. 2021;**22**:1–24.
70. Cheng F, Wu J, Wang X. Genome triplication drove the diversification of *Brassica* plants. *Hortic Res*. 2014;**1**:14024.
71. Chalhoub B, Denoeud F, Liu S *et al.* Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014;**345**:950–3.
72. Yoo M, Szadkowski E, Wendel JH. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Mol Plant*. 2013;**110**:171–80.
73. Dong S, Adams KL. Differential contributions to the transcriptome of duplicated genes in response to abiotic stresses in natural and synthetic polyploids. *New Phytol*. 2011;**190**:1045–57.
74. Buggs RJ, Wendel JF, Doyle JJ *et al.* The legacy of diploid progenitors in allopolyploid gene expression patterns. *Phil Trans Royal Soc B*. 2014;**369**:20130354.
75. Kryvokhyzha D, Milesi P, Duan T *et al.* Towards the new normal: transcriptomic convergence and genomic legacy of the two subgenomes of an allopolyploid weed (*Capsella bursa-pastoris*). *PLoS Genet*. 2019;**15**:e1008131.
76. Han J, Zhou B, Shan W *et al.* A and D genomes spatial separation at somatic metaphase in tetraploid cotton: evidence for genomic disposition in a polyploid plant. *Plant J*. 2015;**84**:1167–77.
77. Jordan KW, Wang S, Lun Y *et al.* A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. *Genome Biol*. 2015;**16**:1–18.
78. Yang JH, Liu D, Wang X *et al.* The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. *Nat Genet*. 2016;**48**:1225–32.
79. Ge Y, Ramchiary N, Wang T *et al.* Mapping quantitative trait loci for leaf and heading-related traits in Chinese cabbage *Brassica rapa* L. *Hortic Environ Biotechnol*. 2011;**52**:494–501.
80. Wang Y, Liu X, Ji X *et al.* Identification and validation of a major QTL controlling the presence/absence of leaf lobes in *Brassica rapa* L. *Euphytica*. 2015;**205**:761–71.
81. Sun X, Basnet RK, Yan Z *et al.* Genome-wide transcriptome analysis reveals molecular pathways involved in leafy head formation of Chinese cabbage (*Brassica rapa*). *Hortic Res*. 2019;**6**:130.
82. Gao Y, Lu Y, Li X *et al.* Development and application of SSR markers related to genes involved in leaf adaxial-abaxial polarity establishment in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Front Genet*. 2020;**11**:773.
83. Su T, Wang W, Li P *et al.* Natural variations of BrHISN2 provide a genetic basis for growth-flavour trade-off in different *Brassica rapa* subspecies. *New Phytol*. 2021;**231**:2186–99.
84. Wu Y, Zhang S, Zhang H *et al.* QTL mapping and candidate gene identification of swollen root formation in turnip. *Int J Mol Sci*. 2021;**22**:653.
85. Powell JJ, Fitzgerald TL, Stiller J *et al.* The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. *Plant Biotechnol J*. 2017;**15**:533–43.
86. Lu SJ, Dong L, Fang C *et al.* Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nat Genet*. 2020;**52**:428.
87. Yu H, Lin T, Meng X *et al.* A route to de novo domestication of wild allotetraploid rice. *Cell*. 2021;**184**:1156–1170.e14.
88. Xie Y, Zhang T, Huang X *et al.* A two-in-one breeding strategy boosts rapid utilization of wild species and elite cultivars, *Plant Biotechnol J*. 2022;**20**:800–2.