



Unveiling rapid plant diversification: Exploring the “Fast-Subspeciation Kingdom” of *Brassica rapa*

Chengcheng Cai^a, Sanwen Huang^b, Feng Cheng^{a,*}

^a State Key Laboratory of Vegetable Biobreeding, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops of the Ministry of Agriculture and Rural Affairs, Sino-Dutch Joint Laboratory of Horticultural Genomics, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, 100081, China

^b State Key Laboratory of Tropical Crop Breeding, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518120, China

ARTICLE INFO

Keywords

Rapid diversification
Coevolution
Brassica rapa
Subspecies
Morphotype
Chinese cabbage
Leaf heading

ABSTRACT

The mechanisms driving rapid plant diversification—often referred to as Darwin's “abominable mystery”—remain a central question in evolutionary biology. Elucidating the genomic basis of rapid speciation and morphological innovation, such as those observed in the Brassicaceae, not only addresses fundamental evolutionary questions but also provides valuable insights for crop improvement. Recent advances in telomere-to-telomere genome assembly and graph-based pangenomics have enabled systematic dissection of these processes. Here, we highlight a recent study that uses *Brassica rapa*—a species that has radiated into a ‘kingdom’ of subspecies and vegetable crops within millennia—as a model to decode rapid diversification at the genomic level. By integrating 11 complete genomes with resequencing data from 1720 accessions, the study reveals how centromere remodeling driven by young satellite arrays, widespread structural variants (SVs) linked to subspecies-specific traits, and mutation of a single gene (*BrLH1*), which dominantly controls leafy head formation in Chinese cabbage (*B. rapa* ssp. *pekinensis*), collectively contributed to diversification. The work proposes an integrated SV–gene–centromere coevolution model, offering a new framework for understanding rapid radiation in plants. Beyond evolutionary insights, the study provides a high-resolution pangenome that will accelerate precision breeding in *B. rapa*, demonstrating how cutting-edge genomics can bridge evolutionary research and agricultural innovation.

The evolutionary success and rapid diversification of angiosperms since the mid-Cretaceous represent a fundamental question in evolutionary biology, famously described by Darwin as an “abominable mystery” [1,2]. Since Darwin's time, multiple hypotheses have been proposed to explain the drivers of rapid speciation and morphological innovation, including complex interactions between environmental factors and plant traits and genetic variation [3–5], as well as polyploidization–rediploidization events [6–8]. Despite these advances, the precise genomic mechanisms underlying this mystery remain elusive [1]. Recent developments in gap-free genome assembly and high-quality pangenomics provide unprecedented opportunities to address this question [9]. However, progress has been limited by the lack of high-quality genomic resources spanning diverse morphotypes within rapidly radiating lineages.

The Brassicaceae is a large, globally distributed plant family of great scientific and economic importance, notable for its considerable

morphological diversity [10]. Within this family, the genus *Brassica* includes several agriculturally important species, many of which are widely cultivated as vegetables. During evolution, the ancestor of this genus underwent whole-genome triplication, and the resulting hexaploid ancestor gave rise to diploid species such as *B. rapa* (AA genome) and *B. oleracea* (CC genome). Combinations of the ancestors of these diploids then produced tetraploid species such as *B. napus* (AACC genome), relationships described by the Triangle of U [11], in which the diploids occupy the vertices and the tetraploids arise from pairwise combinations of the diploid ancestors along the sides. Among the *Brassica* species, *B. rapa* in particular has undergone a remarkable diversification within just a few millennia, evolving into multiple subspecies with distinct morphotypes, including leafy greens that form heads (heading Chinese cabbage, ssp. *pekinensis*), leafy greens with thick stems (pak choi, ssp. *chinensis*), tuberous roots (turnips, ssp. *rapa*), and oil-rich seeds (rapeseed, ssp. *oleifera*) [12]. How this ‘kingdom’ of diverse

* Corresponding author.

E-mail address: chengfeng@caas.cn (F. Cheng).

<https://doi.org/10.1016/j.abiote.2026.100049>

Received 24 February 2026; Received in revised form 2 April 2026; Accepted 8 April 2026

Available online 16 April 2026

2662-1738/© 2026 The Authors. Published by Elsevier B.V. on behalf of Agricultural Information Institute of CAAS. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

vegetable forms arose from a common ancestor over such a short evolutionary timescale remains unclear.

In the study that reported the first *B. rapa* reference genome (ssp. *pekinensis*), the authors showed that extensive over-retention of phytohormone-related gene families—duplicated during the *Brassica* hexaploidization event—may have contributed to morphotype diversification [13]. Population-scale variome analysis of 199 core *B. rapa* accessions further revealed parallel selection on homoeologous genes across subgenomes, associated with traits specific to each subspecies [14]. Moreover, a pangenome study of 16 representative *B. rapa* accessions, along with resequencing data from 524 accessions, demonstrated that genomic structural variants (SVs) can be used to track morphotype domestication [15]. Collectively, these variome and pangenome studies identified a set of candidate genes and genomic regions associated with the domestication of leaf-heading and tuberous-root morphotypes, with other studies reporting similar and complementary findings [16–19]. However, these findings have not yet fully elucidated the mechanisms underlying subspeciation in *B. rapa*, thus providing only limited insights into the broader question of rapid angiosperm diversification. Examination of ultra-high-quality genome assemblies and analysis of large sample sets is likely to shed further light on this question.

In a recent study published in *Science* [20], the authors address this knowledge gap by leveraging *B. rapa* as a powerful model system. By integrating 11 high-quality, gap-free *B. rapa* genomes with a large resequencing dataset comprising 1720 *B. rapa* accessions, the study deciphers the genomic architecture underlying rapid subspeciation and diversification in *B. rapa*, providing a framework to revisit the long-standing mystery of angiosperm expansion.

This study presents a comprehensive genomic investigation of the drivers underlying rapid diversification in *B. rapa* (Fig. 1). First, whole-genome resequencing of a diverse panel of accessions, combined with *de novo* assembly of representative genomes from key subspecies, generated a rich genomic resource. Second, the authors constructed the first high-resolution map of *B. rapa* centromeres and discovered novel centromeric satellites, revealing dynamic patterns linked to morphotype differentiation. Third, the authors constructed a graph-based pangenome incorporating SV data and associated specific SVs with important traits such as leaf heading and flowering time. Finally, the authors mapped the *BrLH1* gene and demonstrated its role as a key regulator of leafy head formation. Together, these findings reveal connections among cryptic genomic elements (centromeric satellites), SVs, and key

genes that underlie the rapid diversification of *B. rapa* morphotypes.

The foundation of this research lies in the quality and scale of its genomic resources. The authors generated whole-genome resequencing data from 1720 *B. rapa* accessions and assembled 11 gapless, telomere-to-telomere (T2T) genomes representing all seven major subspecies, including elite breeding lines and landraces with distinct traits. These assemblies were produced using a multi-platform sequencing strategy that integrates PacBio HiFi, Oxford Nanopore ultralong reads, Hi-C and short-read sequencing, enabling complete resolution of telomeres, centromeres, and other complex repetitive regions that were inaccessible in earlier *B. rapa* pangenomes. High-quality annotation of these assemblies identified 6992 genes absent from previously reported genomes. Together with 20 previously published genomes [15,21–23], the authors constructed a saturated, graph-based pangenome that captures species-wide genomic diversity. Supported by population-level data from 1720 accessions, this pangenome provides an accurate representation of population structure and the genomic landscape of *B. rapa*. Continuous generation of additional high-quality genomes and large, integrated variome datasets will further enable the dissection of these previously intractable biological questions.

A particularly notable contribution of this work is the detailed characterization of centromere evolution, long considered as genomic ‘dark matter’. Using BrCENH3 ChIP-seq, the authors mapped 110 functional centromeres across representative *B. rapa* accessions. Their results reveal that centromeres are highly dynamic, with most undergoing extensive repositioning and structural reshaping across subspecies, a process termed centromere remodeling. The discovery of five novel centromeric or pericentromeric satellite arrays (pBrSTRs) and the detailed analysis of their copy number variation—most strikingly, the near absence of three pBrSTRs in Chinese cabbage—provide molecular markers for tracing subspecies differentiation. The authors propose that rapid contraction and expansion of these young, dynamic satellite arrays represent an important force driving centromere evolution and are closely associated with subspeciation. In contrast, the centromere on chromosome A10 is highly conserved across accessions. This centromere harbors genes that are essential for growth and development, suggesting a balance between genomic innovation and the maintenance of core biological functions. The evolutionary trajectory of these satellites—from their absence in diploid species such as *B. nigra* (BB genome) to their proliferation in *B. rapa* (AA genome) and retention in allotetraploids—offers molecular insights into genome evolution within the *Brassica* Triangle of U. Together, these results highlight the rapid

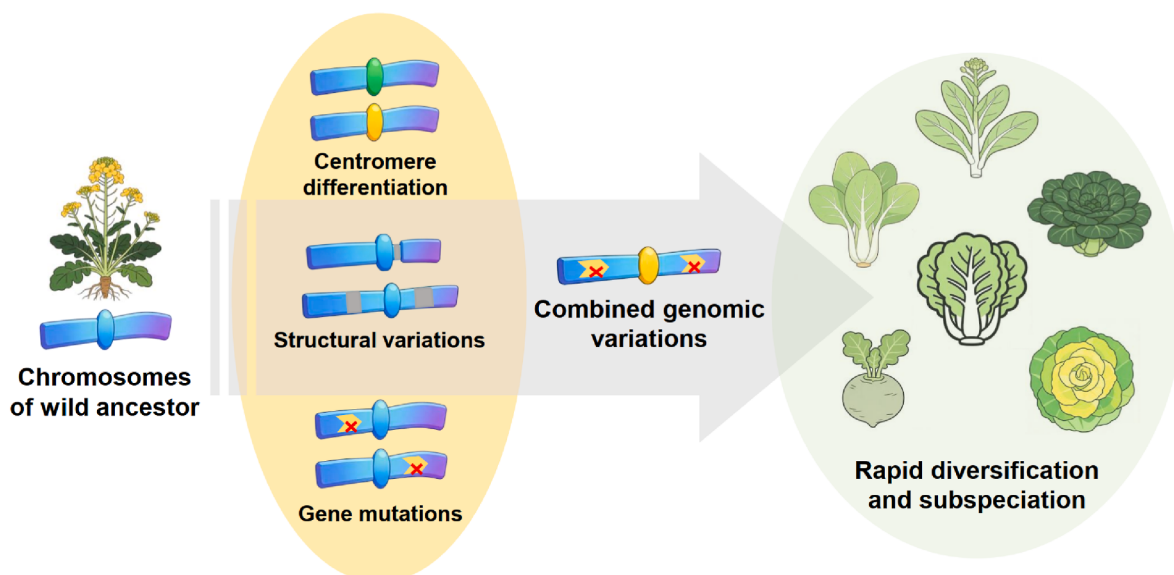


Fig. 1. Coevolution of centromere differentiation, structural variation, and gene mutation drives the rapid diversification of *Brassica rapa*.

differentiation of centromeres during the evolution of *B. rapa* subspecies. However, further studies are needed to investigate their fundamental role in driving rapid speciation.

Beyond centromeres, the study revealed that SVs are central players in subspecies divergence, further highlighting their importance as basic drivers of genetic diversity and evolution [24,25]. By constructing a graph pangenome, the authors identified more than 276,700 SVs, the majority of which are rare and unique to single or a few accessions, consistent with their recent and potentially adaptive origin. The authors further linked specific SVs to the divergence of key agronomic trait, including an ~8.6-kb presence/absence variation (PAV) on chromosome A04 harboring an RZ-1B homolog associated with leafy head formation, and an ~1.1-kb PAV containing *LBD31* linked to the non-vernalization requirement in Caixin (*B. rapa* ssp. *parachinensis*). Moreover, over 40% of PAVs overlap with transposable elements (TEs), reinforcing the role of TEs as a major source of genomic variation [24]. Using the ancestral subspecies *B. rapa* ssp. *rapa* (turnip) as an outgroup, the authors further suggest that derived SVs might have contributed to recent morphotype diversification. Together, these findings support a model in which SVs and subspecies-specific genes provide the primary genetic material for the rapid emergence of distinct morphotypes in *B. rapa*.

Leafy head formation is a key domestication trait of Chinese cabbage. The identification of *BrLH1* as a causal morphogenetic regulator of this trait in *B. rapa* demonstrates the power of these fundamental genomic resources to yield biological insight. Through pangenome-based genome-wide association studies (GWAS) integrating single-nucleotide polymorphisms (SNPs), small insertions/deletions, and PAVs, the authors pinpointed *BrLH1* on chromosome A07 as the major locus controlling leafy head formation. The genetic evidence is compelling: haplotypes defined by three nonsynonymous SNPs in *BrLH1* show complete differentiation and fixation between Chinese cabbage and other morphotypes. This finding is particularly striking, as previous studies had implicated multiple loci and selection signals rather than a single gene with a dominant effect on morphotype formation [14,15,24,26]. Functional validation provides further support: a loss-of-function EMS mutant of this gene abolished head formation, albeit with a concomitant dwarf phenotype. *BrLH1* encodes a protein that contains multiple C2 and transmembrane domains and is homologous to *Arabidopsis* QUIRKY, which interacts with the receptor kinase STRUBBELIG (SUB) to regulate cell patterning [27]. Consistently, the authors demonstrate that *BrLH1* interacts with *BrSUB*, suggesting that a conserved module was co-opted during *B. rapa* domestication to regulate leaf curvature. Together, this work provides direct genetic and mechanistic evidence for a key domestication trait, highlighting the power of integrating high-quality genomes with a large-scale pangenome variome resource in linking genomic variation to complex trait formation.

By integrating gap-free T2T assemblies, repetitive DNA landscapes, and population-level structural variation, this study proposes a unified SV–gene–centromere coevolution model for rapid phenotypic diversification. From evolutionary perspective, *B. rapa* emerges as a tractable system for studying rapid radiation, a process that resembles the early rapid diversification of angiosperms, offering genomic insights into Darwin's "abominable mystery". From applied perspective, the gapless pangenome and pan-SV graph provide powerful references for allele discovery and molecular breeding, with trait-associated SVs and genes such as *BrLH1* offering direct targets for crop improvement.

In conclusion, this work demonstrates that the key to understanding rapid plant diversification lies within the repetitive and structurally variable regions of the genome. By illuminating how centromere remodeling, structural variation, and gene innovation collectively shape the extraordinary diversity of *B. rapa*, this study elucidates how a 'kingdom' of rich vegetable forms could arise from a common ancestor over a short evolutionary timescale. Looking ahead, a promising direction for future research is the systematic comparison of plant lineages exhibiting contrasting levels of diversification as more high-quality

genomic data become available. Such efforts will enable deeper investigation of the underlying mechanisms and the conditions that activate the genomic drivers of species diversification.

CRediT authorship contribution statement

Chengcheng Cai: Writing – review & editing. **Sanwen Huang:** Writing – review & editing. **Feng Cheng:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors have no conflict of interests to declare.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 32402574), the Agricultural Science and Technology Innovation Program (grant no. JCKJ2025-CG-03), and the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture and Rural Affairs, China.

Data availability

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

References

- [1] Buggs RJ. The deepening of Darwin's abominable mystery. *Nat Ecol Evol* 2017;1(6):169.
- [2] Darwin C. More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters. D. Appleton; 1903.
- [3] Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proc Natl Acad Sci* 2004;101(7):1904–9.
- [4] Condamine FL, Silvestro D, Koppelhus EB, Antonelli A. The rise of angiosperms pushed conifers to decline during global cooling. *Proc Natl Acad Sci* 2020;117(46):28867–75.
- [5] Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, et al. Phylogenomics and the rise of the angiosperms. *Nature* 2024;629(8013):843–50.
- [6] Schranz ME, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and diversification: the WGD radiation lag-time model. *Curr Opin Plant Biol* 2012;15(2):147–53.
- [7] Karl R, Koch MA. A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. *Ann Bot* 2013;112(6):983–1001.
- [8] Zhang K, Wang X, Cheng F. Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic Plant J* 2019;5(6):231–9.
- [9] Yan H, Han J, Jin S, Han Z, Si Z, Yan S, et al. Post-polyploidization centromere evolution in cotton. *Nat Genet* 2025;1–10.
- [10] Hendriks KP, Kiefer C, Al-Shehbaz IA, Bailey CD, van Huysduynen AH, Nikolov LA, et al. Global brassicaceae phylogeny based on filtering of 1,000-gene dataset. *Curr Biol* 2023;33(19):4052–68. e6.
- [11] Nagaharu U, Nagaharu N. Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Jpn J Bot* 1935;7(7):389–452.
- [12] Cheng F, Wu J, Wang X. Genome triplication drove the diversification of *Brassica* plants. *Hortic Res* 2014;1:14024.
- [13] Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, et al. The genome of the mesopolyploid crop species *Brassica rapa*. *Nat Genet* 2011;43(10):1035–9.
- [14] Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, et al. Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*. *Nat Genet* 2016;48(10):1218–24.
- [15] Cai X, Chang L, Zhang T, Chen H, Zhang L, Lin R, et al. Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa*. *Genome Biol* 2021;22(1):166.
- [16] Su T, Wang W, Li P, Zhang B, Li P, Xin X, et al. A genomic variation map provides insights into the genetic basis of spring Chinese cabbage (*Brassica rapa* ssp. *pekinensis*) selection. *Mol Plant* 2018;11(11):1360–76.
- [17] Sun X, Basnet RK, Yan Z, Bucher J, Cai C, Zhao J, et al. Genome-wide transcriptome analysis reveals molecular pathways involved in leafy head formation of Chinese cabbage (*Brassica rapa*). *Hortic Res* 2019;6:130.
- [18] Sun X, Gao Y, Lu Y, Zhang X, Luo S, Li X, et al. Genetic analysis of the "head top shape" quality trait of Chinese cabbage and its association with rosette leaf variation. *Hortic Res* 2021;8(1):106.

- [19] Zhao Y, Liu C, Fang B, Huang S, Wang N, Tan C, et al. Identification of a leafy head formation related gene in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Horticulturae* 2022;8(11):1086.
- [20] Ma W, Liu Y, Wei X, Zhang X, Li X, Liu Z, et al. Gapless pangenome analyses reveal fast *Brassica rapa* subspeciation. *Science* 2026;391(6785). eady7590.
- [21] Zhou Y, Ye H, Liu E, Tian J, Song L, Ren Z, et al. The complexity of structural variations in *Brassica rapa* revealed by assembly of two complete T2T genomes. *Sci Bull* 2024;69(15):2346–51.
- [22] Zhang L, Liang J, Chen H, Zhang Z, Wu J, Wang X. A near-complete genome assembly of *Brassica rapa* provides new insights into the evolution of centromeres. *Plant Biotechnol J* 2023;21(5):1022–32.
- [23] Belser C, Istace B, Denis E, Dubarry M, Baurens F-C, Falentin C, et al. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nat Plants* 2018;4(11):879–87.
- [24] Li X, Wang Y, Cai C, Ji J, Han F, Zhang L, et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in *Brassica oleracea*. *Nat Genet* 2024;56(3):517–29.
- [25] Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* 2022;606(7914):527–34.
- [26] Chen R, Chen K, Yao X, Zhang X, Yang Y, Su X, et al. Genomic analyses reveal the stepwise domestication and genetic mechanism of curd biogenesis in cauliflower. *Nat Genet* 2024;56(6):1235–44.
- [27] Trehin C, Schrempp S, Chauvet A, Berne-Dedieu A, Thierry A-M, Faure J-E, et al. QUIRKY interacts with STRUBBELIG and PAL OF QUIRKY to regulate cell growth anisotropy during *Arabidopsis* gynoecium development. *Development* 2013;140(23):4807–17.