

ISSN 1925-2013

<https://genbreedpublisher.com/index.php/pgt>



# Plant Gene and Trait

Vol.17 No.3 2026



*Hydrangea macrophylla* (Thunb.) Ser

OPEN  ACCESS

2026  
03



## Publisher

GenBreed Publisher

## Edited by

Editorial Team of Plant Gene and Trait

Email: [edit@pgt.genbreedpublisher.com](mailto:edit@pgt.genbreedpublisher.com)

Website: <http://genbreedpublisher.com/index.php/pgt>

Address:

11388 Stevenston Hwy,

PO Box 96016,

Richmond, V7A 5J5, British Columbia

Canada

---

Plant Gene and Trait (ISSN 1925-2013) is an open access, peer reviewed journal published online by GenBreed Publisher.

The journal publishes articles that address the fundamental nature of genes and genomes at any level, either experimental or computational approaches, in plants as well as algae, including applications of novel techniques to plant biology and plant trait improvement. All papers chosen for publishing should be innovative research work in fields of plant genes or traits, plant protection, plant breeding, particular in the areas of functional genomics, genomic tools, genome technologies, transgene, genome sequencing analysis, molecular genetics, proteomics, genetic diversity, heterosis, genetic characteristics, genetic modification, genotype-phenotype relationships, stress resistance characteristics, QTL analysis, biochemistry, physiology and morphology.



GenBreed Publisher is an international Open Access publisher specializing in plant protection, plant breeding, molecular genetics, proteomics and genetic diversity registered at the publishing platform that is operated by Sophia Publishing Group (SPG), founded in British Columbia of Canada.

## Open Access

All the articles published in Plant Gene and Trait are Open Access, and are distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



GenBreed Publisher uses CrossCheck service to identify academic plagiarism through the world's leading plagiarism prevention tool, iParadigms, and to protect the original authors' copyrights.



## Latest Content

[A Unified Framework for Causal Inference in Statistical Genetics: Integrating GWAS, Molecular QTL, Colocalization, and Mendelian Randomization](#)

Xuanjun Fang

Plant Gene and Trait, 2026, Vol. 17, No. 3, 156-172

[MADS-Box Transcription Factors in \*Arabidopsis thaliana\*: Molecular Control of Floral Organ Identity and Development](#)

Rimjhim Chandra

Plant Gene and Trait, 2026, Vol. 17, No. 3, 173-181

[Improving Berry Uniformity in Grape \(\*Vitis vinifera\*\): Trait-Based Evaluation and Selection Perspectives](#)

Chunmei Zhu, Yunlong Mao

Plant Gene and Trait, 2026, Vol. 17, No. 3, 182-196

[Yield-Related Traits in Cucumber \(\*Cucumis sativus\*\) and Their Variation and Formation](#)

Wenjing Xu, Mengting Luo

Plant Gene and Trait, 2026, Vol. 17, No. 3, 197-215

[Fruit Quality Traits and Cultivation Practices in Chinese Bayberry \(\*Myrica rubra\*\): Current Status and Research Progress](#)

Yaya Jin


Plant Gene and Trait, 2026, Vol. 17, No. 3, 216-234

## Research Report

## Open Access

**A Unified Framework for Causal Inference in Statistical Genetics: Integrating GWAS, Molecular QTL, Colocalization, and Mendelian Randomization**Xuanjun Fang 

Hainan Provincial Key Laboratory of Crop Molecular Breeding, Hainan Institute of Tropical Agricultural Resources (HITAR), Sanya, 572025, Hainan, China

 Corresponding email: [xuanjunfang@hitar.org](mailto:xuanjunfang@hitar.org)Plant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0011](https://doi.org/10.5376/pgt.2026.17.0011)

Received: 30 Mar., 2026

Accepted: 26 Apr., 2026

Published: 20 May, 2026

**Copyright** © 2026 Fang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**Preferred citation for this article:**Fang X.J., 2026, A unified framework for causal inference in statistical genetics: integrating GWAS, molecular QTL, colocalization, and Mendelian randomization, Plant Gene and Trait, 17(3): 156-172 (doi: [10.5376/pgt.2026.17.0011](https://doi.org/10.5376/pgt.2026.17.0011))

**Abstract** Genome-wide association studies (GWAS) have identified thousands of loci associated with complex traits, yet translating these statistical signals into biological mechanisms remains a major challenge. A key difficulty lies in distinguishing between association, shared genetic architecture, and causal relationships across multiple layers of molecular regulation. In this study, we present a unified analytical framework for causal inference in statistical genetics that integrates GWAS, molecular quantitative trait loci (QTL), transcriptome-wide association studies (TWAS), colocalization analysis, and Mendelian randomization (MR). Within this framework, different methods address distinct inferential targets: GWAS identifies variant–trait associations; molecular QTL and TWAS link genetic variation to intermediate phenotypes; colocalization evaluates the consistency of signals across datasets; and MR estimates the direction and magnitude of potential effects under explicit assumptions. We emphasize that these components should not be interpreted in isolation but as part of a sequential process of evidence refinement. In particular, colocalization is necessary for prioritizing candidate mechanisms but does not establish causality, while MR provides effect estimates that remain sensitive to instrument validity, pleiotropy, and data heterogeneity. We further discuss practical considerations for implementation, including instrument selection, diagnostic evaluation, and cross-population validation, as well as challenges arising from pleiotropy, tissue specificity, and environmental interactions. Finally, we extend this framework to plant systems and emerging multi-omics contexts, highlighting the role of single-cell and epigenomic data in refining causal interpretation. By clarifying the roles and limitations of individual methods within an integrated framework, this study provides a structured approach for moving from genetic associations toward biologically interpretable and experimentally testable hypotheses.

**Keywords** Statistical genetics; Causal inference; Genome-wide association study (GWAS); Molecular QTL (eQTL, sQTL, pQTL); Transcriptome-wide association study (TWAS); Colocalization; Mendelian randomization; Multi-omics integration; Pleiotropy; Complex traits

**1 Introduction**

Genome-wide association studies (GWAS) have generated an unprecedented scale of statistical associations in complex trait genetics. However, these signals fundamentally represent association estimands—statistical relationships between genetic variants and traits—rather than direct evidence of biological causality. This distinction underlies a central gap in the field: many GWAS loci reside in non-coding regions and are influenced by linkage disequilibrium (LD) and multilayer regulatory architectures, such that a single association peak often implicates multiple candidate genes and mechanisms. Even when fine-mapping reduces signals to smaller credible sets, the resulting inference remains a causal probability estimand, rather than a direct estimate of causal direction or effect (Liu et al., 2019; Wainberg et al., 2019; Xie et al., 2021; Mostafavi et al., 2023).

From a unified statistical genetics perspective, the analysis of complex traits can be conceptualized as a multi-layer inferential chain defined by distinct statistical targets (estimands): GWAS characterizes association evidence, fine-mapping quantifies posterior probabilities of causality, and polygenic risk scores (PRS) translate these signals into individual-level predictive functionals. Yet a critical gap remains in this chain: how to move from causal probability to causal pathways and causal effect estimands. This gap defines the role of functional integration and causal inference methods.

To bridge this divide, expression quantitative trait loci (eQTL) and transcriptome-wide association studies (TWAS) introduce molecular phenotypes as intermediate layers, extending inference from the variant level to the gene expression level. In this framework, eQTL analyses characterize the mapping from genotype to gene expression, while TWAS integrates GWAS summary statistics with expression prediction models to generate gene-level association signals and prioritize candidate genes (Wainberg et al., 2019; Xie et al., 2021; Zhao et al., 2022).

However, it is essential to emphasize that, TWAS remains an association-based projection under LD structure, not a causal estimand. Due to co-regulation among nearby genes, LD contamination, tissue mismatch, and genetic confounding, TWAS signals may reflect non-causal variants, leading to the prioritization of “bystander genes” (Liu et al., 2019; Zhao et al., 2022; Tambets et al., 2024). Thus, TWAS provides necessary but insufficient evidence for causality, and its statistical target remains a gene-level reparameterization of the association estimand.

Colocalization analysis provides a critical interface at this stage. Rather than constituting an independent association test, colocalization evaluates whether GWAS and molecular QTL signals share the same underlying causal variant by comparing their posterior distributions. In this sense, colocalization can be understood as an inference on shared causal configuration estimands derived from fine-mapping posterior distributions. This probabilistic framework enables the integration of association and functional evidence across data domains.

Building upon this structure, Mendelian randomization (MR) advances inference from causal probability to causal effect estimands. By using genetic variants as instrumental variables (IVs), MR estimates the causal effect of an exposure (e.g., gene expression or protein level) on an outcome trait, thereby mitigating confounding and reverse causation under three core assumptions: relevance, independence, and exclusion restriction (Hemani et al., 2018; Jiang et al., 2022).

In practice, however, MR is subject to several challenges, including horizontal pleiotropy, complex LD structure, and weak instruments, all of which can introduce bias and inflate false positives (Barfield et al., 2018; Tambets et al., 2024). Moreover, the limited availability of strong and cross-tissue stable cis-eQTL instruments constrains the applicability and reproducibility of MR in causal gene identification (Lu et al., 2024). Recent methodological developments-such as MR-Egger, weighted median estimators, and MR-PRESSO-provide partial robustness to these violations, but their validity critically depends on the structural evidence provided by upstream eQTL/TWAS and colocalization analyses (Hemani et al., 2018; Zhao et al., 2022).

Based on these considerations, we propose a unified causal inference layer in statistical genetics, in which different methods correspond to distinct estimands within a coherent inferential hierarchy:

- GWAS: association estimand
- Fine-mapping: causal probability estimand
- eQTL/TWAS: mediation mapping estimand
- Colocalization: shared causal configuration estimand
- MR: causal effect estimand

Within this framework, complex trait analysis can be formalized as a “functionally integrated causal chain”: GWAS → fine-mapping → eQTL/TWAS → colocalization → MR. This pipeline represents a progressive refinement from statistical association to causal effect estimation, where each layer is defined by its estimand, assumptions, and sources of uncertainty. This estimand-driven perspective can also be extended to multi-trait genetics, where shared genetic architecture is distinguished from causal interpretation through structural, locus-level, and pattern-level inference (Fang, 2026).

In this study, we systematically develop this unified framework by clarifying the statistical foundations, assumptions, and limitations of each component, with particular emphasis on their interfaces and error

propagation across layers. We further propose an operational workflow and reporting standards applicable to multi-ancestry and multi-tissue settings. By reframing functional integration as an estimand-driven inference system rather than a collection of tools, this framework establishes a coherent theoretical foundation for bridging association discovery, causal probability, and mechanistic interpretation in complex trait genetics.

## 2 Integration of eQTL with Functional Phenotypes

In studies of complex traits, eQTL analyses provide an essential intermediate layer that links genetic variation to molecular phenotypes. Unlike GWAS, which captures statistical associations between loci and traits, eQTL focuses on how genetic variants influence gene expression or related molecular traits, thereby offering clues about potential regulatory mechanisms. In this sense, eQTL analyses do not by themselves establish causality, but instead help delineate the pathways through which genetic variants may act.

Within integrative frameworks, eQTL data are typically used in two ways. First, they provide molecular signals that can be compared with GWAS results through colocalization analyses. Second, they serve as a source of candidate instruments for downstream causal inference methods, such as Mendelian randomization. As such, eQTL analyses form a critical bridge between statistical association and functional interpretation.

### 2.1 cis-eQTL and trans-eQTL

eQTL are commonly categorized into cis- and trans-acting variants based on their genomic proximity to the target gene. Cis-eQTL are located near the gene they regulate and typically exert their effects through local regulatory elements such as promoters, enhancers, or untranslated regions. Large-scale datasets across multiple tissues have shown that cis-regulatory effects are widespread and relatively reproducible, with a substantial proportion overlapping GWAS loci (Liu et al., 2019; Wainberg et al., 2019). For this reason, cis-eQTL are often prioritized in integrative analyses as plausible links between genetic variants and gene expression.

In practice, cis-eQTL play two major roles. They can be used in colocalization analyses to assess whether GWAS and expression signals are likely driven by the same underlying variant. They also tend to provide more stable instruments for Mendelian randomization, enabling the evaluation of relationships between gene expression and phenotypic traits. At the same time, interpretation requires caution. Linkage disequilibrium may cause signals to spread across neighboring variants, and allelic heterogeneity can complicate the assignment of effects to specific loci. It is therefore common to combine eQTL results with fine-mapping, allele-specific expression, and functional annotations to strengthen the evidence.

Trans-eQTL, in contrast, affect genes located at a distance and often operate through indirect regulatory mechanisms, such as transcription factors, microRNAs, or chromatin interactions (Kirsten et al., 2015). These effects are typically weaker and more sensitive to cellular composition, environmental influences, and population structure, which makes their detection and interpretation more challenging. Recent work suggests that some trans effects arise through hierarchical regulatory relationships, where a cis-regulated gene acts as an upstream driver influencing downstream targets (Kvamme et al., 2025). This observation has practical implications for analysis strategies. One approach is to first identify candidate cis-regulatory variants, then explore their downstream impact using network-based methods or mediation analyses. In some cases, stepwise Mendelian randomization can be applied to evaluate relationships across multiple layers (Figure 1).

In plant systems, these challenges are often amplified by extended linkage disequilibrium, structural variation, and polyploid genomes. Incorporating multi-parent populations (such as NAM or MAGIC) and pangenome references can help reduce misattribution and improve robustness.

Image caption: Cis-eQTLs typically influence gene expression through proximal regulatory elements, such as promoters or enhancers, and are therefore more readily aligned with GWAS signals in integrative analyses. In contrast, trans-eQTLs act on distal genes through indirect mechanisms involving transcription factors, miRNAs, or chromatin interactions, often reflecting multi-layer regulatory processes and increased sensitivity to cellular and

population context. The figure illustrates the differences in regulatory scope and pathways between these two classes, highlighting the transition from local regulatory effects to broader network-level influences.

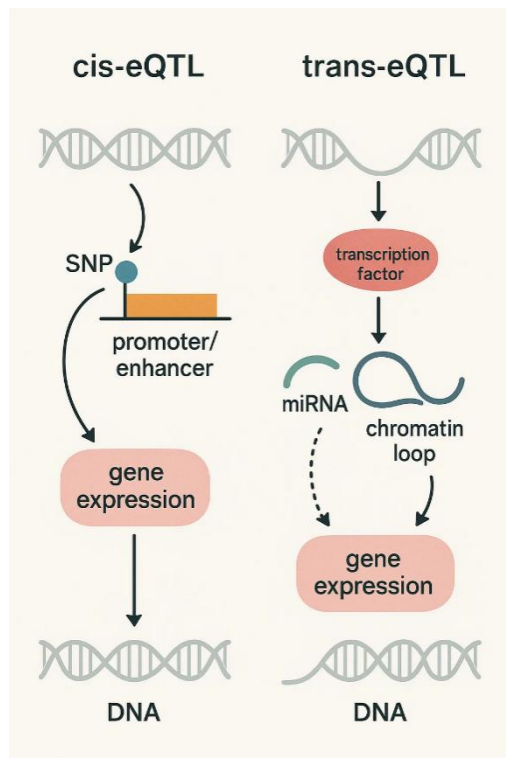


Figure 1 Regulatory patterns of cis- and trans-eQTL and their roles in integrative analyses

## 2.2 Tissue and cell-type specificity

eQTL effects often vary substantially across tissues and developmental stages (Fagny et al., 2017). Some loci exhibit consistent effects across multiple tissues, whereas others are highly tissue-specific, reflecting differences in chromatin states, regulatory programs, and cellular composition. For integrative analyses, prioritizing tissues that are relevant to the trait of interest generally improves interpretability. At the same time, reporting both shared and tissue-specific effects can help distinguish robust signals from context-dependent ones.

Cellular heterogeneity represents an important source of confounding in bulk tissue data. To address this, approaches such as deconvolution, interaction models, and environment-specific analyses can be used to better characterize context-dependent effects (Zhang and Zhao, 2023).

Single-cell eQTL (sc-eQTL) analyses further increase resolution by identifying regulatory effects at the level of individual cell types or states. Studies in immune and brain tissues have revealed a large number of cell-type-specific signals, as well as dynamic changes across conditions (Bryois et al., 2022). From a statistical perspective, pseudo-bulk aggregation and hierarchical modeling are often used to balance resolution and power. These data can provide more precise information about where regulatory effects are likely to act, which is valuable for downstream interpretation.

## 2.3 Extended QTL types: sQTL, pQTL, and meQTL

Beyond expression QTL, additional layers of molecular QTL provide further insight into regulatory mechanisms. Splicing QTL (sQTL) capture variation in transcript structure and may operate independently of total expression levels, allowing regulatory effects to be examined from both quantitative and structural perspectives (Zheng et al., 2020). Analyses at the transcript level or using multi-exposure models can help disentangle these contributions.

Protein QTL (pQTL) measure the effects of genetic variants on protein abundance and are often closer to the functional endpoints of many traits. While some cis-pQTL can serve as strong instruments, their relationship with eQTL is not always straightforward, reflecting additional layers of post-transcriptional regulation.

Methylation QTL (meQTL) describe the influence of genetic variation on DNA methylation patterns. When combined with data on chromatin accessibility and histone modifications, they contribute to a broader view of regulatory architecture. In multi-tissue settings, meQTL signals often only partially overlap with eQTL or GWAS loci, suggesting that different regulatory layers may act through distinct pathways.

In plant systems, particular attention should be given to differences among methylation contexts (CG, CHG, and CHH) and their regulatory mechanisms. Incorporating tissue- or environment-specific analyses can improve the interpretability of these signals.

### 3 TWAS: From GWAS to Gene-Level Associations

Within integrative analyses, transcriptome-wide association studies (TWAS) provide a framework for translating GWAS signals from the variant level to the gene level. By incorporating gene expression as an intermediate phenotype, TWAS enables the evaluation of whether genetically regulated expression is associated with complex traits, thereby offering a structured link between genetic variation and downstream phenotypes.

It is important to note, however, that TWAS does not directly establish causality. Rather, it reorganizes SNP-level association signals into gene-level statistics under a specified expression prediction model and LD structure. In this sense, TWAS is best viewed as a structured projection of GWAS signals, rather than an independent causal inference approach.

#### 3.1 Basic principles

The central idea of TWAS is to use reference datasets that contain both genotype and expression data to train predictive models of gene expression, and then apply these models to GWAS data to assess gene-trait associations (Li and Ritchie, 2021; Evans et al., 2024). For a given gene  $g$ , the predicted expression can be written as:

$$\hat{E}_g = \sum_{j \in L_g} w_{gj} G_j$$

where  $L_g$  typically denotes SNPs within a cis region,  $w_{gj}$  represents weights estimated from reference data (e.g., using Elastic Net, BLUP, or BSLMM), and  $G_j$  denotes SNP dosage. Association is then tested between  $\hat{E}_g$  and the phenotype  $Y$ , with the goal of evaluating whether genetically driven variation in expression is related to the trait (Mai et al., 2023).

When only GWAS summary statistics are available, TWAS can be implemented using LD-based transformations. Let  $Z$  denote the vector of SNP-level GWAS Z-scores and  $R$  the corresponding LD matrix. The gene-level statistic can be approximated as:

$$Z_g \approx \frac{w_g^T Z}{\sqrt{w_g^T R w_g}}$$

The significance of this statistic depends on the effective information carried by the weights under the LD structure. Two practical considerations follow from this formulation. First, the LD reference panel should closely match the target GWAS population. Second, the choice of tissue or cell type used to train the expression model plays a critical role in determining both interpretability and effect direction (Li and Ritchie, 2021).

#### 3.2 Common methods

Most TWAS methods follow a two-step strategy involving model training and downstream association testing, but differ in how expression weights are constructed and how multi-tissue information is incorporated. PrediXcan and its summary-based extension S-PrediXcan use elastic net regression to estimate cis-regulatory weights in reference data, which are then applied to individual-level or summary-level GWAS data. MultiXcan and UTMOST extend this framework by integrating information across tissues to distinguish shared from tissue-specific effects.

FUSION adopts a more flexible approach by integrating multiple prediction models-including BLUP, LASSO, Elastic Net, and BSLMM-within a unified framework, allowing direct computation of gene-level statistics from

summary data while accounting for model uncertainty and LD structure (Evans et al., 2024). This class of methods emphasizes robustness in model selection and statistical inference.

More recent developments include approaches that incorporate nonparametric or Bayesian modeling strategies (e.g., TIGAR and its extensions), as well as methods that integrate multiple priors to improve power (Parrish et al., 2022; Liang et al., 2025). These extensions broaden the applicability of TWAS across diverse data settings.

In practice, different methods reflect distinct priorities. Some emphasize the portability of expression weights across datasets, whereas others focus on model integration and uncertainty control. Regardless of the approach, TWAS results are typically interpreted in conjunction with fine-mapping and colocalization analyses, which help refine gene-level signals into more credible candidate regions (Li and Ritchie, 2021; Mai et al., 2023).

### **3.3 Limitations**

Despite its utility in linking GWAS signals to functional interpretation, TWAS has several important limitations. First, the transferability of expression prediction models is often constrained. The estimated weights depend on the ancestry, LD structure, and tissue context of the reference dataset. When these differ from those of the target GWAS population, prediction accuracy may decline, leading to reduced statistical power and potential bias (Li and Ritchie, 2021; Mai et al., 2023). Multi-tissue approaches and expanded reference resources such as GTEx can mitigate this issue to some extent, but do not fully resolve it.

Second, TWAS results remain fundamentally associative. Because of LD, the weights used to predict expression may capture signals from variants that are correlated with, but not identical to, the true causal variant. In addition, co-regulation and unobserved confounding can cause non-causal genes to appear significantly associated. As a result, interpreting TWAS findings as evidence of causal effects can be misleading (Wainberg et al., 2019; Evans et al., 2024). Simulation and methodological studies have shown that such interpretations may inflate false positive rates if not carefully controlled (Zhu and Zhou, 2020; De Leeuw et al., 2023).

For this reason, TWAS findings are typically evaluated alongside locus-level evidence. Colocalization analyses can be used to assess whether GWAS and expression signals are consistent with a shared underlying variant, while Mendelian randomization can provide additional support for potential causal relationships. This layered approach helps reduce misinterpretation and improves the reliability of downstream inference.

Finally, the scope of TWAS is limited by the availability and coverage of reference datasets. Current eQTL resources provide incomplete representation of rare variants, trans-regulatory effects, and noncoding RNAs, which constrains the comprehensiveness of the models. Future developments are likely to focus on more flexible modeling strategies, expanded multi-ancestry and multi-tissue datasets, and explicit modeling of genotype-by-environment interactions, particularly in plant and multi-environment studies (Parrish et al., 2022; Liang et al., 2025).

## **4 The Role and Limitations of Colocalization Analysis**

In moving from GWAS signals toward functional interpretation, a central question is whether association signals observed in different data sources—such as GWAS and molecular QTL—reflect the same underlying genetic factors within a given genomic region. Colocalization analysis was developed to address this question by evaluating the consistency of signals across datasets and providing a basis for downstream interpretation.

Unlike association analysis within a single dataset, colocalization focuses on the correspondence between signals from different sources. The goal is to assess whether two signals can plausibly be explained by the same underlying variant, given the local LD structure and statistical uncertainty. In practice, this step serves to refine candidate regions, narrowing the focus from general associations to loci that are more likely to support coherent biological interpretation.

### **4.1 Statistical framework of colocalization**

Widely used approaches such as COLOC adopt a Bayesian framework to compare a set of mutually exclusive hypotheses, including no association, association in only one dataset, independent associations in both datasets,

and a shared underlying variant. Posterior probabilities are assigned to each scenario, with PPH4 commonly used as a summary measure of support for the shared-variant hypothesis (Zuber et al., 2022).

As genomic regions often contain multiple independent signals, the assumption of a single causal variant is frequently violated. To address this, more recent methods incorporate multi-signal modeling and integrate colocalization with fine-mapping, such as coloc combined with SuSiE or FINEMAP, as well as approaches like eCAVIAR and fastENLOC (Foley et al., 2021; Wallace, 2021). These developments improve performance in regions with complex signal structures and allow colocalization to be more closely aligned with locus-level inference.

In practice, reliable colocalization requires careful data harmonization. This includes aligning allele orientation, standardizing effect sizes and standard errors, and using LD reference panels that match the ancestry of the GWAS dataset. Because multiple independent signals may be present within a region, it is often advisable to perform conditional analysis or fine-mapping prior to colocalization, or to apply models that explicitly account for multiple signals. In multi-tissue settings, analyses can be conducted separately for each tissue and then integrated with functional annotations to form a more complete interpretation (Wallace, 2021; Zuber et al., 2022).

Although thresholds such as  $PPH4 > 0.8$  are often used in practice, their interpretation depends on model assumptions, prior choices, and data quality. Sensitivity analyses and conditional results should therefore be considered when evaluating the robustness of findings (Figure 2) (Rasooly et al., 2022).

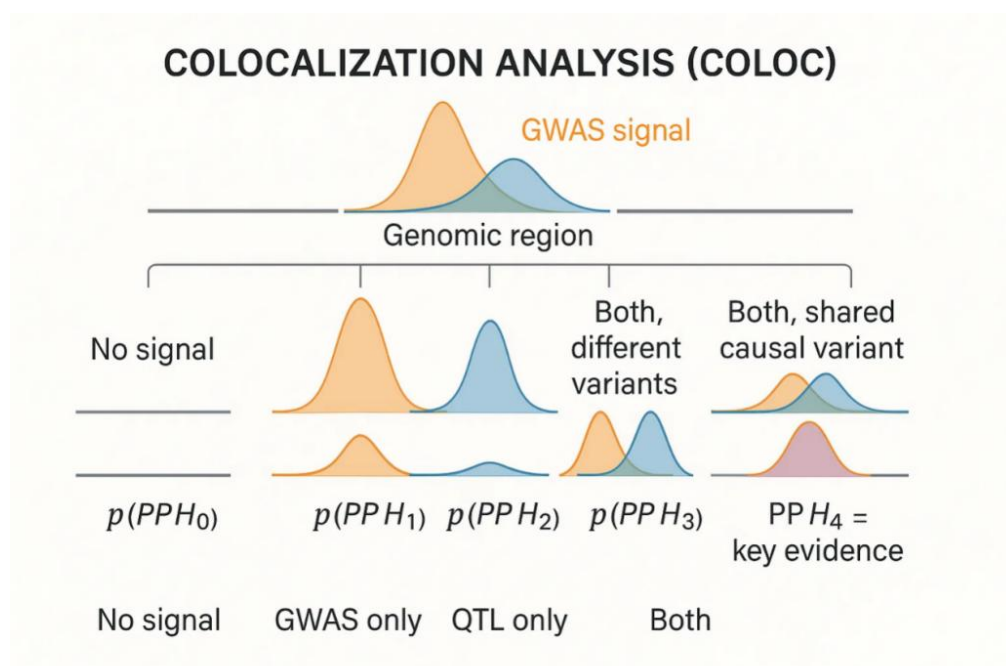


Figure 2 Statistical decision framework of colocalization analysis and its role in cross-dataset integration

Image caption: Colocalization analysis evaluates whether association signals from GWAS and molecular QTL (e.g., eQTL) within the same genomic region are consistent with a shared underlying genetic variant. Under a Bayesian framework, methods such as COLOC compare five mutually exclusive scenarios, including no signal, GWAS-only, QTL-only, independent signals, and shared signals, thereby quantifying alternative explanations for the observed patterns. The figure illustrates these scenarios and the corresponding signal configurations, with PPH4 representing the posterior support for the shared-variant hypothesis. It should be noted that PPH4 reflects statistical evidence for signal concordance rather than direct inference of causal mechanisms or mediation

## 4.2 Interpreting colocalization results

While colocalization is a useful tool for prioritizing candidate loci, its results should not be interpreted as direct evidence of causal relationships. Even when the posterior probability for a shared signal is high, this only indicates that two associations are consistent with the same underlying variant; it does not establish how that variant influences the trait.

One common scenario is horizontal pleiotropy, in which a single variant affects multiple traits through independent pathways. In such cases, colocalization may still detect a shared signal, even though no direct mediation relationship exists between the molecular phenotype and the complex trait (Rasooly et al., 2022). Evidence from epigenomic studies further suggests that many disease-associated loci involve complex regulatory architectures with multiple parallel pathways (Shikov et al., 2020; Boix et al., 2021; Khan et al., 2024).

LD structure and allelic heterogeneity can also complicate interpretation, particularly in regions with multiple signals where the true causal variant may be obscured by correlated variants (Wallace, 2021). For these reasons, colocalization is best viewed as a filtering step rather than a definitive test.

In analytical workflows, loci with strong colocalization support are often prioritized for further evaluation using complementary approaches. For example, Mendelian randomization can be applied to assess the direction and magnitude of potential effects, whereas loci with weak or inconsistent evidence may require re-examination at the level of fine-mapping or data harmonization before further interpretation.

#### **4.3 Applications in plant systems**

In plant systems, colocalization analysis has proven useful for disentangling tissue-specific and environment-dependent regulatory effects. Studies across multiple tissues and environmental conditions have shown that the effects of regulatory variants can vary substantially depending on developmental stage or external stimuli, leading to context-dependent contributions to complex traits.

For instance, analyses in crops and model plants have identified distinct regulatory patterns across organs such as leaves, roots, and fruits, as well as across developmental stages. Work in tomato and other species has further demonstrated that conserved developmental genes may exhibit pleiotropic effects, while still showing variation in regulatory behavior across environments or species (Hendelman et al., 2021).

In practice, it is often beneficial to construct eQTL maps in trait-relevant tissues and under representative environmental conditions, followed by stratified colocalization analyses across these contexts. The use of multi-parent populations, such as NAM or MAGIC, can improve resolution by reducing LD and helping to distinguish multiple signals. In polyploid crops, additional attention is required to differentiate homologous gene copies and to accurately quantify expression, in order to minimize ambiguity in gene assignment.

Candidate loci identified through colocalization can then be further evaluated using downstream approaches, including Mendelian randomization, near-isogenic lines, genome editing, and expression assays, ultimately contributing to a more complete characterization of the relationship between genetic variation and phenotypic traits.

### **5 Mendelian Randomization as A Framework for Causal Inference**

Within integrative analysis pipelines, Mendelian randomization (MR) is typically applied at a later stage, once genetic associations and molecular evidence have been established. Its primary purpose is to evaluate the direction and magnitude of potential effects by using genetic variants as instruments and treating molecular traits-such as gene expression-as exposures.

Compared with earlier steps, which focus on mapping or aligning signals across datasets, MR aims to quantify relationships under a set of assumptions. As a result, the interpretation of MR findings depends critically on how instruments are selected and on the plausibility of the underlying assumptions.

#### **5.1 Core assumptions of instrumental variables**

MR analyses rely on three basic conditions. First, the selected genetic variants must be sufficiently associated with the exposure, ensuring that they carry informative signal. Second, these variants should be independent of confounding factors, an assumption that is partly justified by the approximate random allocation of alleles at conception. Third, the effect of the instruments on the outcome should operate primarily through the exposure of interest, rather than through alternative pathways.

In practice, these conditions are not directly testable and may be violated in subtle ways. For example, a variant may influence multiple biological processes, giving rise to additional pathways that complicate interpretation. For this reason, instrument selection is often informed by upstream analyses. Colocalization, in particular, can be used to identify regions where GWAS and molecular QTL signals are consistent with a shared underlying variant, thereby increasing confidence that the selected instruments reflect a common source of variation (Zuber et al., 2022).

## 5.2 Common estimation approaches

Among available methods, inverse-variance weighted (IVW) regression is most commonly used as the primary estimator. It combines ratio estimates from multiple instruments-defined as the effect of a variant on the outcome divided by its effect on the exposure-using inverse-variance weighting. When all instruments are valid, or when biases average out, IVW provides efficient estimates. In the case of a single instrument, the Wald ratio is typically used.

When heterogeneity is present across instruments, random-effects formulations can be applied to account for additional variance. A number of complementary approaches have been developed to improve robustness. MR-Egger regression introduces an intercept term to detect directional bias and, under certain conditions, provides a corrected estimate, although at the cost of reduced precision. The weighted median estimator remains consistent when a subset of instruments is invalid, while mode-based estimators offer further robustness under specific assumptions about the distribution of effects.

In settings where multiple exposures may contribute jointly-such as gene expression, splicing, or protein abundance-multivariable MR (MVMR) can be used to disentangle their contributions. Although this approach can provide more detailed insight into complex regulatory relationships, it also requires stronger assumptions and higher data quality.

## 5.3 Weak instruments and pleiotropy

The strength of the instruments plays a central role in determining the stability of MR estimates. Weak associations between instruments and exposure can lead to biased estimates that are closer to observational correlations, along with inflated uncertainty. The F statistic is commonly used as a diagnostic measure, with a value of around 10 often considered a practical benchmark in univariable analyses. In multivariable settings, instrument strength needs to be evaluated separately for each exposure.

When weak instruments are detected, several strategies may help improve performance, including applying stricter selection thresholds, restricting analyses to cis-regulatory variants, or using reference datasets that better match the target population. Issues such as sample overlap and winner's curse can further weaken instruments and should be considered during study design and interpretation.

Pleiotropy introduces an additional layer of complexity. Overall heterogeneity can be assessed using statistics such as Cochran's Q, while the MR-Egger intercept provides a test for directional bias. MR-PRESSO offers procedures for identifying and correcting outlier instruments, and radial MR provides a useful visualization framework for detecting influential points. Sensitivity analyses, including leave-one-out procedures and directionality checks such as the Steiger test, can further help evaluate the robustness of the results.

In reporting, it is generally advisable to consider multiple diagnostics together. When estimates from different methods (e.g., IVW, weighted median, MR-Egger) are broadly consistent and diagnostic tests do not indicate major violations, the findings are more likely to be reliable. When discrepancies arise, it is often necessary to revisit earlier steps, including instrument selection, data harmonization, and upstream evidence.

## 5.4 Position within integrative analyses

In a broader analytical workflow, MR is typically applied after an initial round of signal prioritization. Colocalization can be used to identify loci where signals from different data sources are consistent, thereby

guiding the selection of instruments. MR is then used to further evaluate whether these signals are compatible with a directional relationship between a molecular trait and a complex phenotype.

For loci with stronger supporting evidence, replication across independent datasets or conditions can help assess robustness. In multi-tissue or multi-omics settings, multivariable models may provide additional insight into overlapping pathways. Conversely, when evidence is limited or inconsistent, it is often more appropriate to return to earlier stages of analysis, such as fine-mapping or data harmonization, rather than proceeding directly to causal interpretation.

Overall, MR is best understood as part of a sequence of analytical steps rather than a standalone method. When combined with association analyses, colocalization, and functional annotation, it contributes to a gradual refinement of evidence, moving from statistical association toward more directed interpretation of genetic effects (Zuber et al., 2022).

## **6 An Integrated Pathway for Causal Inference**

In studies of complex traits, different data types and analytical approaches each provide only partial information. Effective integration therefore requires a coherent analytical path through which initial genetic associations can be progressively refined into more interpretable results. Rather than functioning as independent tools, these methods operate in sequence, with each step narrowing the set of candidates and adding complementary evidence.

In practice, analyses often begin with GWAS signals and proceed by incorporating molecular and statistical constraints, gradually focusing from broad genomic regions to more specific genes or regulatory mechanisms.

### **6.1 From GWAS signals to candidate genes**

Analyses typically start with GWAS summary statistics. Initial steps include harmonizing allele orientation, standardizing effect sizes, and selecting LD reference panels matched to the study population. Fine-mapping can then be applied to reduce the set of candidate variants and concentrate the analysis on more localized regions (Hormozdiari et al., 2016).

At this stage, incorporating molecular QTL data provides an additional layer of information. Colocalization analysis is used to assess whether GWAS and molecular signals within a region are consistent with a shared underlying variant, thereby helping to prioritize loci for further investigation. When such consistency is observed, expression-based models can be used to translate locus-level signals into gene-level associations, further narrowing the list of candidate genes (Porcu et al., 2019; Wainberg et al., 2019; Zhang et al., 2024).

Following this prioritization, selected variants can be used as instruments to evaluate relationships between molecular traits and complex phenotypes. Mendelian randomization is commonly applied at this stage to examine potential directionality and estimate effect sizes. Through this sequence, initial GWAS signals are progressively refined into more specific hypotheses, such as the involvement of particular genes or regulatory processes (Lessard et al., 2024).

In situations where data are incomplete—for example, when high-quality eQTL data are unavailable—the analytical path may need to be adapted. TWAS can provide an initial prioritization of candidate genes, which can later be revisited using external or newly generated datasets. In such cases, however, interpretation should remain cautious, particularly when moving toward causal claims (Wainberg et al., 2019).

Across the entire process, replication in independent populations or environments can help assess robustness. Combining multiple MR methods and diagnostic measures further contributes to a more comprehensive evaluation of the evidence (Porcu et al., 2019; Zuber et al., 2022).

### **6.2 Practical considerations in analysis**

In applied settings, the choice of analytical strategy depends on data availability and quality. A key consideration is whether high-quality molecular QTL data are available in tissues relevant to the trait of interest, along with

appropriate LD reference panels. When these conditions are not met, the scope of interpretation becomes more limited.

During candidate prioritization, locally acting signals are often easier to interpret and are therefore commonly prioritized. In contrast, distal regulatory signals typically require additional supporting evidence, such as network-based analyses or multi-step approaches, to support interpretation.

Functional annotation also plays an important role. When candidate variants align with known regulatory features-such as open chromatin regions or transcription factor binding sites-the biological plausibility of the findings is strengthened. Conversely, when evidence is limited or multiple signals coexist within a region, additional refinement or data integration may be required (Hormozdiari et al., 2016).

Under favorable conditions-where molecular QTL signals align with GWAS results and suitable instruments can be identified-further analyses can be carried out, with primary estimates reported alongside sensitivity analyses using complementary methods (Porcu et al., 2019; Zuber et al., 2022). When evidence is inconsistent, however, it is often preferable to revisit earlier steps rather than proceed with interpretation.

The final output of such analyses typically combines statistical results with diagnostic measures and functional annotations, allowing candidate genes to be ranked according to the strength of evidence and prioritized for experimental validation (Figure 3) (Votava and Parks, 2021; Lessard et al., 2024).

Image caption: This figure illustrates an integrated analytical pathway for causal inference in statistical genetics. The workflow begins with GWAS signals, followed by fine-mapping to refine candidate variants. Molecular QTL data are incorporated to link genetic variation with intermediate phenotypes. TWAS translates variant-level signals into gene-level associations, and colocalization evaluates whether signals across datasets are consistent with a shared underlying variant. Mendelian randomization is then applied to assess the direction and magnitude of potential effects. Diagnostic procedures and replication across populations or environments are used to evaluate robustness. The framework represents a progressive refinement of evidence rather than a strictly linear sequence

### **6.3 Examples of application**

This integrative approach has been applied across a range of biological systems. In human studies, such as those focusing on lipid-related traits, significant GWAS loci can be examined in relevant tissues (e.g., liver) to identify candidate regions. Colocalization analyses can then be used to prioritize loci showing consistent signals, followed by expression-based analyses and MR to evaluate potential relationships. When results are stable and supported by multiple lines of evidence, they can be further interpreted in the context of known biological pathways or potential therapeutic targets (Porcu et al., 2019; Wainberg et al., 2019; Votava and Parks, 2021; Lessard et al., 2024; Zhang et al., 2024).

In plant systems, similar strategies can be applied to complex traits such as disease resistance. Analyses across tissues and environmental conditions can help identify context-dependent regulatory signals. These can then be integrated with genetic and functional data to prioritize candidate genes. Given the complexity of plant genomes, including extended LD and structural variation, the use of multi-parent populations and pangenome references can improve resolution and interpretation (Zhang et al., 2024).

Together, these examples illustrate that integration across data types is not a matter of simply combining results, but of progressively refining evidence. Through this process, genetic associations can be translated into more specific and testable hypotheses about the mechanisms underlying complex traits.

## **7 Discussion**

### **7.1 Linking molecular associations with causal evaluation**

A central challenge in the study of complex traits lies in connecting genetic associations to mechanistic interpretation. Molecular QTL analyses and TWAS contribute by constraining association signals within a functional context, enabling signals dispersed across the genome to be interpreted at the level of genes or

regulatory processes. Mendelian randomization, in turn, builds on this information to evaluate potential directionality and effect magnitude. Rather than acting as independent components, these approaches are linked through intermediate steps—most notably colocalization—which help align signals across datasets and establish continuity between different layers of evidence.

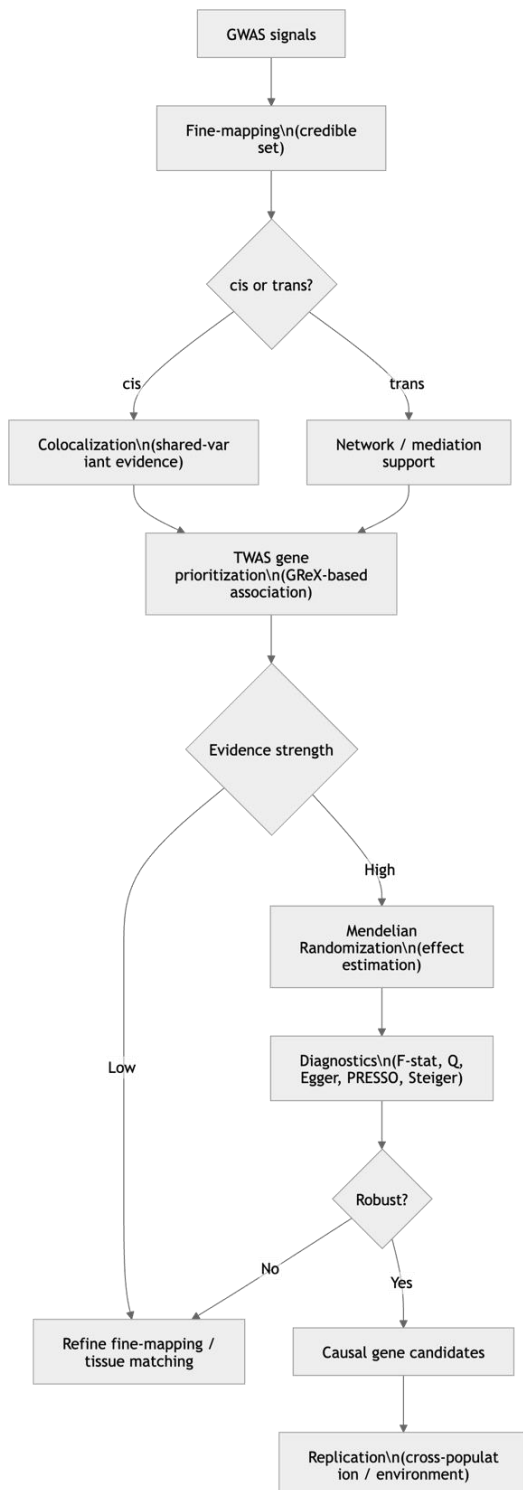


Figure 3 An integrative framework for causal inference in statistical genetics

In practice, locally acting regulatory signals, particularly cis-eQTLs derived from trait-relevant tissues, tend to be more stable and interpretable. TWAS further aggregates these signals into gene-level associations, narrowing the pool of candidates. However, such associations remain statistical in nature. Colocalization provides a way to

assess whether signals from different sources are compatible with a shared genetic origin, thereby reducing the likelihood of advancing non-causal genes into downstream analyses. Under these conditions, effect estimates obtained from MR become more interpretable in a biological context (Porcu et al., 2019).

With the increasing availability of diverse datasets, this integrative process has extended to multi-tissue and single-cell contexts, allowing regulatory effects to be examined at finer spatial and contextual resolution. For example, cell type-specific regulatory mechanisms have been shown to play a central role in certain disease processes, highlighting the value of incorporating such information into integrative analyses (Gleason et al., 2021).

At the implementation level, the choice of instruments remains a critical factor. Prioritizing cis-regulatory variants, combined with ancestry-matched LD reference panels and tissue-specific models, can improve stability. The use of multiple estimation methods alongside systematic diagnostic procedures further strengthens the robustness of findings (Hemani et al., 2018; Hu et al., 2022). In settings involving multiple molecular layers or tissues, multivariable models can help disentangle overlapping signals and avoid attributing shared regulation to a single pathway (Zuber et al., 2022).

## **7.2 The impact of pleiotropy and heterogeneity**

Despite these advances, integrative analyses remain sensitive to pleiotropy and heterogeneity. Horizontal pleiotropy presents a major challenge, as a single genetic variant may influence multiple traits through independent pathways, complicating interpretation based on a single assumed mechanism. In such cases, effect estimates may deviate from the underlying biological process (Hemani et al., 2018).

Differences across datasets-such as mismatches in tissue relevance, population structure, or environmental context-can further contribute to variability in results. Even when colocalization indicates concordance between signals, this may reflect shared genetic architecture rather than a specific mechanistic pathway (Zuber et al., 2022). Consequently, colocalization should be interpreted as a filtering step rather than as evidence of causality.

Addressing these challenges requires both careful instrument selection and the use of complementary analytical strategies. Restricting analyses to locally acting variants and accounting for LD structure can reduce confounding. At the same time, comparing results across multiple methods helps identify inconsistencies that may indicate violations of assumptions. Diagnostic and sensitivity analyses play a key role in detecting influential or invalid instruments and in assessing the robustness of conclusions (Hu et al., 2022).

Recent methodological developments have sought to model pleiotropy and heterogeneity explicitly, particularly in multi-tissue and multi-context settings. While these approaches show promise, their performance remains dependent on data quality and appropriate model specification (Gleason et al., 2021; Lu et al., 2024).

## **7.3 Extensions to plant systems**

In plant systems, integrative analyses face additional layers of complexity. Environmental effects often play a dominant role, such that the same genetic variant may exhibit different effects across developmental stages, tissues, or stress conditions. In addition, genomic features such as extended LD, structural variation, and polyploidy complicate the interpretation of association signals.

These characteristics necessitate adjustments to analytical strategies. For example, regulatory maps should ideally be constructed under trait-relevant tissues and environmental conditions, with stratified analyses used to capture context-dependent effects. Instrument selection must also account for gene copy number and homology, to avoid ambiguity in signal assignment (Porcu et al., 2019).

Population design is another important consideration. Multi-parent populations can improve resolution by reducing LD and enabling better separation of multiple signals. Replication across environmental conditions helps identify stable regulatory relationships, while multivariable approaches can be used to separate baseline and

inducible effects (Lu et al., 2024). Importantly, results that do not support a given hypothesis should also be retained, as they contribute to refining candidate prioritization and avoiding overinterpretation.

#### **7.4 Integration with emerging multi-omics and dynamic systems**

The expansion of multi-omics data has made it possible to characterize genetic effects across multiple biological layers. Epigenomic and three-dimensional genome data provide direct evidence for regulatory mechanisms, allowing the physical relationships between variants, regulatory elements, and target genes to be examined. For example, different classes of QTL and chromatin interaction data can jointly describe the connections between enhancers and promoters (Hu et al., 2018; Bhattacharya et al., 2021).

Incorporating these data into integrative frameworks allows for more detailed interpretation of regulatory pathways. When multiple molecular layers-such as expression, splicing, and protein abundance-are considered simultaneously, multivariable approaches can help distinguish their relative contributions.

Single-cell and multimodal data further extend this framework by enabling analyses at the level of cell types and cellular states. These approaches have already revealed highly specific regulatory patterns in several systems, offering new perspectives on the mechanisms underlying complex traits.

Future directions are likely to include the incorporation of temporal and perturbation data, allowing dynamic processes to be modeled more explicitly, as well as network-based approaches that consider groups of genes or regulatory modules. Coupling statistical analyses with high-throughput experimental validation may ultimately lead to integrated systems in which data-driven inference and experimental testing inform one another, advancing the transition from statistical associations to mechanistic understanding and, eventually, to targeted intervention (Colomé-Tatché and Theis, 2018; Bhattacharya et al., 2021).

### **8 Conclusion**

A central challenge in complex trait genetics is not the application of individual methods, but the establishment of a coherent analytical path that connects statistical associations to biological interpretation. In this context, colocalization analysis occupies a critical intermediate position, serving to evaluate the consistency of signals across data sources and to guide the transition from molecular association to downstream inference.

When GWAS and molecular QTL signals show stable correspondence within a genomic region, this provides a basis for prioritizing candidate genes and regulatory elements. However, such evidence reflects compatibility at the level of shared signal rather than direct insight into underlying mechanisms. In other words, colocalization supports entry into further analysis but does not, on its own, establish how a genetic variant influences a trait. Interpreting it as evidence of causality without additional support risks conflating shared genetic architecture with specific biological pathways.

Building on this foundation, Mendelian randomization offers a means to evaluate potential relationships in terms of direction and magnitude. By leveraging genetic variants as instruments, MR extends the analysis from signal alignment toward effect estimation. At the same time, its conclusions remain contingent on a set of assumptions, including the strength and validity of the instruments and the absence of alternative pathways. As a result, MR findings should be interpreted alongside diagnostic measures and complementary methods, and discrepancies should prompt re-examination of earlier analytical steps rather than isolated interpretation.

Taken together, the framework outlined here is best understood not as a fixed pipeline, but as a process of progressive refinement. Starting from GWAS signals, fine-mapping reduces the candidate variant space; molecular QTL data and colocalization analyses help identify signals that are consistent across datasets; expression-based models further narrow the focus to gene-level candidates; and, where appropriate, MR is used to evaluate potential relationships. The outputs of this process extend beyond lists of candidate genes, incorporating levels of supporting evidence and consistency that can guide experimental prioritization.

Although this general strategy is applicable across biological systems, its implementation must be adapted to the characteristics of the data. In human studies, ancestry matching and LD structure play a central role in interpretation, whereas in plant systems, environmental variation, genomic complexity, and gene copy number introduce additional challenges. These differences do not alter the overall framework but influence how individual steps are carried out and weighted.

Looking forward, advances in epigenomics, single-cell technologies, and multimodal datasets will enable the relationships between genetic variation and phenotypic outcomes to be examined across multiple biological layers. Integrating these data into existing analytical frameworks will allow regulatory pathways to be characterized with greater precision. Coupled with high-throughput experimental approaches, such developments have the potential to establish a more continuous link between data-driven inference and mechanistic validation, ultimately advancing the translation of statistical findings into actionable biological insights.

### Author Contributions

Xuanjun Fang conducted this study, including literature review, data analysis, and the writing and revision of the manuscript. The author has read and approved the final version of the manuscript.

### Acknowledgements

This work was supported by a Major Project of the National Natural Science Foundation of China (Grant No. 30490254).

### References

- Barfield R., Feng H., Gusev A., Wu L., Zheng W., Pasaniuc B., and Kraft P., 2018, Transcriptome-wide association studies accounting for colocalization using Egger regression, *Genetic Epidemiology*, 42(5): 418-433.  
<https://doi.org/10.1002/gepi.22131>
- Bhattacharya A., Li Y., and Love M.I., 2021, MOSTWAS: multi-omic strategies for transcriptome-wide association studies, *PLoS Genetics*, 17(3): e1009398.  
<https://doi.org/10.1371/journal.pgen.1009398>
- Boix C.A., James B.T., Park Y. P., Meuleman W., and Kellis M., 2021, Regulatory genomic circuitry of human disease loci by integrative epigenomics, *Nature*, 590(7845): 300-307.  
<https://doi.org/10.1038/s41586-020-03145-z>
- Broyois J., Calini D., Macnair W., Foo L., Urich E., Ortmann W., Iglesias V.A., Selvaraj S., Nutma E., Marzin M., Amor S., Williams A., Castelo-Branco G., Menon V., De Jager P., and Malhotra D., 2022, Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders, *Nature Neuroscience*, 25(8): 1104-1112.  
<https://doi.org/10.1038/s41593-022-01128-z>
- Colomé-Tatché M., and Theis F.J., 2018, Statistical single cell multi-omics integration, *Current Opinion in Systems Biology*, 7: 54-59.  
<https://doi.org/10.1016/j.coisb.2018.01.003>
- De Leeuw C., Werme J., Savage J.E., Peyrot W.J., and Posthuma D., 2023, On the interpretation of transcriptome-wide association studies, *PLoS Genetics*, 19(9): e1010517.  
<https://doi.org/10.1371/journal.pgen.1010921>
- Evans P., Nagai T., Konkashbaev A., Zhou D., Knapik E.W., and Gamazon E.R., 2024, Transcriptome-wide association studies (TWAS): methodologies, applications, and challenges, *Current Protocols*, 4(2): e981.  
<https://doi.org/10.1002/cpz1.981>
- Fagny M., Paulson J.N., Kuijjer M.L., Sonawane A.R., Chen C.Y., Lopes-Ramos C.M., Glass K., Quackenbush J., and Platig J., 2017, Exploring regulation in tissues with eQTL networks, *Proceedings of the National Academy of Sciences*, 114(37): E7841-E7850.  
<https://doi.org/10.1073/pnas.1707375114>
- Fang X.J., 2026, A hierarchical inference framework for multi-trait genetics integrating genomic SEM, PLEIO, and Primo, *Tree Genetics and Molecular Breeding*, 16(1): xx-xx.
- Foley C.N., Staley J.R., Breen P.G., Sun B.B., Kirk P.D., Burgess S., and Howson J.M., 2021, A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits, *Nature Communications*, 12(1): 764.  
<https://doi.org/10.1038/s41467-020-20885-8>
- Gleason K.J., Yang F., and Chen L.S., 2021, A robust two-sample transcriptome-wide Mendelian randomization method integrating GWAS with multi-tissue eQTL summary statistics, *Genetic Epidemiology*, 45(4): 353-371.  
<https://doi.org/10.1002/gepi.22380>
- Hemani G., Bowden J., and Davey Smith G., 2018, Evaluating the potential role of pleiotropy in Mendelian randomization studies, *Human Molecular Genetics*, 27(R2): R195-R208.  
<https://doi.org/10.1093/hmg/ddy163>

- Hendelman A., Zebell S., Rodriguez-Leal D., Dukler N., Robitaille G., Wu X., Kostyun J., Tal L., Wang P., Bartlett M.E., Eshed Y., Efroni I., and Lippman Z.B., 2021, Conserved pleiotropy of an ancient plant homeobox gene uncovered by cis-regulatory dissection, *Cell*, 184(7): 1724-1739.  
<https://doi.org/10.1016/j.cell.2021.02.001>
- Hormozdiari F., Van De Bunt M., Segre A.V., Li X., Joo J.W.J., Bilow M., Sul J.H., Sankararaman S., Pasaniuc B., and Eskin E., 2016, Colocalization of GWAS and eQTL signals detects target genes, *The American Journal of Human Genetics*, 99(6): 1245-1260.  
<https://doi.org/10.1016/j.ajhg.2016.10.003>
- Hu X., Zhao J., Lin Z., Wang Y., Peng H., Zhao H., Wang X., and Yang C., 2022, Mendelian randomization for causal inference accounting for pleiotropy and sample structure using genome-wide summary statistics, *Proceedings of the National Academy of Sciences*, 119(28): e2106858119.  
<https://doi.org/10.1073/pnas.2106858119>
- Hu Y., An Q., Sheu K., Trejo B., Fan S., and Guo Y., 2018, Single cell multi-omics technology: methodology and application, *Frontiers in Cell and Developmental Biology*, 6: 28.  
<https://doi.org/10.3389/fcell.2018.00028>
- Jiang L., Miao L., Yi G., Li X., Xue C., Li M. J., Huang H., and Li M., 2022, Powerful and robust inference of complex phenotypes' causal genes with dependent expression quantitative loci by a median-based Mendelian randomization, *The American Journal of Human Genetics*, 109(5): 838-856.  
<https://doi.org/10.1016/j.ajhg.2022.04.004>
- Khan M., Ludl A.A., Bankier S., Björkegren J.L., and Michael T., 2024, Prediction of causal genes at GWAS loci with pleiotropic gene regulatory effects using sets of correlated instrumental variables, *PLoS Genetics*, 20(11): e1011473.  
<https://doi.org/10.1371/journal.pgen.1011473>
- Kirsten H., Al-Hasani H., Holdt L., Gross A., Beutner F., Krohn K., Horn K., Ahnert P., Burkhardt R., Reiche K., Hackermüller J., Löffler M., Teupser D., Thierry J., and Scholz M., 2015, Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci, *Human Molecular Genetics*, 24(16): 4746-4763.  
<https://doi.org/10.1093/hmg/ddv194>
- Kvamme J., Badsha M.B., Martin E.A., Wu J., Wang X., and Fu A.Q., 2025, Causal network inference of cis- and trans-gene regulation of expression quantitative trait loci across human tissues, *Genetics*, 230(2): iyaf064.  
<https://doi.org/10.1093/genetics/iyaf064>
- Lessard S., Chao M., Reis K., FinnGen and Estonian Biobank Research Team, Beauvais M., Rajpal D.K., Sloane J., Palta P., Klinger K., de Rinaldis E., Shameer K., and Chatelain C., 2024, Leveraging large-scale multi-omics evidences to identify therapeutic targets from genome-wide association studies, *BMC Genomics*, 25(1): 1111.  
<https://doi.org/10.1186/s12864-024-10971-2>
- Li B., and Ritchie M.D., 2021, From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries, *Frontiers in Genetics*, 12: 713230.  
<https://doi.org/10.3389/fgene.2021.713230>
- Liang Y., Wang H., and Zhang Y.D., 2025, A-TWAS: an aggregated transcriptome-wide association study model incorporating multiple Bayesian priors, *bioRxiv*, 2025-01.  
<https://doi.org/10.1101/2025.01.27.635054>
- Liu B., Gloudeamans M.J., Rao A.S., Ingelsson E., and Montgomery S.B., 2019, Abundant associations with gene expression complicate GWAS follow-up, *Nature Genetics*, 51(5): 768-769.  
<https://doi.org/10.1038/s41588-019-0404-0>
- Lu Y., Xu K., Maydanchik N., Kang B., Pierce B.L., Yang F., and Chen L.S., 2024, An integrative multi-context Mendelian randomization method for identifying risk genes across human tissues, *The American Journal of Human Genetics*, 111(8): 1736-1749.  
<https://doi.org/10.1016/j.ajhg.2024.06.012>
- Mai J., Lu M., Gao Q., Zeng J., and Xiao J., 2023, Transcriptome-wide association studies: recent advances in methods, applications and available databases, *Communications Biology*, 6(1): 899.  
<https://doi.org/10.1038/s42003-023-05279-y>
- Mostafavi H., Spence J.P., Naqvi S., and Pritchard J.K., 2023, Systematic differences in discovery of genetic effects on gene expression and complex traits, *Nature Genetics*, 55(11): 1866-1875.  
<https://doi.org/10.1038/s41588-023-01529-1>
- Parrish R.L., Gibson G.C., Epstein M.P., and Yang J., 2022, TIGAR-V2: efficient TWAS tool with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8, *Human Genetics and Genomics Advances*, 3(1): 100078.  
<https://doi.org/10.1016/j.xhgg.2021.100068>
- Porcu E., Rüeger S., Lepik K., Santoni F.A., Reymond A., and Kutalik Z., 2019, Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits, *Nature Communications*, 10(1): 3300.  
<https://doi.org/10.10101/377267>
- Rasooly D., Peloso G.M., and Giambartolomei C., 2022, Bayesian genetic colocalization test of two traits using coloc, *Current Protocols*, 2(12): e627.  
<https://doi.org/10.1002/cpz1.627>
- Shikov A.E., Skitchenko R.K., Predeus A.V., and Barbitoff Y.A., 2020, Phenome-wide functional dissection of pleiotropic effects highlights key molecular pathways for human complex traits, *Scientific Reports*, 10(1): 1037.  
<https://doi.org/10.1038/s41598-020-58040-4>

- Tambets R., Kolde A., Kolberg P., Love M.I., and Alasoo K., 2024, Extensive co-regulation of neighboring genes complicates the use of eQTLs in target gene prioritization, *Human Genetics and Genomics Advances*, 5(4): 100187.  
<https://doi.org/10.1016/j.xhgg.2024.100348>
- Votava J.A., and Parks B.W., 2021, Cross-species data integration to prioritize causal genes in lipid metabolism, *Current Opinion in Lipidology*, 32(2): 141-146.  
<https://doi.org/10.1097/MOL.0000000000000742>
- Wainberg M., Sinnott-Armstrong N., Mancuso N., Barbeira A.N., Knowles D.A., Golan D., Ermel R., Ruusalepp A., Quertermous T., Hao K., Björkegren J.L.M., Im H.K., Pasaniuc B., Rivas M.A., and Kundaje A., 2019, Opportunities and challenges for transcriptome-wide association studies, *Nature Genetics*, 51(4): 592-599.  
<https://doi.org/10.1038/s41588-019-0385-z>
- Wallace C., 2021, A more accurate method for colocalisation analysis allowing for multiple causal variants, *PLoS Genetics*, 17(9): e1009440.  
<https://doi.org/10.1371/journal.pgen.1009440>
- Xie Y., Shan N., Zhao H., and Hou L., 2021, Transcriptome-wide association studies: general framework and methods, *Quantitative Biology*, 9(2): 141-150.  
<https://doi.org/10.15302/J-QB-020-0228>
- Zhang J., and Zhao H., 2023, eQTL studies: from bulk tissues to single cells, *Journal of Genetics and Genomics*, 50(12): 925-933.  
<https://doi.org/10.1016/j.jgg.2023.05.003>
- Zhang Y., Wang M., Li Z., Yang X., Li K., Xie A., Dong F., Wang S., Yan J., and Liu J., 2024, An overview of detecting gene-trait associations by integrating GWAS summary statistics and eQTLs, *Science China Life Sciences*, 67(6): 1133-1154.  
<https://doi.org/10.1007/s11427-023-2522-8>
- Zhao S., Crouse W., Qian S., Luo K., Stephens M., and He X., 2022, Adjusting for genetic confounders in transcriptome-wide association studies leads to reliable detection of causal genes, *bioRxiv*, 2022(9): 1-46.  
<https://doi.org/10.1101/2022.09.27.509700>
- Zheng Z., Huang D., Wang J., Zhao K., Zhou Y., Guo Z., Zhai S., Xu H., Cui H., Yao H., Wang Z., Yi X., Zhang S., Sham P.C., and Li M.J., 2020, QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes, *Nucleic Acids Research*, 48(D1): D983-D991.  
<https://doi.org/10.1093/nar/gkz888>
- Zhu H., and Zhou X., 2020, Transcriptome-wide association studies: a view from Mendelian randomization, *Quantitative Biology*, 2020: 1-15.
- Zuber V., Grinberg N.F., Gill D., Manipur I., Slob E.A., Patel A., Wallace C., and Burgess S., 2022, Combining evidence from Mendelian randomization and colocalization: review and comparison of approaches, *The American Journal of Human Genetics*, 109(5): 767-782.  
<https://doi.org/10.1016/j.ajhg.2022.04.001>



---

#### Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

## Review Article

## Open Access

# MADS-Box Transcription Factors in *Arabidopsis thaliana*: Molecular Control of Floral Organ Identity and Development

Rimjhim Chandra ✉

Department of Botany, Government Degree College, Chamba, Himachal Pradesh, 176314, India

✉ Corresponding email: [rimjhimchandra6@gmail.com](mailto:rimjhimchandra6@gmail.com)Plant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0012](https://doi.org/10.5376/pgt.2026.17.0012)

Received: 17 Apr., 2026

Accepted: 15 May, 2026

Published: 3 Jun., 2026

**Copyright** © 2026 Chandra, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**

Chandra R, 2026, MADS-box transcription factors in *Arabidopsis thaliana*: molecular control of floral organ identity and development, Plant Gene and Trait, 17(3): 173-181 (doi: [10.5376/pgt.2026.17.0012](https://doi.org/10.5376/pgt.2026.17.0012))

**Abstract** MADS-box transcription factors are central regulators of plant development, particularly in reproductive processes. This review integrates molecular, phylogenetic, transcriptomic, and functional analyses of the complete MADS-box gene family in *Arabidopsis thaliana*. These transcription factors regulate gene expression through binding to CArG-box cis-elements, with specificity determined by combinatorial protein interactions and higher-order complex formation. Genome-wide studies reveal diversification of MADS-box genes into distinct lineages with specialized developmental roles. Transcript profiling during silique development demonstrates coordinated expression patterns associated with embryogenesis, seed maturation, and fruit development. Protein interaction mapping supports the quartet model, highlighting the importance of multimeric complexes in floral organ specification. Functional analyses of B-sister genes confirm their conserved roles in ovule and seed development, particularly in endothelium differentiation. Additionally, transcriptional networks controlled by AGAMOUS reveal feedback regulation and direct target gene activation during floral organogenesis. Advances in chromatin immunoprecipitation and protein tagging approaches have enabled in vivo identification of target genes. Overall, MADS-box transcription factors act as key integrators of gene regulatory networks, coordinating plant developmental processes. This review uniquely integrates molecular, genomic, and regulatory network perspectives of MADS-box transcription factors and highlights emerging research gaps and applications in crop improvement.

**Keywords** MADS-box genes; *Arabidopsis thaliana*; Transcription factors; CArG-box; Gene regulation; Floral development; Protein-protein interactions; Gene regulatory networks; Chromatin immunoprecipitation (ChIP); Plant development

## 1 Introduction

Transcriptional regulation is a key mechanism governing plant growth and development, mediated by transcription factors that bind specific cis-regulatory elements in target genes. Among these, the MADS-box transcription factor family represents one of the most important regulators of plant developmental processes, particularly in reproductive organ formation (Riechmann and Meyerowitz, 1997; Theissen et al., 2000). In *Arabidopsis thaliana*, MADS-box genes control diverse processes including floral transition, organ identity determination, ovule development, and fruit formation, highlighting their central role in plant developmental biology (Ng and Yanofsky, 2001; Ferrario et al., 2004).

MADS-domain proteins regulate gene expression through binding to conserved CArG-box cis-elements with the consensus sequence CC(A/T)<sub>6</sub> GG. However, DNA-binding specificity is not solely determined by this motif but is influenced by combinatorial interactions among transcription factors, their spatial and temporal expression patterns, and chromatin context (Shore and Sharrocks, 1995; Kaufmann et al., 2005). Despite possessing a conserved DNA-binding domain, different MADS-box proteins regulate distinct target genes, while multiple factors may act on the same regulatory regions, reflecting complex transcriptional control (de Folter and Angenent, 2006). These regulatory mechanisms are summarised in Figure 1. Importantly, these regulatory layers do not function independently. DNA-binding specificity emerges from the coordinated interplay between chromatin accessibility, higher-order MADS-box protein complex formation, cofactor recruitment, and developmental context. This integrative regulatory framework helps explain how closely related transcription factors achieve distinct developmental outputs despite recognizing similar CArG-box motifs. Furthermore, comparative studies increasingly suggest that while core regulatory principles are conserved across angiosperms, species-specific diversification of interaction networks contributes to developmental innovation in crop plants.

Genome-wide analyses have identified more than 100 MADS-box genes in *Arabidopsis*, classified into type I and type II lineages, indicating significant evolutionary diversification and functional specialization. Expression profiling studies further demonstrate coordinated yet distinct patterns during key developmental stages such as embryogenesis, seed maturation, and fruit development (Alvarez-Buylla et al., 2000; Parenicová et al., 2003).

A characteristic feature of MADS-box proteins is their ability to form dimers and higher-order complexes, which enhance regulatory specificity. These interactions underpin the “quartet model,” where multimeric protein complexes bind to multiple cis-elements to regulate floral organ identity (Egea-Cortines et al., 1999; Theissen and Saedler, 2001). Such combinatorial control mechanisms significantly expand the functional diversity of the MADS-box family (Honma and Goto, 2001; Jack, 2001). The molecular framework underlying these interactions is illustrated in Figure 1.

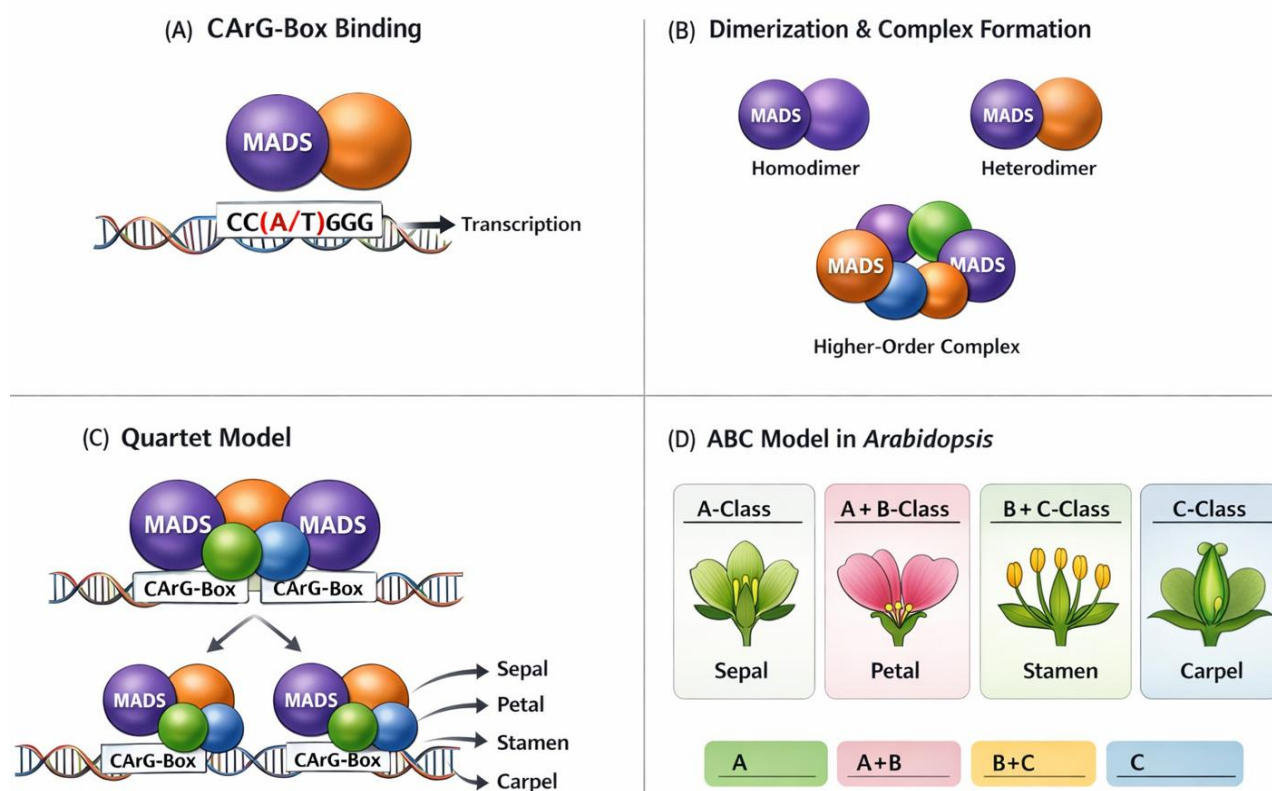


Figure 1 Regulatory framework of MADS-box transcription factors in plant development

Note: (A) MADS-box transcription factors bind to conserved CArG-box cis-regulatory elements [CC(A/T)<sub>6</sub>GG] located in the promoter regions of target genes, initiating transcriptional regulation of downstream developmental processes. (B) MADS-box proteins undergo homo- and heterodimerization and can assemble into higher-order multimeric complexes, which enhance DNA-binding specificity and enable combinatorial control of gene expression. (C) Quartet model of floral organ specification, illustrating tetrameric complexes of MADS-box proteins binding to two adjacent CArG-box elements to regulate target genes involved in the determination of floral organ identity, including sepals, petals, stamens, and carpels. (D) ABC model of floral organ identity in *Arabidopsis thaliana*, showing that A-class genes specify sepals, A- and B-class genes specify petals, B- and C-class genes specify stamens, and C-class genes specify carpels

Functional studies have demonstrated that specific MADS-box genes, such as B-sister genes and AGAMOUS, play crucial roles in ovule, seed, and floral organ development through transcriptional regulation and feedback loops (Ito et al., 2004; Gómez-Mena et al., 2005). Additionally, advances in techniques such as chromatin immunoprecipitation and transcriptomic analyses have enabled the identification of direct target genes and regulatory networks (Wang et al., 2002; Taverner et al., 2004). Recent studies have further expanded our understanding of MADS-box transcription factors, highlighting their roles in complex gene regulatory networks and evolutionary diversification (Smaczniak et al., 2012; Chen et al., 2018).

Overall, MADS-box transcription factors act as central integrators of gene regulatory networks in plants. Understanding their molecular mechanisms and regulatory interactions is essential for elucidating plant developmental processes and improving crop traits. Although numerous studies have characterized MADS-box transcription factors in *Arabidopsis thaliana*, existing reviews primarily focus on either molecular function or evolutionary aspects in isolation. This review provides an integrative perspective by linking functional genomics, transcriptional regulatory networks, and epigenetic regulation, while also highlighting their translational potential in crop improvement. Furthermore, this review identifies key knowledge gaps in understanding target specificity, chromatin context dependency, and higher-order complex formation, thereby providing directions for future research.

## 2 Classification of MADS-box Genes

MADS-box transcription factors in plants are broadly classified into two major lineages: Type I and Type II, based on phylogenetic relationships and structural features. Type I MADS-box genes are further subdivided into  $M\alpha$ ,  $M\beta$ , and  $M\gamma$  groups, and are primarily associated with roles in gametophyte and seed development. These genes generally lack conserved protein-protein interaction domains and exhibit relatively simple gene structures. In contrast, Type II MADS-box genes, also known as MIKC-type proteins, are characterized by a modular domain structure consisting of the MADS (M), Intervening (I), Keratin-like (K), and C-terminal (C) domains. These genes are extensively involved in floral organ identity, meristem determination, and reproductive development. The MIKC-type genes are further divided into  $MIKC^C$  and  $MIKC^*$  subgroups, with  $MIKC^C$  playing dominant roles in the ABC model of floral development. The structural complexity and combinatorial interaction capacity of Type II proteins underpin their functional diversification. The classification and developmental functions of major MADS-box groups are summarised in Figure 2.

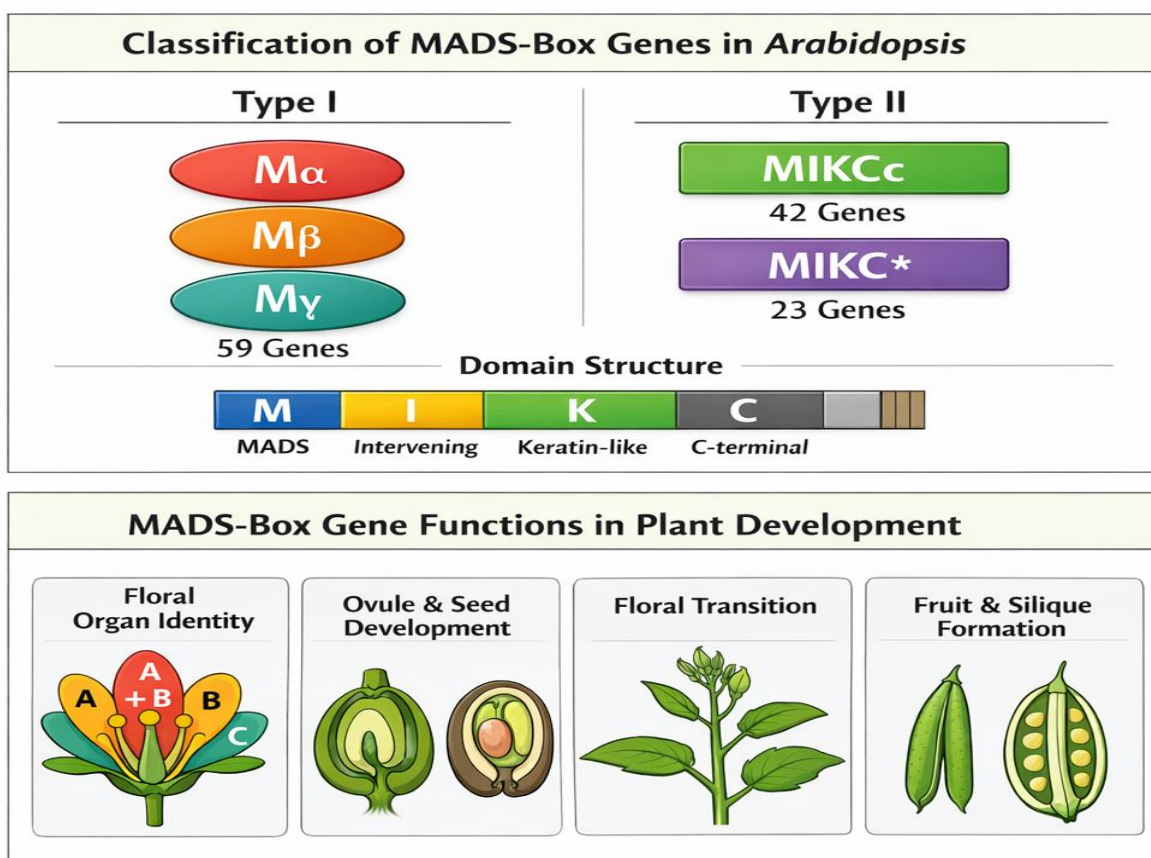


Figure 2 Classification and functions of MADS-box transcription factors in *Arabidopsis thaliana*. MADS-box genes are divided into Type I ( $M\alpha$ ,  $M\beta$ ,  $M\gamma$ ) and Type II ( $MIKC^C$  and  $MIKC^*$ ) groups. Type II proteins exhibit a conserved MIKC domain structure. These transcription factors regulate key developmental processes, including floral organ identity, ovule and seed development, floral transition, and fruit formation

## 2.1 Evolutionary diversification of MADS-box genes

The MADS-box gene family has undergone extensive expansion and diversification throughout plant evolution, primarily driven by gene duplication events such as whole-genome duplications and tandem duplications. Phylogenetic analyses indicate that ancestral MADS-box genes existed prior to the divergence of major eukaryotic lineages, followed by lineage-specific expansions in plants. In angiosperms, the diversification of MIKC-type genes has enabled the evolution of complex floral structures. Functional divergence following duplication has resulted in subfunctionalization and neofunctionalization, allowing paralogous genes to acquire specialized roles in distinct developmental pathways. Comparative genomics studies reveal conservation of core regulatory functions alongside species-specific innovations, highlighting the evolutionary plasticity of this gene family. Genome-wide analyses have further refined classification and functional annotation of MADS-box genes across plant species. Recent phylogenomic studies have revealed dynamic patterns of gene retention and loss, emphasizing the evolutionary plasticity of MIKC-type MADS-box genes (Gramzow and Theissen, 2015; Ruelens et al., 2017).

## 3 Functional Roles in Plant Development

### 3.1 Floral development

MADS-box transcription factors are central regulators of floral organ identity through the ABC model, where combinatorial gene activity determines the formation of sepals, petals, stamens, and carpels. Key genes such as *APETALA1*, *PISTILLATA*, and *AGAMOUS* coordinate organ specification via transcriptional regulation. The regulatory role of *AGAMOUS* and associated complexes has been further elucidated through recent molecular studies (Ó'Maoiléidigh et al., 2013). Furthermore, experimental evidence supporting the quartet model suggests that higher-order protein complexes provide an additional level of specificity. Recent evidence further suggests that quartet complex activity may depend on local chromatin accessibility and developmental stage, indicating that floral organ identity is regulated through dynamic integration of transcription factor assembly and epigenetic context rather than fixed combinatorial codes alone. However, the stability, composition, and in vivo dynamics of these complexes remain insufficiently characterized.

Importantly, discrepancies between mutant phenotypes and predicted ABC model outcomes highlight that redundancy and network buffering play significant roles in floral development. This suggests that floral organ identity is governed not by linear gene interactions, but by a robust and highly interconnected regulatory network.

### 3.2 Seed and ovule development

B-sister MADS-box genes have been implicated in ovule and seed coat development; however, their functional conservation across species remains only partially understood. While studies in *Arabidopsis* demonstrate a clear role in endothelium differentiation, comparative analyses suggest divergence in regulatory function in crop species.

This raises an important limitation in current research: the over-reliance on *Arabidopsis* as a model system. Although it provides valuable mechanistic insights, translating these findings to agriculturally relevant species is not straightforward due to differences in gene regulation and developmental context.

Consequently, future studies must adopt a comparative framework to determine whether observed functions represent conserved mechanisms or species-specific adaptations.

### 3.3 Fruit development

MADS-box genes also regulate fruit formation and ripening processes. They control tissue differentiation, cell expansion, and hormonal signalling pathways, contributing to fruit morphology and reproductive success.

### 3.4 Regulatory networks and epigenetic control

MADS-box transcription factors function within complex gene regulatory networks (GRNs), integrating multiple signalling pathways to control plant development. Their activity is modulated through protein-protein interactions, enabling the formation of dimers and higher-order complexes such as those described in the quartet model. Epigenetic mechanisms further refine MADS-box gene regulation. Chromatin remodelling, histone modifications, and DNA methylation influence gene accessibility and transcriptional activity. Additionally, long non-coding RNAs

and microRNAs have been implicated in post-transcriptional regulation of MADS-box genes. An important emerging concept is that transcription factor complex formation and chromatin regulation are mechanistically interconnected. MADS-box complexes may preferentially bind accessible chromatin regions, while simultaneously recruiting chromatin modifiers that reinforce transcriptional states. This reciprocal relationship likely contributes to regulatory specificity and developmental robustness. These multilayered regulatory systems ensure precise spatial and temporal gene expression during development. Advances in systems biology have enabled reconstruction of gene regulatory networks governing floral development (Chen et al., 2018). Epigenetic regulation, including chromatin modifications, plays a crucial role in modulating MADS-box gene activity. This gap highlights a broader issue in plant molecular biology: the need to move from static interaction models to dynamic, systems-level understanding. An overview of these regulatory interactions and developmental pathways is presented in Figure 3.

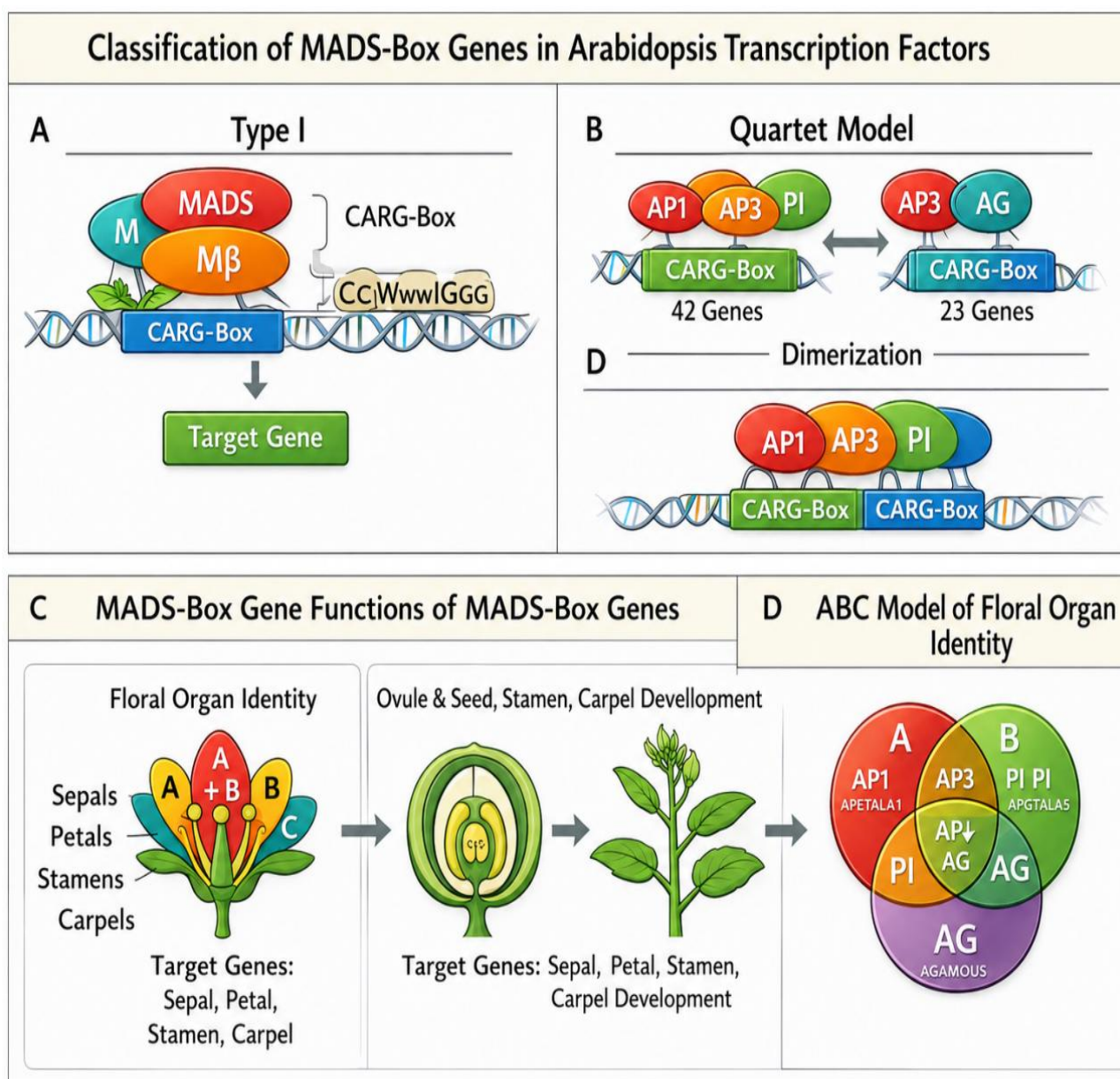


Figure 3 Regulatory roles of MADS-box transcription factors in *Arabidopsis thaliana*. MADS-box proteins bind CARG-box elements, form dimers and higher-order complexes (quartet model), and regulate floral organ identity through the ABC model, controlling key developmental processes

Importantly, these regulatory layers do not function independently but operate as an interconnected system that determines developmental specificity. The formation of distinct MADS-box transcription factor complexes influences DNA-binding selectivity, while chromatin accessibility and epigenetic modifications further determine

whether target loci can be transcriptionally activated or repressed. Thus, regulatory specificity emerges from the combined effects of protein complex composition, cis-regulatory context, and chromatin state. Insights from *Arabidopsis* have therefore provided a foundational framework for understanding conserved transcriptional regulatory mechanisms in plants, with increasing relevance for translational studies aimed at improving flowering behaviour, seed development, fruit traits, and environmental adaptability in crop species.

### 3.5 Applications in crop improvement

Understanding MADS-box gene function has significant implications for agricultural biotechnology. Manipulation of these genes can enhance traits such as flowering time, fruit quality, seed development, and yield. For example, modification of floral regulators can optimise flowering for different climatic conditions. Targeting fruit-related MADS genes can improve ripening and shelf life. Engineering seed development pathways can increase productivity. Advances in genome editing technologies, particularly CRISPR/Cas systems, have enabled precise functional studies and targeted manipulation of MADS-box genes. These approaches provide powerful tools for developing improved crop varieties with enhanced agronomic traits.

## 4 Knowledge Gaps and Future Perspectives

### 4.1 Challenges in understanding MADS-box regulatory specificity

Despite substantial conceptual and technological advances in understanding MADS-box transcription factors in *Arabidopsis thaliana*, several fundamental mechanistic questions remain unresolved that limit a comprehensive understanding of their regulatory mechanisms and functional specificity.

One of the major unresolved questions is the specificity of CARG-box binding. Although the consensus sequence CC (A/T)<sub>6</sub> GG is well established, it is widely distributed throughout the genome, raising questions about how MADS-box proteins selectively regulate target genes. Emerging evidence suggests that binding specificity is influenced by chromatin context, cofactor interactions, and higher-order protein complex formation; however, these factors remain insufficiently characterized. Collectively, current research has firmly established that MADS-box transcription factors act through combinatorial protein interactions, higher-order complex formation, and binding to CARG-box cis-elements to regulate floral and reproductive development. It is also well supported that chromatin accessibility and epigenetic modifications influence transcriptional activity. However, major uncertainties remain regarding how these regulatory layers are integrated *in vivo* to achieve developmental specificity. In particular, the dynamic assembly of transcription factor complexes, the contribution of local chromatin architecture to target selection, and the extent to which these mechanisms are conserved across plant species remain incompletely understood. Emerging approaches including single-cell multi-omics, spatial transcriptomics, live-cell imaging, chromatin conformation analyses, and AI-assisted network modelling are expected to provide deeper mechanistic insights into these unresolved questions.

Another important limitation is the lack of high-resolution spatiotemporal data. Most transcriptomic studies are based on bulk tissue analysis, which obscures cell-type-specific regulatory dynamics. The application of single-cell RNA sequencing and spatial transcriptomics could provide deeper insights into the precise roles of MADS-box genes during different developmental stages, particularly in floral organogenesis and seed development.

Furthermore, while protein-protein interactions and regulatory networks have been extensively proposed, functional validation of these networks remains incomplete. Advanced genome editing technologies, such as CRISPR/Cas systems, offer powerful tools for targeted mutagenesis and gene function analysis, yet their application in systematically dissecting MADS-box regulatory circuits is still limited.

The role of epigenetic regulation and chromatin architecture also represents an emerging frontier. Although studies have highlighted the importance of histone modifications, DNA methylation, and chromatin remodelling in modulating MADS-box gene activity, the integration of these epigenetic layers with transcription factor binding and gene regulatory networks is not fully understood.

Finally, translating knowledge from *Arabidopsis* to crop species remains a significant challenge. While MADS-box genes are evolutionarily conserved, their functional divergence across species necessitates comparative and translational studies to effectively harness their potential for crop improvement.

In the future, integrative approaches combining multi-omics technologies, high-resolution imaging, genome editing, and computational modeling will be essential to unravel the complexity of MADS-box-mediated regulatory networks. A major future challenge will be integrating transcription factor occupancy, chromatin architecture, protein interaction dynamics, and developmental signalling into predictive regulatory models capable of explaining context-dependent developmental outcomes. Such efforts will not only advance fundamental plant biology but also facilitate the development of improved crop varieties with desirable agronomic traits.

## 5 Conclusion

MADS-box transcription factors represent a central regulatory module governing plant developmental processes, particularly those associated with reproductive organ formation. As illustrated in Figure 1, their function is based on a hierarchical and highly coordinated framework that integrates cis-trans interactions, combinatorial protein complex formation, and gene regulatory network dynamics. The binding of MADS-domain proteins to CArG-box cis-elements constitutes the primary mechanism of transcriptional control; however, specificity is further refined through dimerization and higher-order complex assembly, as demonstrated by the quartet model (Honma and Goto, 2001; Jack, 2001).

Genome-wide and phylogenetic analyses highlight the extensive diversification of the MADS-box gene family in *Arabidopsis thaliana*, enabling functional specialization across multiple developmental pathways. Transcriptomic studies further reveal coordinated and stage-specific expression patterns associated with embryogenesis, seed maturation, and fruit development. Functional characterization of key genes, including B-sister MADS-box members and AGAMOUS, underscores their critical roles in ovule, seed, and floral organ development through complex transcriptional networks and feedback regulation.

Advances in molecular techniques, such as chromatin immunoprecipitation and protein tagging, have significantly improved the identification of direct target genes and in vivo protein-DNA interactions, providing deeper insights into regulatory mechanisms. Nevertheless, the widespread occurrence of CArG-box-like sequences across the genome suggests that additional factors, including chromatin context, cofactor interactions, and protein complex composition, are essential determinants of binding specificity.

In conclusion, MADS-box transcription factors act as key integrators of plant gene regulatory networks, coordinating developmental signalling through dynamic and multi-layered regulatory mechanisms. Future research focusing on high-resolution mapping of transcriptional networks, identification of cofactors, and real-time analysis of protein-DNA interactions will be critical to fully elucidate the complexity and specificity of MADS-mediated regulation. Such insights will not only advance fundamental plant biology but also provide opportunities for targeted manipulation of traits relevant to crop improvement. Importantly, this review bridges the gap between molecular mechanisms and systems-level regulatory networks of MADS-box transcription factors, emphasising the need for integrative approaches combining genomics, epigenetics, and functional validation. Such a framework is essential for translating fundamental insights from *Arabidopsis* into crop improvement strategies.

## Acknowledgements

The author acknowledges the contributions of researchers and collaborators whose work on MADS-box transcription factors has significantly advanced understanding of plant developmental biology. Insights synthesised in this review are based on studies conducted at Plant Research International, Wageningen-UR, and benefited from the scientific guidance and support of colleagues and collaborators in the field. The author extends appreciation to Prof. Gerco C. Angenent for his contributions to MADS-box research and to all researchers whose experimental and conceptual work formed the basis of this review.

## Conflict of Interest Disclosure

The author declares that there are no conflicts of interest regarding the publication of this review paper. The author confirms that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Alvarez-Buylla E.R., Pelaz S., Liljegren S.J., Gold S.E., Burgeff C., Ditta G.S., Ribas de Pouplana L., Martínez-Castilla L., and Yanofsky M.F., 2000, An ancestral MADS-box gene duplication occurred before the divergence of plants and animals, *Proceedings of the National Academy of Sciences*, 97(10): 5328-5333.  
<https://doi.org/10.1073/pnas.97.10.5328>
- Chen D., Yan W., Fu L.Y., and Kaufmann K., 2018, Architecture of gene regulatory networks controlling flower development in *Arabidopsis thaliana*, *Nature Communications*, 9(1): 4534.  
<https://doi.org/10.1038/s41467-018-06772-3>
- de Folter S., and Angenent G.C., 2006, Trans meets cis in MADS science, *Trends in Plant Science*, 11(5): 224-231.  
<https://doi.org/10.1016/j.tplants.2006.03.008>
- Egea-Cortines M., Saedler H., and Sommer H., 1999, Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in *Antirrhinum majus*, *The EMBO Journal*, 18(19): 5370-5379.  
<https://doi.org/10.1093/emboj/18.19.5370>
- Ferrario S., Immink R.G., and Angenent G.C., 2004, Conservation and diversity in flower land, *Current Opinion in Plant Biology*, 7(1): 84-91.  
<https://doi.org/10.1016/j.pbi.2003.11.003>
- Gómez-Mena C., de Folter S., Costa M.M.R., Angenent G.C., and Sablowski R., 2005, Transcriptional program controlled by the floral homeotic gene AGAMOUS during early organogenesis, *Development*, 132(3): 429-438.  
<https://doi.org/10.1242/dev.01600>
- Gramzow L., and Theissen G., 2015, Phylogenomics reveals surprising sets of essential and dispensable clades of MIKCC-group MADS-box genes in flowering plants, *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 324(4): 353-362.  
<https://doi.org/10.1002/jez.b.22598>
- Honma T., and Goto K., 2001, Complexes of MADS-box proteins are sufficient to convert leaves into floral organs, *Nature*, 409(6819): 525-529.  
<https://doi.org/10.1038/35054083>
- Ito T., Wellmer F., Yu H., Das P., Ito N., Alves-Ferreira M., Riechmann J.L., and Meyerowitz E.M., 2004, The homeotic protein AGAMOUS controls microsporogenesis by regulation of SPOROCTELESS, *Nature*, 430(6997): 356-360.  
<https://doi.org/10.1038/nature02733>
- Jack T., 2001, Relearning our ABCs: new twists on an old model, *Trends in Plant Science*, 6(7): 310-316.  
[https://doi.org/10.1016/S1360-1385\(01\)01987-2](https://doi.org/10.1016/S1360-1385(01)01987-2)
- Kaufmann K., Melzer R., and Theissen G., 2005, MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants, *Gene*, 347(2): 183-198.  
<https://doi.org/10.1016/j.gene.2004.12.014>
- Ng M., and Yanofsky M.F., 2001, Function and evolution of the plant MADS-box gene family, *Nature Reviews Genetics*, 2(3): 186-195.  
<https://doi.org/10.1038/35056041>
- Ó'Maoléidigh D.S., Wuest S.E., Rae L., Raganelli A., Ryan P.T., Kwaśniewska K., Das P., Lohan A.J., Loftus B., Graciet E., and Wellmer F., 2013, Control of reproductive floral organ identity specification in *Arabidopsis* by the C function regulator AGAMOUS, *The Plant Cell*, 25(7): 2482-2503.  
<https://doi.org/10.1105/tpc.113.113209>
- Parenicová L., De Folter S., Kieffer M., Horner D.S., Favalli C., Busscher J., Cook H.E., Ingram R.M., Kater M.M., Davies B., Angenent G.C., and Colombo L., 2003, Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world, *The Plant Cell*, 15(7): 1538-1551.  
<https://doi.org/10.1105/tpc.011544>
- Riechmann J.L., and Meyerowitz E.M., 1997, MADS domain proteins in plant development, *Biological Chemistry*, 378(10): 1079-1101.
- Ruelens P., Zhang Z., Van Mourik H., Maere S., Kaufmann K., and Geuten K., 2017, The origin of floral organ identity quartets, *The Plant Cell*, 29(2): 229-242.  
<https://doi.org/10.1105/tpc.16.00366>
- Shore P., and Sharrocks A.D., 1995, The MADS-box family of transcription factors, *European Journal of Biochemistry*, 229(1): 1-13.  
<https://doi.org/10.1111/j.1432-1033.1995.tb20430.x>
- Smaczniak C., Immink R.G., Angenent G.C., and Kaufmann K., 2012, Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies, *Development*, 139(17): 3081-3098.  
<https://doi.org/10.1242/dev.074674>
- Taverner N.V., Smith J.C., and Wardle F.C., 2004, Identifying transcriptional targets, *Genome Biology*, 5(3): 210.  
<https://doi.org/10.1186/gb-2004-5-3-210>
- Theissen G., and Saedler H., 2001, Floral quartets, *Nature*, 409(6819): 469-471.  
<https://doi.org/10.1038/35054172>

- Theissen G., Becker A., Di Rosa A., Kanno A., Kim J.T., Münster T., Winter K., and Saedler H., 2000, A short history of MADS-box genes in plants, *Plant Molecular Biology*, 42(1): 115-149.  
<https://doi.org/10.1023/A:1006332105728>
- Wang H., Tang W., Zhu C., and Perry S.E., 2002, A chromatin immunoprecipitation (ChIP) approach to isolate genes regulated by AGL15, a MADS domain protein that preferentially accumulates in embryos, *The Plant Journal*, 32(5): 831-843.  
<https://doi.org/10.1046/j.1365-3113X.2002.01455.x>



#### **Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

## Research Insight

## Open Access

# Improving Berry Uniformity in Grape (*Vitis vinifera*): Trait-Based Evaluation and Selection Perspectives

Chunmei Zhu, Yunlong Mao ✉

Changxing Heping Hanxiangmi Family Farm, Huzhou, 313103, Zhejiang, China

✉ Corresponding email: [361822731@qq.com](mailto:361822731@qq.com)Plant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0013](https://doi.org/10.5376/pgt.2026.17.0013)

Received: 25 Apr., 2026

Accepted: 28 May, 2026

Published: 13 Jun., 2026

**Copyright** © 2026 Zhu and Mao, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**

Zhu C.M., and Mao Y.L., 2026, Improving berry uniformity in grape (*Vitis vinifera*): trait-based evaluation and selection perspectives, Plant Gene and Trait, 17(3): 182-196 (doi: [10.5376/pgt.2026.17.0013](https://doi.org/10.5376/pgt.2026.17.0013))

**Abstract** This study explores the conceptual framework and evaluation methods of grape berry uniformity, elucidating its multidimensional nature arising from the coordinated contributions of berry size, shape, and cluster structure. Quantitative evaluation approaches based on the coefficient of variation, composite multi-trait indices, and high-throughput phenotyping technologies are systematically summarized. On this basis, key factors influencing berry uniformity are further analyzed, including genetic background, pollination and fertilization processes, berry developmental dynamics, plant growth regulator treatments, and water-nutrient environmental conditions. Integrating breeding strategies with production practices, a framework for improving berry uniformity is proposed, centered on “multi-trait selection, marker-assisted selection, and cultivation regulation.” Meanwhile, with the advancement of machine vision, high-throughput phenotyping, and multi-source data integration technologies, the evaluation of berry uniformity is shifting toward automation, precision, and intelligence. However, challenges remain in the standardization of evaluation systems, elucidation of molecular mechanisms, and integration of multi-source data. Future research directions toward data-driven precision improvement are discussed. This study aims to provide theoretical foundations and technical support for enhancing the quality and standardized production of table grapes.

**Keywords** Grapevine; Berry uniformity; Phenotypic evaluation; Cluster architecture; Precision breeding

## 1 Introduction

Grapevine (*Vitis vinifera* L.) is one of the most widely cultivated fruit crops worldwide, and its table grapes are highly favored by consumers due to their attractive appearance and desirable flavor. In the quality evaluation system of table grapes, in addition to intrinsic attributes such as soluble solids content, flavor, and taste, external traits also play a central role. Previous studies have shown that consumers are highly sensitive to visual characteristics such as berry size, shape, color, and overall cluster architecture, among which visual uniformity often plays a dominant role in purchasing decisions (De Oliveira et al., 2026). In this context, berry uniformity, as an important composite indicator reflecting the visual harmony of grape clusters, directly influences consumers' first impressions and product acceptance, and is therefore a key trait for assessing the commercial quality and market value of table grapes.

Clusters with high uniformity typically exhibit consistent berry size, orderly arrangement, and well-coordinated structure, which not only significantly enhance visual quality but also reflect the combined effects of genetic traits and cultivation management practices (Dobrei and Sala, 2025). In contrast, uneven berry size or irregular distribution is often regarded as an indicator of reduced commercial quality and may negatively affect market evaluation, even when intrinsic quality remains high (De Oliveira et al., 2026). In production practice, such heterogeneity is relatively common and results from the combined influence of genetic factors and environmental regulation. For example, different cultivars show significant variation in fruit set, berry enlargement capacity, and cluster architecture, while seed number, floral characteristics, and cultivation practices (such as thinning, application of plant growth regulators, and water and nutrient management) also affect berry development (Gharate et al., 2025; Milišić et al., 2025). In addition, trade-offs among cluster compactness, berry number, and individual berry size further exacerbate differences in uniformity within and between clusters (Meneses et al., 2025; Sharma et al., 2025).

Berry uniformity not only affects appearance but also leads to spatial heterogeneity in fruit quality. Differences in berry size and structure can alter the ratio of pulp to skin, thereby influencing sugar-acid composition, coloration, and antioxidant capacity, ultimately resulting in uneven eating quality within the same cluster. This heterogeneity increases the difficulty of grading, reduces packaging and transportation efficiency, and may diminish commercial value and cause economic losses. Although grape germplasm exhibits abundant genetic diversity in berry size, shape, and cluster structure (Gharate et al., 2025), most existing studies focus on individual traits such as berry diameter or cluster compactness, while systematic quantitative evaluation of “uniformity” as an integrated trait remains limited. Moreover, inconsistencies in evaluation methods among studies restrict the comparability and practical application of research findings.

This study focuses on berry uniformity as a key trait in grapevine, systematically reviewing its conceptual framework, evaluation methods, and influencing factors, and further exploring its application in cultivar selection and cultivation management strategies. The aim is to provide theoretical foundations and technical references for improving table grape quality and promoting standardized production. In recent years, advances in phenomics and quantitative genetics have provided new approaches for the precise assessment and genetic improvement of berry uniformity. High-throughput quantification of berry size distribution and cluster architecture can be achieved through digital image analysis and two- and three-dimensional segmentation techniques. Meanwhile, QTL mapping and genome-wide association studies have identified multiple genetic loci associated with berry size and cluster structure. Combined with marker-assisted selection and optimized cultivation practices, these approaches offer promising opportunities for the coordinated improvement of berry uniformity.

## **2 Conceptual Framework of Grape Berry Uniformity**

### **2.1 Conceptual components of berry uniformity**

Grape berry uniformity is a comprehensive visual trait that not only reflects the consistency of individual berry size but also involves the coordination of spatial distribution within a cluster. At the intra-cluster scale, uniformity mainly refers to the low variability in berry length, width or diameter, weight, and shape, resulting in high visual consistency among berries located at different positions within the cluster. At the whole-cluster scale, uniformity also encompasses the spatial arrangement of berries along the rachis and its branches, including whether the distribution is balanced and whether local overcrowding, sparsity, or berry deformation due to compression occurs (Torres-Lomas et al., 2024).

In practical production, berry uniformity is the result of the combined effects of genetic background, reproductive development, and cultivation environment. Factors such as pollination quality, fruit set rate, seed development, berry growth rate, and the balance of assimilate distribution all influence berry size and the synchronization of ripening. Studies have shown that berries from different genotypes or with different seed numbers exhibit significant differences in size and uniformity, indicating that uniformity has a strong genetic dependency and developmental basis (De Oliveira et al., 2026). When berry development is synchronized and spatial distribution is well balanced, clusters typically display a full, orderly, and marketable appearance; otherwise, problems such as mixed berry sizes, local crowding, or excessive gaps may occur.

In addition, different grape cultivars exhibit substantial variation in cluster structure and berry development patterns. Some cultivars naturally produce compact clusters, which may enhance visual fullness but excessive compactness can lead to berry compression and deformation. In contrast, loosely structured clusters may reduce compression but can result in uneven spatial distribution and reduced visual coordination (Torres-Lomas et al., 2024). Therefore, berry uniformity should not be simply interpreted as “larger berries are better” or “more compact clusters are better,” but rather as a comprehensive expression of size consistency, shape uniformity, and coordinated cluster architecture.

### **2.2 Key phenotypic traits of berry uniformity**

Among the factors contributing to berry uniformity, berry size traits represent the most direct and fundamental basis for evaluation. Berry length, width (or diameter), and single-berry weight are commonly used descriptors of berry size, and these traits are often positively correlated, meaning that a berry large in one dimension is typically

large in others as well (Dobrei and Sala, 2025). When variability in these traits is low within a cluster, the overall appearance is visually uniform; otherwise, noticeable size heterogeneity may occur, reducing uniformity.

Beyond size, berry shape provides an important complementary dimension for evaluating uniformity. The shape index, typically defined as the ratio of berry length to width, is used to distinguish morphological types and reflect shape consistency. More advanced descriptors, such as eccentricity, contour curvature, and elliptical Fourier descriptors, allow a more precise characterization of berry outlines, thereby improving the accuracy of shape evaluation (De Sousa Moreira et al., 2024). Previous studies have shown significant differences in shape stability among grape materials, indicating that shape consistency is also an important component of uniformity.

At the structural level, cluster compactness serves as a key link between individual berry traits and overall cluster appearance. This trait integrates berry number, berry size, and rachis structure, and can be interpreted as the degree of space filling within the cluster (Torres-Lomas et al., 2024). Moderately compact clusters contribute to a full and orderly appearance, whereas excessive compactness may lead to berry deformation and uneven development, and overly loose clusters may result in poor arrangement and visual inconsistency. Therefore, berry size, shape, and cluster structure collectively form the core phenotypic basis of berry uniformity, which is essentially the result of the coordinated interaction of multiple traits.

### **2.3 Quantitative evaluation methods of berry uniformity**

To achieve an objective assessment of berry uniformity, phenotypic traits such as berry size, shape, and cluster structure must be transformed into quantifiable indicators. Among these, the coefficient of variation (CV) is the most widely used statistical metric and is employed to describe the dispersion of traits such as berry weight, length, width, and area within a cluster (Dobrei and Sala, 2025; Milišić et al., 2025). A lower CV value indicates smaller differences among berries and thus higher uniformity, whereas a higher CV reflects greater variability. Due to its simplicity and comparability, CV serves as a fundamental quantitative tool for uniformity evaluation.

Building upon this, composite uniformity indices can be constructed to integrate multiple traits into a comprehensive evaluation. Such indices are typically derived by standardizing variables with different units and assigning weights according to research objectives or commercial grading requirements, thereby reflecting both size and shape consistency. Compared with a single CV metric, composite indices provide a more holistic characterization of uniformity, particularly for comparisons among cultivars and for breeding selection. In addition, structural parameters such as compactness indices, berry number per unit cluster length, and spatial distribution descriptors can be incorporated to quantify berry arrangement within clusters.

With the advancement of digital image analysis and high-throughput phenotyping technologies, the evaluation of berry uniformity is transitioning from manual measurement to automated and intelligent approaches. Two-dimensional image analysis can extract parameters such as berry area, length, width, and shape, while three-dimensional reconstruction techniques can provide additional information on berry volume, cluster volume, and spatial distribution. In recent years, instance segmentation methods based on vision models such as the Segment Anything Model (SAM) have enabled automatic identification of individual berries and extraction of multidimensional phenotypic parameters, providing robust support for high-throughput analysis of uniformity-related traits (Torres-Lomas et al., 2024; Sharma et al., 2025). Overall, berry uniformity evaluation is evolving toward a “multi-indicator integration+high-throughput measurement” paradigm, providing an important technical foundation for precision breeding and standardized production.

## **3 Evaluation Methods of Grape Berry Uniformity**

### **3.1 Traditional evaluation methods**

In grape production, postharvest sorting, and market circulation, the evaluation of berry uniformity has long relied on visual grading methods based on cluster appearance. This approach typically assesses the overall coordination of the cluster by observing whether berry size is uniform, whether the arrangement is orderly, whether cluster structure is appropriate, and whether mixed berry sizes occur, thereby classifying clusters into different commercial grades. Clusters with high uniformity usually exhibit consistent berry size, balanced spatial

distribution, a clean appearance, and strong market appeal, whereas clusters with obvious size variation, local overcrowding, or excessive sparsity are more likely to be downgraded.

Within scientific evaluation systems, the descriptors established by the International Organisation of Vine and Wine (OIV) provide a relatively standardized basis for visual grading of grape clusters and berries. For example, OIV 204 is used to evaluate cluster compactness, and OIV 221 is used for berry size classification. Evaluators typically assign compactness levels ranging from very loose to very compact based on cluster appearance and determine berry size grades according to berry diameter ranges. This method is simple to operate, low in cost, and requires minimal instrumentation, and thus remains widely used in field surveys, germplasm characterization, and production grading.

However, visual grading is inherently semi-quantitative and experience-based, making it susceptible to evaluator expertise, interpretation of scoring scales, and subjective bias. Particularly in populations with small phenotypic differences, different evaluators may assign different scores to the same cluster, leading to reduced reproducibility and comparability of data (Sharma et al., 2025). In addition, traditional berry size grading often employs relatively coarse classification scales, making it difficult to detect subtle differences in berry size distribution within clusters and unable to accurately reflect spatial distances among berries, internal void proportions, or local crowding within clusters. Therefore, although visual grading retains practical value in production, it is insufficient for detailed analysis of berry uniformity under the demands of modern grape production, which emphasizes standardization, precision, and efficient breeding selection.

### 3.2 Quantitative evaluation methods

To overcome the subjectivity of traditional visual grading, the evaluation of grape berry uniformity has gradually shifted from qualitative description to quantitative analysis based on continuous variables. This approach constructs a multi-indicator evaluation system by measuring individual berry traits such as length, width, diameter, weight, area, and volume, as well as structural traits including berry number per cluster, cluster length, width, and weight. Among these, berry size distribution serves as the foundation for assessing size consistency. By analyzing the mean, standard deviation, and range of berry size or weight, the degree of concentration and dispersion within a berry population can be directly quantified, providing a preliminary basis for evaluating uniformity.

Among various statistical indicators, the coefficient of variation (CV) is one of the most widely used parameters for uniformity evaluation. CV describes trait dispersion as the ratio of standard deviation to mean, effectively eliminating the influence of different measurement scales and mean values, thereby improving comparability among materials. For a given cluster, smaller CV values for berry diameter, weight, or volume indicate more concentrated distributions and higher uniformity, whereas larger CV values indicate greater variability. Somogyi et al. (2021), in a study on ‘Italia’ grapes, found significant differences in the CV of berry weight and circumference among berries with different seed numbers, with seedless berries showing higher variability and seeded berries generally exhibiting greater uniformity, highlighting the important biological role of seed number in determining berry uniformity.

Beyond berry size, cluster compactness provides a critical structural dimension for evaluating uniformity. Compactness reflects the integrated relationship among berry number, berry volume, and rachis structure, and can be quantified using metrics such as berry number per unit cluster length, the ratio of berry area to projected cluster area, or the ratio of total berry volume to cluster volume (Meneses et al., 2025). Compared with single size metrics, compactness better captures the spatial arrangement of berries. For instance, clusters with similar average berry size may exhibit different levels of uniformity due to differences in berry number or internal void space. Therefore, uniformity evaluation should be based on multi-indicator integration, considering both size consistency and spatial structural coordination.

### 3.3 Modern technological approaches

With the development of computer vision and intelligent agricultural technologies, high-throughput phenotyping methods based on digital image analysis have become important tools for evaluating berry uniformity.

Two-dimensional image analysis typically uses RGB images to identify berry contours and extract parameters such as berry number, projected area, length, width, aspect ratio, and cluster dimensions. The tools such as the Berry Analysis Tool (BAT) and Cluster Analysis Tool (CAT) can achieve automatic berry counting and size estimation, with results highly consistent with manual measurements, providing an efficient and objective basis for uniformity evaluation.

Building upon 2D image analysis, deep learning-based segmentation models have significantly improved the accuracy of berry detection under complex conditions. Instance segmentation methods such as Mask R-CNN, as well as foundation vision models like the Segment Anything Model (SAM), can automatically identify individual berries under varying lighting, occlusion, and background interference, and extract key parameters such as size distribution and compactness (Figure 1) (Kim et al., 2023; Torres-Lomas et al., 2024; Sharma et al., 2025). These approaches enable the transition from manual sampling to high-throughput measurement at the whole-cluster or population scale, providing more reliable phenotypic data for breeding selection and genetic analysis.

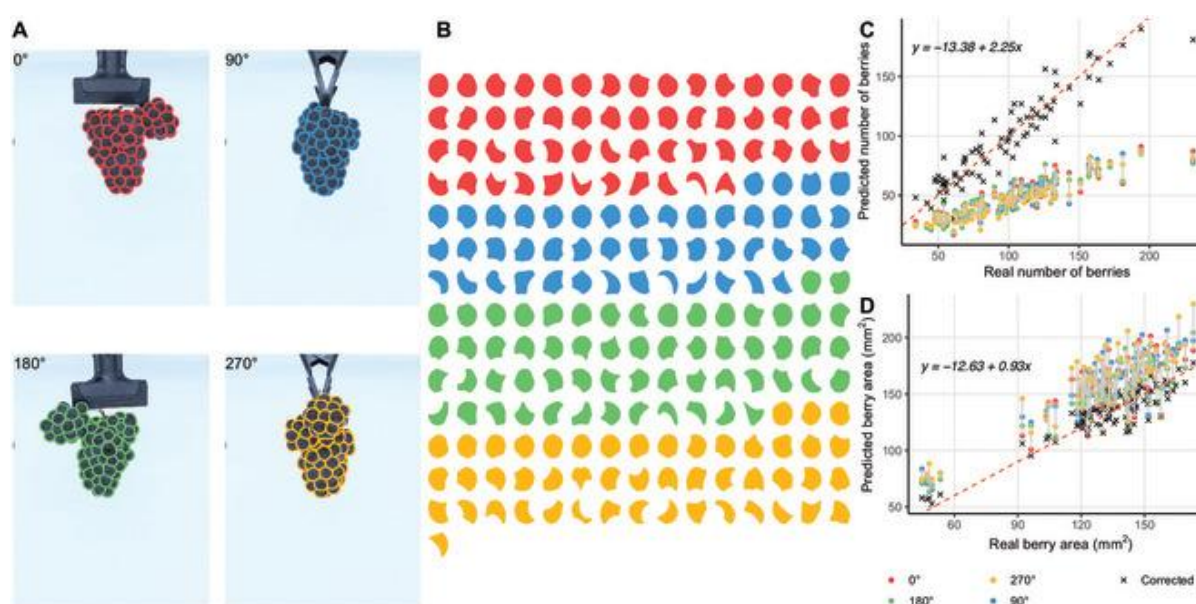


Figure 1 Prediction of berry number using SAM from cluster images (Adopted from Torres-Lomas et al., 2024)

Image caption: (A) Identification of individual berries from 4 angles on the same cluster. (B) Berry masks from cluster images in panel A, color-coded by angle view. (C) Correlation between real and predicted berry counts from SAM; predicted counts for each angle view in panel A are displayed. Points marked with an X represent corrected counts using the angle view with the maximum berries, adjusted with a linear model. (D) Correlation between real and predicted berry area; color and shape patterns are similar to panel C; corrected points were generated with a linear model of the form  $y \sim \beta_0 + \beta_1 x$ . The vertical red line indicates a one-to-one relationship between variables (Adopted from Torres-Lomas et al., 2024)

Furthermore, three-dimensional modeling techniques overcome the limitations of 2D image analysis in representing spatial structure. Through 3D scanning, stereo vision, or point cloud reconstruction, parameters such as berry number, average diameter, individual berry volume, cluster envelope volume, and spatial compactness can be obtained. Compared with 2D methods, 3D analysis provides a more accurate description of inter-berry distances, internal voids, and cluster closure, and is suitable for dynamic monitoring of cluster development (Trivedi et al., 2023). Combined with mobile devices and field platforms, these technologies are driving berry uniformity evaluation toward automation, scalability, and intelligent applications.

## 4 Factors Affecting Grape Berry Uniformity

### 4.1 Genetic factors

Grape berry uniformity is primarily influenced by genetic background. Different cultivars exhibit inherent differences in berry size, shape, and cluster structure, which arise from genetic traits such as berry developmental potential, fruit set stability, seed formation capacity, and cluster architectural formation. Studies have shown that berry size and cluster structure display extensive variation within grape germplasm, with berry weight ranging

from less than 1 g to approximately 10 g, closely associated with cell division and expansion before and after flowering as well as pericarp tissue development. Therefore, the performance of berry uniformity among cultivars has a strong genetic basis.

In table grape breeding, genetic selection plays a decisive role in improving berry size and uniformity standards. Breeding programs represented by ‘Kyoho’ and its derivatives have significantly increased berry size, while ‘Shine Muscat’ has further achieved a combination of large berries, excellent flavor, and good resistance. This cultivar has an average berry weight of approximately 10-12 g and can achieve seedless production through gibberellic acid (GA<sub>3</sub>) treatments at full bloom and post-bloom stages, resulting in high berry consistency through the synergistic effects of genetic potential and cultivation practices. This demonstrates that the selection of superior cultivars is one of the core approaches to improving berry uniformity.

In addition to varietal differences, clonal variation is also an important genetic source of uniformity. Studies have shown that different clones may exhibit stable differences in seed number, fruit set rate, berry size, and cluster density, with seedless or low-seed types often displaying distinct berry developmental patterns and structural characteristics (Alañón-Sánchez et al., 2026). Overall, berry uniformity is a typical quantitative trait controlled by multiple genes and influenced by genotype-environment interactions, requiring multi-trait selection and long-term breeding efforts to achieve stable improvement.

#### **4.2 Physiological factors**

Berry uniformity largely depends on key physiological processes from flowering to early fruit development, among which the adequacy and synchrony of pollination and fertilization are fundamental for uniform berry development. Insufficient pollen viability or uneven pollination conditions may result in poor fertilization of some flowers, leading to the formation of underdeveloped small berries and increased variability within clusters. Sabir et al. (2020) demonstrated that supplementary or cross-pollination can significantly improve fruit set and berry development, indicating a direct effect of pollination quality on uniformity.

Pollination and fertilization further influence berry enlargement by regulating hormonal signaling and cellular development processes. Dauelsberg et al. (2011) reported that successfully pollinated berries exhibit larger diameters and faster flesh expansion, whereas berries formed from unpollinated flowers remain smaller and developmentally restricted. These differences are closely associated with the expression of genes related to gibberellins, auxins, and cytokinins, indicating that fertilization activates hormone-mediated regulation and cell division required for early fruit development. In addition, differences in pollen source and viability can influence berry size and uniformity through metaxenia effects (Dhakad et al., 2024).

During the berry enlargement stage, developmental synchrony becomes a key determinant of uniformity. Berry growth depends on assimilate supply, water transport, and hormonal regulation, and differences among berries in seed number, hormone levels, and competitive ability for resources can lead to asynchronous development, resulting in size variability. Therefore, berry uniformity is the cumulative outcome of multiple developmental stages, including pollination, fertilization, seed development, and berry enlargement, and is fundamentally determined by the synchrony of berry development.

#### **4.3 Cultivation and environmental factors**

In production practice, cultivation management is the most direct and controllable factor affecting berry uniformity. Practices such as flower thinning, cluster thinning, and berry thinning reduce berry number, optimize the source-sink relationship, and decrease competition among berries, thereby promoting balanced development of the remaining berries. Khalil et al. (2023) reported that cluster thinning can significantly increase berry weight and diameter in certain cultivars, although responses vary among genotypes, indicating that cultivation practices must be adapted to genetic background. For compact clusters, thinning also improves spatial distribution, reduces compactness, and enhances visual quality (Alshallash et al., 2023; Choi et al., 2023).

Plant growth regulators, particularly gibberellic acid (GA<sub>3</sub>), play an important role in regulating berry uniformity. Appropriate GA<sub>3</sub> treatments can promote berry enlargement and improve uniformity, whereas improper applications may exacerbate size variability. Studies have shown that multi-stage GA<sub>3</sub> applications enhance vascular development and sugar transport, thereby promoting berry growth (Cai et al., 2024), while combined treatments of GA<sub>3</sub> and CPPU regulate endogenous hormone levels to influence cell division and expansion in berries (Figure 2) (Choi et al., 2023; Chen et al., 2025). However, such regulation is highly cultivar-dependent. For example, Acharya et al. (2025) reported that low concentrations of GA<sub>3</sub> had limited effects in ‘Cabernet Sauvignon’, indicating the need for cultivar-specific optimization.



Figure 2 The shape of ‘Shine Muscat’ grapes at harvest according to PGR treatment (Adopted from Choi et al., 2023)

Image caption: The combinations of PGR treatments applied at full bloom (F) and 12 days after full bloom (DAFB) are as follows; Group 1: F: GA<sub>3</sub> 12.5+TDZ 2.5, 12 DAFB: GA<sub>3</sub> 25. Group 2: F: GA<sub>3</sub> 25 + CPPU 5, 12 DAFB: GA<sub>3</sub> 25. Group 3: F: GA<sub>3</sub> 25 + TDZ 5, 12 DAFB: GA<sub>3</sub> 25. Group 4: GA<sub>3</sub> 25 + CPPU 5, 12 DAFB: GA<sub>3</sub> 25+ CPPU 5. Group 5: F: GA<sub>3</sub> 25 + CPPU 5, 12 DAFB: untreated. The numbers that appear with the PGR are concentrations and their unit is mg/L. Abbreviations: gibberellic acid 3 (GA<sub>3</sub>), thidiazuron (TDZ), forchlorfenuron (CPPU) (Adopted from Choi et al., 2023)

Water and nutrient management, as well as climatic conditions, also have significant effects on berry uniformity. Stable water and nutrient supply helps maintain synchronized berry growth, whereas water stress or nutrient fluctuations may disrupt developmental balance. Environmental factors such as temperature, light, and precipitation not only affect pollination and seed formation but also influence berry growth rate and cluster structure, leading to inter-annual variation in uniformity. Therefore, improving berry uniformity requires an integrated consideration of genetic background, physiological processes, and cultivation environment, with coordinated regulation of multiple factors to achieve stable optimization.

## 5 Selection Strategies for Improving Grape Berry Uniformity

### 5.1 Trait selection

The first step in improving grape berry uniformity is to establish clear, quantifiable, and selection-oriented trait criteria. For table grapes, the combination of large berry size, uniform appearance, and coordinated cluster structure plays a crucial role in consumer preference, commercial grading, and market competitiveness. Therefore, selection should not focus solely on average berry size but should emphasize berry size uniformity, stability of berry morphology, and the coordination of spatial structure within clusters.

At the berry level, the coefficient of variation (CV) of berry size can serve as a key indicator for assessing size uniformity. A lower CV indicates a more concentrated distribution of berry size within a cluster and thus higher uniformity. Consequently, in cultivar screening, germplasm evaluation, and progeny selection, priority should be

given to materials with higher mean values of berry weight, length, width, and volume, combined with lower CV values. Previous studies have shown that grape germplasm resources and breeding populations exhibit wide variation in berry weight, length, diameter, seed number, and cluster structural traits, providing a genetic basis for selecting genotypes with uniform berries and coordinated cluster architecture (Güler and Karadeniz, 2023; Gharate et al., 2025).

At the cluster level, an ideal cluster should maintain a balance between compactness and openness. Overly loose clusters may exhibit large gaps and uneven berry distribution, whereas excessively compact clusters are prone to berry compression, deformation, and uneven development. Therefore, selection for uniformity should comprehensively consider cluster length, width, berry number, berry density, compactness, and overall cluster regularity (Thorat et al., 2024; Sharma et al., 2025). Genome-wide association studies (GWAS) and QTL analyses have shown that berry size and cluster structure traits are controlled by multiple genes but generally exhibit moderate to high heritability, enabling gradual improvement through continuous selection and the utilization of stable genetic loci (De Oliveira et al., 2026).

## **5.2 Breeding strategies**

From a genetic improvement perspective, enhancing grape berry uniformity requires an integrated strategy combining high-quality germplasm evaluation, hybridization design, and marker-assisted selection. It is essential to systematically identify superior genetic resources from existing cultivars, local germplasm, and breeding materials that exhibit high berry uniformity, stable cluster structure, and consistent fruit set performance, and to use these as parental lines in hybrid breeding. Germplasm studies indicate that grape berry size, cluster density, berry number, and seed traits show substantial variation, providing a foundation for targeted selection of uniformity-related traits (Güler and Karadeniz, 2023; Gharate et al., 2025).

Because berry uniformity is a complex quantitative trait, relying solely on phenotypic selection is often time-consuming and susceptible to environmental interference. In recent years, technologies such as QTL mapping, GWAS, RNA-Seq, and high-density SNP genotyping have provided powerful tools for dissecting the genetic basis of uniformity-related traits. Multiple studies have identified QTLs associated with berry weight, length, diameter, seed traits, and cluster structure, with some loci showing stable expression across years and populations (García-Abadillo et al., 2024). For example, QTLs associated with berry weight have been located on chromosomes 11, 17, and 18, while major loci for berry length and width have been identified on chromosomes 14 and 3-5 (Figure 3) (Thorat et al., 2024; De Oliveira et al., 2026).

Marker-assisted selection (MAS) enables the early identification of progeny carrying favorable alleles, thereby improving breeding efficiency for uniformity-related traits. Studies based on RNA-Seq and GWAS have identified multiple SNP and InDel markers associated with berry size, some of which co-localize with candidate genes involved in cell number, cell wall modification, and hormone signaling pathways. In practice, a strategy integrating “phenotypic pre-screening + marker-assisted selection + multi-year stability validation” can be adopted: first selecting materials with low CV, large average berry size, and coordinated cluster structure, then enriching favorable alleles through MAS or genomic prediction, and finally validating the stability of uniformity across multiple environments.

## **5.3 Cultivation regulation**

At the production level, cultivation practices represent the most direct and rapidly effective approach for improving berry uniformity. Even when cultivars possess favorable genetic potential, inappropriate cluster load, thinning intensity, growth regulator application, or water and nutrient management may still result in uneven berry size, excessive compactness, or localized developmental imbalance. Therefore, uniformity optimization should be based on cultivar-specific genetic characteristics and achieved through coordinated regulation of cluster shaping, quantitative thinning, and appropriate application of plant growth regulators to ensure balanced and stable berry development.

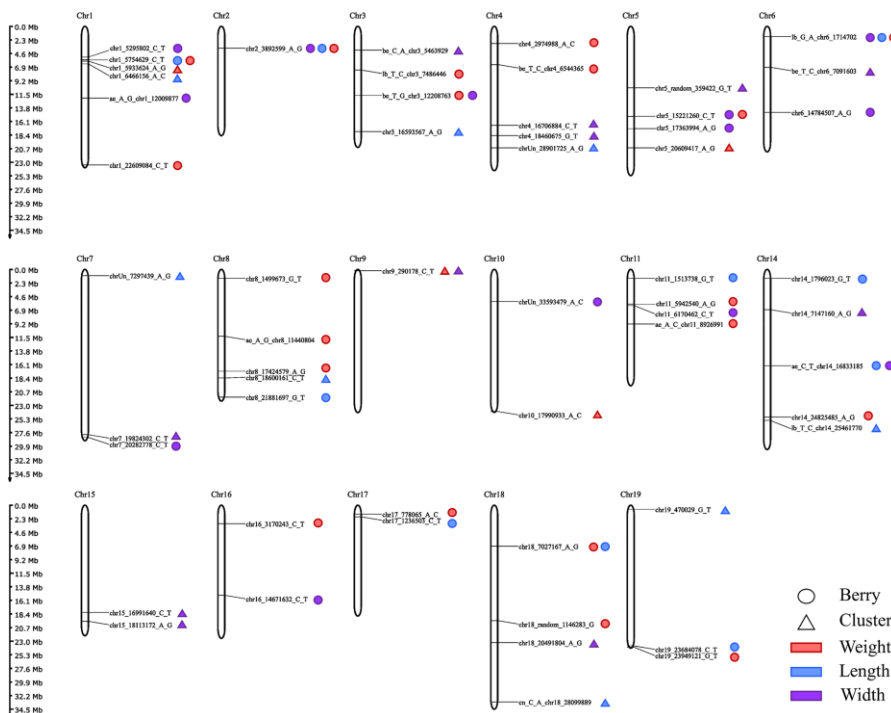


Figure 3 Chromosomal localization of significant SNPs associated with berry- and cluster-related traits (Adopted from De Oliveira et al., 2026)

Image caption: The chromosome number is shown at the top of each chromosome, and chromosome sizes are depicted on a vertical scale (Mb) (Adopted from De Oliveira et al., 2026)

Cluster shaping and quantitative thinning are key techniques for improving cluster structure and berry uniformity. By adjusting inflorescence length, secondary cluster number, and branching density before and after flowering, berry crowding at later stages can be effectively reduced, creating space for uniform berry enlargement. For compact cultivars, shortening inflorescences or removing secondary clusters can reduce berry density, whereas for loose cultivars, excessive thinning should be avoided to maintain cluster fullness. Implementing quantitative berry thinning during the early fruit stage removes underdeveloped or overly dense berries, reduces resource competition, and ensures a more balanced assimilate supply to the remaining berries. Studies have shown that mechanical or chemical thinning can significantly improve berry size and cluster structure in compact cultivars. In ‘Shine Muscat’, moderate thinning promotes berry enlargement and maintains sugar-acid balance, whereas excessive thinning may reduce fruit quality (Choi et al., 2023), indicating that thinning intensity must be carefully adjusted according to cultivar characteristics and production objectives.

Plant growth regulators, particularly gibberellic acid (GA<sub>3</sub>), play an important role in regulating berry uniformity. Appropriate timing and concentration of GA<sub>3</sub> treatments can promote berry enlargement, increase berry diameter and cluster weight, and act synergistically with thinning practices. However, the effects of GA<sub>3</sub> are highly cultivar-dependent and sensitive to dosage, and improper application may result in uneven berry development, altered skin characteristics, or delayed ripening. Therefore, GA<sub>3</sub> application should be integrated with cluster shaping, thinning, and water and nutrient management to form a systematic regulation strategy. Overall, the coordinated optimization of multiple cultivation practices can effectively enhance developmental synchrony and represents a key pathway for achieving stable improvement in berry uniformity.

## 6 Digitalization and Intelligent Development Trends

### 6.1 Machine vision-based automatic identification and evaluation systems for berry uniformity

With the rapid development of artificial intelligence, machine vision, and high-throughput phenotyping technologies, image-based automatic evaluation systems for grape berry uniformity have become a key direction for improving efficiency, objectivity, and reproducibility. These systems typically acquire cluster information using high-resolution RGB images, mobile device images, or field close-range images, and integrate image

segmentation, object detection, edge reconstruction, and geometric analysis to achieve automatic identification and parameter extraction of individual berries, including berry number, diameter, area, shape, spatial coordinates, and distribution density.

Early machine vision approaches mainly relied on two-dimensional image processing. For example, methods based on edge detection and geometric analysis can identify berry contours and estimate berry diameter, with an average error of approximately 2-3 mm and good stability across different grape types (Luo et al., 2021). Automated frameworks based on conditional random fields can classify approximately circular structures in images as “berries” or “non-berries,” with high correlation between image-derived and manually measured diameters ( $\rho \approx 0.88$ ) (Roscher et al., 2017). In addition, field-scale berry size mapping systems can control diameter estimation errors within 6% and show strong correlations with berry weight ( $R^2 \approx 0.96$ ), providing a foundation for automated evaluation of spatial variation and uniformity (Mirbod et al., 2016).

In recent years, deep learning models have further improved berry detection accuracy under complex backgrounds. The Segment Anything Model (SAM) enables high-precision segmentation of individual berries in large sets of 2D cluster images, showing strong agreement with manual annotations (Pearson correlation  $r = 0.96$ ), and can generate over 150,000 berry masks with spatial coordinates for analyzing berry size distribution, cluster compactness, and spatial structure (Torres-Lomas et al., 2024). Moreover, instance segmentation models based on AS-SwinT and end-to-end berry counting algorithms can automatically detect and count berries before thinning, supporting intelligent thinning decisions in high-value table grapes such as ‘Shine Muscat’ (Du and Liu, 2023). Mobile vision systems integrating Mask R-CNN and calibration objects can achieve sub-millimeter accuracy in berry diameter measurement and dynamically track berry growth, providing a technical basis for temporal monitoring of berry uniformity (Upadhyaya et al., 2023).

## 6.2 Trait association analysis and predictive modeling based on big data

In the context of digitalization, research on grape berry uniformity is shifting from single-point measurements to multi-source data integration and large-scale modeling. By integrating phenotypic data across different cultivars, years, ecological regions, and cultivation conditions, it is possible to systematically analyze the relationships between berry uniformity and yield, fruit quality, stress resistance, and marketable fruit rate. For instance, questions such as whether lower berry CV corresponds to a higher proportion of marketable fruit, whether optimal cluster compactness reduces disease risk, or whether uniform berry size affects sugar and acid accumulation can be addressed through big data modeling.

High-throughput phenotyping platforms provide the foundation for such analyses. Automated berry imaging systems can extract over 100 traits per fruit within approximately one second, including size, shape, and color parameters, and store the data in standardized formats for subsequent genetic analysis and model training. When combined with genome-wide SNP data, these high-density phenotypic datasets can be used for GWAS, haplotype analysis, and multi-trait selection, enabling the dissection of the genetic basis of complex traits such as berry shape, sugar content, organic acids, and stress tolerance (Zhang et al., 2025).

At the vineyard scale, multimodal sensing and machine learning models can be used to establish predictive relationships among environment, plant status, yield, and fruit quality. By integrating hyperspectral vegetation indices, thermal infrared indices, photosynthetically active radiation interception, stem water potential, chlorophyll content, and gas exchange parameters, and applying algorithms such as random forest and gradient boosting, it is possible to accurately predict traits such as average berry weight, berry number per cluster, cluster weight, total yield, soluble solids content, pH, titratable acidity, and maturity index, with some models achieving  $R^2$  values greater than 0.9 (Jewan et al., 2024). UAV-based multispectral and thermal remote sensing studies have also shown that vegetation indices are positively correlated with yield and berry weight, while canopy temperature is related to berry pH, polyphenol content, and anthocyanin levels, providing a basis for zone management and selective harvesting (Lee et al., 2024). In addition, artificial neural networks can use CIELab color parameters to accurately predict berry physicochemical properties, with correlation coefficients reaching  $R \approx 0.98-0.99$ , indicating that color information is also an important variable for modeling fruit maturity and quality.

### **6.3 Application of intelligent management in precision regulation**

Digital and intelligent technologies are not only used for evaluating berry uniformity but are also transforming vineyard resource management. Precision viticulture systems integrate satellite remote sensing, UAVs, ground-based sensors, soil moisture monitoring, and plant physiological data to spatially identify variations in plant growth, water stress, nutrient supply, and canopy status within vineyards, thereby enabling variable-rate irrigation, fertilization, and zone-based management strategies (Mucalo et al., 2024). This management approach helps reduce internal variability in resource supply and minimizes uneven berry development caused by water or nutrient stress.

Internet of Things (IoT)-based fertigation systems represent an important tool for precise regulation of berry uniformity. These systems automatically adjust irrigation volumes and nutrient solution compositions based on soil moisture, irrigation water quality, nutrient status, and plant demand, thereby maintaining stable water and nutrient supply during the berry enlargement stage. Smart fertigation platforms such as NutriBalance can automatically calculate optimal nutrient formulations based on water source quality and reduce fertilizer input by approximately 40% while maintaining nutrient supply accuracy (Imbernón-Mulero et al., 2023). Furthermore, irrigation and fertilization decision models based on mathematical optimization and genetic algorithms can improve economic returns while reducing environmental impacts, demonstrating the dual benefits of intelligent water-fertilizer management.

Intelligent environmental monitoring systems can also support regulation during critical growth stages. Temperature, humidity, and wind speed during flowering affect pollination uniformity; water supply and canopy light conditions during berry enlargement influence developmental synchrony; and microenvironmental variation during veraison and ripening may affect sugar-acid accumulation and maturity consistency. By continuously monitoring these parameters and integrating predictive models for early warning and regulation, it is possible to reduce asynchronous berry development at the source. In the future, improvements in grape berry uniformity will increasingly rely on an intelligent closed-loop system of “perception-diagnosis-prediction-decision-execution.” This involves evaluating berry uniformity through machine vision, predicting risks through multi-source data modeling, implementing regulation via intelligent water-fertilizer and environmental control systems, and continuously optimizing models based on feedback data. Such a framework enables the transition from experience-based management to data-driven precision management, providing sustainable technical support for the production of high-quality table grapes.

## **7 Challenges and Future Perspectives**

### **7.1 Lack of a unified evaluation system**

Currently, research on grape berry uniformity still lacks a unified and standardized evaluation system. Existing grape trait description systems mainly focus on individual traits such as berry size, berry shape, cluster compactness, and cluster color, while the composite trait of “intra-cluster berry uniformity” lacks a dedicated definition and standardized quantitative framework. Different studies and germplasm databases often employ OIV descriptors, self-defined trait systems, or output indicators from high-throughput phenotyping platforms, leading to inconsistencies in indicator definitions, measurement methods, and grading standards. These discrepancies limit cross-study comparisons and hinder industrial application (García-Abadillo et al., 2024; Liu et al., 2024; Zhang et al., 2025).

In recent years, two-dimensional and three-dimensional high-throughput phenotyping technologies have enabled precise extraction of traits such as berry diameter, berry volume, cluster length, cluster width, berry number, and compactness. However, these technologies are primarily applied to cluster architecture or single morphological traits and have not yet formed a dedicated core indicator system specifically for berry uniformity evaluation. Therefore, future efforts should focus on developing a standardized multi-indicator evaluation framework, while retaining the simplicity of traditional methods, with key components including berry size CV, shape consistency, spatial distribution uniformity, and appropriate cluster compactness.

## 7.2 Insufficient understanding of molecular mechanisms

Although considerable progress has been made in QTL mapping, GWAS, and candidate gene identification for traits related to berry size, shape, and cluster structure, research on the genetic mechanisms underlying the composite trait of berry uniformity remains limited. Most existing studies focus on mean values of traits such as berry weight, length, diameter, cluster compactness, or berry number, whereas relatively few directly use intra-cluster variability (e.g., CV) or spatial uniformity as core phenotypic traits for genetic mapping (García-Abadillo et al., 2024; Thorat et al., 2024).

Previous studies have shown that berry size and cluster structure are controlled by multiple genes, with candidate genes involved in processes such as cell division, cell expansion, hormone signaling, cell wall modification, and stress responses (De Sousa Moreira et al., 2024; Meneses et al., 2025). Pangenome studies further indicate that complex traits in grapevine are influenced not only by single nucleotide polymorphisms (SNPs) but also by structural variations, and that integrating SNP and structural variation analyses can improve the explanation of heritability (Liu et al., 2024). However, how these genetic variations regulate developmental synchrony, intra-cluster resource allocation, and spatial distribution of uniformity remains largely unclear.

Future research should treat berry uniformity as an independent core phenotype and integrate multi-stage developmental phenotypes with transcriptomic, metabolomic, hormonal, and genomic data to elucidate the regulatory networks underlying synchronous and asynchronous berry development. At the same time, functional validation of key candidate genes should be strengthened to facilitate the transition from empirical selection to molecular design breeding for uniformity improvement.

## 7.3 Insufficient integration of multi-source data

With the advancement of high-throughput phenotyping, genomics, and intelligent agriculture technologies, grape uniformity research is entering a data-driven era. However, the integration of phenotypic, genomic, environmental, and management data remains insufficient. Most existing studies focus on a single data type, such as phenotypic measurement, QTL mapping, or GWAS, while integrated modeling of genotype  $\times$  environment  $\times$  management interactions is still limited. This restricts the comprehensive understanding of the stability of berry uniformity across years, regions, and management conditions (Herzog et al., 2025; Zhang et al., 2025).

The integration of high-density genotyping with high-throughput phenotyping has already demonstrated significant value in the study of complex traits such as berry morphology, quality, and stress resistance (Liu et al., 2024; Zhang et al., 2025). Meanwhile, genomic selection and multi-trait selection indices are becoming important approaches to improve breeding efficiency in grapevine (Bharati et al., 2023; Brault et al., 2024). However, in practical production, management variables such as thinning intensity, GA<sub>3</sub> application, water status, microclimate, and vine load are often not recorded and modeled alongside genetic and phenotypic data, limiting the development of predictive models and precision regulation strategies for berry uniformity.

In the future, it will be necessary to establish a multi-source data platform specifically for berry uniformity, standardize data collection protocols, and integrate phenotypic, genotypic, environmental, and management information. By leveraging machine learning, genomic prediction, and digital twin technologies, dynamic predictive models can be developed. On this basis, uniformity-related traits can be incorporated into ideotype design and multi-trait selection indices, enabling the coordinated improvement of high uniformity, high yield, superior quality, and strong adaptability.

## 8 Concluding Remarks

Berry size, shape, and cluster structure are core components of grape quality, directly influencing consumer acceptance, suitability for fresh consumption or processing, and ultimately economic returns. With the advancement of high-throughput image analysis and machine vision technologies, researchers are now able to quantify berry and cluster traits at large scales and to elucidate the relationships between berry size, compactness, structural variation, and yield and quality attributes. Genomic and association studies have demonstrated that most berry-related traits, including size, weight, texture, and shape, are controlled by multiple genes. Nevertheless,

stable genetic loci and selection signals associated with fruit quality and morphology have been identified in grape germplasm, providing a foundation for targeted improvement. Collectively, these findings indicate that berry uniformity is a multidimensional trait with significant economic value that can be clearly defined, accurately measured, and genetically dissected.

Improving berry uniformity requires coordinated efforts in cultivar selection, molecular breeding, and field management. High-density genotyping, genome-wide association studies, and SNP array technologies, combined with high-throughput phenotyping, enable precise identification of genetic loci associated with berry size, cluster structure, and stress resistance, thereby supporting marker-assisted selection and future gene-editing approaches. Meanwhile, advanced breeding strategies such as genomic selection, clonal selection, and polyploidization, together with the integration of multi-omics data, can effectively accumulate favorable alleles and achieve the coordinated improvement of berry traits, fruit quality, and stress tolerance. In addition, digital phenotyping and remote sensing technologies can guide thinning practices, canopy management, and water-nutrient regulation, ensuring that cultivation practices are aligned with genetic potential and promoting synchronized berry development and stable fruit quality. Future progress in improving grape berry uniformity will depend on advancing standardized trait evaluation, precision management, and intelligent production systems. On the one hand, intelligent algorithms based on foundation models such as SAM and deep learning techniques (e.g., AS-SwinT and Mask R-CNN) have enabled automated berry segmentation, counting, and size measurement, laying the groundwork for standardized, machine-readable phenotypic data systems for uniformity and cluster compactness. On the other hand, integrating phenomics, genomics, and environmental monitoring data with machine learning models will enhance the prediction of yield and fruit quality, enabling precision management and selective harvesting based on spatial variability. Furthermore, combining variable-rate irrigation and fertilization technologies with optimization algorithms and IoT-based control systems can improve resource use efficiency and stabilize berry development. Overall, establishing unified data standards, shared trait ontologies, and integrated “perception-decision-execution” systems will be essential pathways for achieving stable improvements in grape berry uniformity and advancing intelligent production systems.

## Acknowledgments

The authors extend their sincere thanks to Mr. Cai for his valuable assistance of this manuscript.

## Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Acharya A., Gautam D.M., Khatri B., Poudel P., and Dahal K., 2025, Effect of gibberellic acid application on berry yield and quality of grapevine cv. Cabernet Sauvignon, *Crop Research*, 60 (1-2): 45-52.  
<https://doi.org/10.31830/2454-1761.2025.cr-1008>
- Alañón-Sánchez N., Ferradás Y., Bezrukov I., Weigel D., Carbonell-Bejerano P., and Ibáñez J., 2026, A complex reciprocal translocation is linked to reduced gamete viability in a loose-bunch grapevine somatic variant, *BMC Plant Biology*, 26: 348.  
<https://doi.org/10.1186/s12870-026-08212-7>
- Alshallash K.S., Fahmy M.A., Tawfeeq A.R., Baghdady G.A., Abdrabboh G.A., Hamdy A.E., and Kabsba E.M., 2023, GA<sub>3</sub> and hand thinning improves physical, chemical characteristics, yield and decrease bunch compactness of Sultanina grapevines (*Vitis vinifera* L.), *Horticulturae*, 9(2): 160.  
<https://doi.org/10.3390/horticulturae9020160>
- Bharati R., Sen M.K., Severová L., Svoboda R., and Fernández-Cusimamani E., 2023, Polyploidization and genomic selection integration for grapevine breeding: a perspective, *Frontiers in Plant Science*, 14: 1248978.  
<https://doi.org/10.3389/fpls.2023.1248978>
- Brault C., Segura V., Roques M., Lamblin P., Bouckennooghe V., Pouzalgues N., Cuntz C., Breil M., Frouin M., Garcin L., Camps L., Ducasse M., Romieu C., Masson G., Julliard S., Flutre T., and Le Cunff L., 2024, Enhancing grapevine breeding efficiency through genomic prediction and selection index, *G3: Genes, Genomes, Genetics*, 14(4): jkae038.  
<https://doi.org/10.1093/g3journal/jkae038>
- Cai Z.H., Li X.J., Forney C.F., Wang Y., Li B., and Xie Z.S., 2024, Effects of GA<sub>3</sub> treatments on fruit vascular structure and water transport of grape, *International Journal of Fruit Science*, 24(1): 200-218.  
<https://doi.org/10.1080/15538362.2024.2363628>

- Chen J.B., Guo Y.F., Hu H.C., Fang C.L., Wang L.R., Hu L.L., Lin Z.H., Zhang D.Y.D., Yang Z.Y., and Wu Y.Y., 2025, Regulation of cell metabolism and changes in berry shape of Shine Muscat grapevines under the influence of different treatments with the plant growth regulators gibberellin A3 and N-(2-chloro-4-pyridyl)-N'-phenylurea, *Horticulturae*, 11(10): 1160.  
<https://doi.org/10.3390/horticulturae11101160>
- Choi S., Ban S., and Choi C., 2023, The impact of plant growth regulators and floral cluster thinning on the fruit quality of 'Shine Muscat' grape, *Horticulturae*, 9(3): 392.  
<https://doi.org/10.3390/horticulturae9030392>
- Dauelsberg P., Matus J.T., Poupin M.J., Leiva-Ampuero A., Godoy F., Vega A., and Arce-Johnson P., 2011, Effect of pollination and fertilization on the expression of genes related to floral transition, hormone synthesis and berry development in grapevine, *Journal of Plant Physiology*, 168(14): 1667-1674.  
<https://doi.org/10.1016/j.jplph.2011.03.006>
- De Oliveira G.A., Francisco F.R., De Moura Y., Niederauer G., Fritsche-Neto R., De Souza A.P., and Furlan M., 2026, Genome-wide association study in a diverse grapevine collection provides insights into the genetic basis of berry size and cluster architecture traits, *PLOS One*, 21(3): e0343491.  
<https://doi.org/10.1101/2025.09.24.678328>
- De Sousa Moreira L., Clark M., Tabb A., Kam A., Londo J., Zou C., Sun Q., Van Zyl S., Prins B., DeLong J., Burhans A., Yang H., and Naegele R., 2024, Identification of novel quantitative trait loci associated with table grape fruit quality characteristics in a segregating population and transferability of existing quality markers, *Journal of the American Society for Horticultural Science*, 149(1): 50-60.  
<https://doi.org/10.21273/JASHS05334-23>
- Dhakad A.K., Patel V.B., Singh S.K., Verma M.K., Mishra G.P., Mhetre V., Kumar C., Singh R., Rajkumar S., and Gaikwad K., 2024, Pollen biology and metaxenic studies on early maturing grape (*Vitis vinifera*) genotypes, *The Indian Journal of Agricultural Sciences*, 94(10): 1094-1099.  
<https://doi.org/10.56093/ijas.v94i10.141725>
- Dobrei A., and Sala F., 2025, Morphological parameters in grape bunches and berries evaluation through multivariate analysis: case study on the 'Feteasca Regala' cultivar, *Life Science and Sustainable Development*, 6(1): 0ppjmz10.  
<https://doi.org/10.58509/0ppjmz10>
- Du W., and Liu P., 2023, Instance segmentation and berry counting of table grape before thinning based on AS-SwinT, *Plant Phenomics*, 5: 0085.  
<https://doi.org/10.34133/plantphenomics.0085>
- García-Abadillo J., Barba P., Carvalho T., Sosa-Zuñiga V., Lozano R., Carvalho H.F., Garcia-Rojas M., Salazar E., and Sánchez J.I., 2024, Dissecting the complex genetic basis of pre-and post-harvest traits in *Vitis vinifera* L. using genome-wide association studies, *Horticulture Research*, 11(2): uhad283.  
<https://doi.org/10.1093/hr/uhad283>
- Gharate P., Somkuwar R., Kumar A., Nilima G., Kakade P., and Karande P., 2025, Morphological and fruit variability of grape (*Vitis vinifera* L.) germplasm under subtropical condition, *Plant Science Today*, 12(sp1): 1-7.  
<https://doi.org/10.14719/pst.7521>
- Güler E., and Karadeniz T., 2023, Discrimination of an untouched autochthonous grapevine (*Vitis vinifera* L.) population by morphological markers and multivariate analyses, *Erwerbs-Obstbau*, 65: 2075-2084.  
<https://doi.org/10.1007/s10341-023-00926-4>
- Herzog K., Kicherer A., Malagol N., Trapp O., and Töpfer R., 2025, High-throughput phenotyping in grapevine breeding research: technologies and applications, *OENO One*, 59(3): 7-11.  
<https://doi.org/10.20870/oeno-one.2025.59.3.8458>
- Imbernón-Mulero A., Maestre-Valero J.F., Martínez-Álvarez V., García-García F., Jódar-Conesa F.J., and Gallego-Elvira B., 2023, Evaluation of an autonomous smart system for optimal management of fertigation with variable sources of irrigation water, *Frontiers in Plant Science*, 14: 1149956.  
<https://doi.org/10.3389/fpls.2023.1149956>
- Jewan S.Y.Y., Gautam D., Sparkes D.L., Singh A., Billa L., Cogato A., Murchie E.H., and Pagay V., 2024, Integrating hyperspectral, thermal, and ground data with machine learning algorithms enhances the prediction of grapevine yield and berry composition, *Remote Sensing*, 16(23): 4539.  
<https://doi.org/10.3390/rs16234539>
- Khalil U., Rajwana I.A., Razzaq K., Brecht J.K., and Sarkhosh A., 2023, The impact of fruit thinning on size and quality of fresh-market muscadine berries, *Journal of the Science of Food and Agriculture*, 104(4): 2198-2203.  
<https://doi.org/10.1002/jsfa.13105>
- Kim E., Lee C.H., Park S.M., Hong S.J., Kim S.H., and Kim G., 2023, A shine muscat grape berry detection and grape cluster compactness estimation for assessment of grape quality based on instance segmentation methods, *Journal of the ASABE*, 66(5): 1175-1185.  
<https://doi.org/10.13031/ja.15503>
- Lee L., Reynolds A., Dorin B., and Shemrock A., 2025, A feasibility study on utilizing remote sensing data to monitor grape yield and berry composition for selective harvesting, *Plants*, 14(1): 88.  
<https://doi.org/10.3390/plants14010088>
- Liu Z.J., Wang N., Su Y., Long Q.M., Peng Y.L., Guan S., Zhang F., Cao S., Wang X., Ge M.Q., Xue H., Ma Z.Y., Liu W.W., Xu X.D., Li C.C., Cao X.J., Ahmad B., Su X.G., Liu Y.T., Huang G.Z., Du M.R., Liu Z.Y., Gan Y., Sun L., Fan X.C., Zhang C., Zhong H.X., Leng X., Ren Y.P., Dong T.Y., Pei D., Wu X.Y., Jin Z.X., Wang Y.W., Liu C.H., Chen J.F., Gaut B.S., Huang S.W., Fang J.G., Xiao H., and Zhou Y.F., 2024, Grapevine pangenome facilitates trait genetics and genomic breeding, *Nature Genetics*, 56: 2804-2814.  
<https://doi.org/10.1038/s41588-024-01967-5>

- Luo L.F., Liu W.T., Lu Q.H., Wang J.H., Wen W.C., Yan D., and Tang Y.C., 2021, Grape berry detection and size measurement based on edge image processing and geometric morphology, *Machines*, 9(10): 233.  
<https://doi.org/10.3390/machines9100233>
- Meneses M., Muñoz-Espinoza C., Reyes-Impellizzeri S., Salazar E., Meneses C., Herzog K., and Hinrichsen P., 2025, Characterization of bunch compactness in a diverse collection of *Vitis vinifera* L. genotypes enriched in table grape cultivars reveals new candidate genes associated with berry number, *Plants*, 14(9): 1308.  
<https://doi.org/10.3390/plants14091308>
- Milišić K., Banjanin T., Vasić Z., Matijašević S., and Jančić R., 2025, Yield and cluster characteristics of grapevine varieties grown in the gene bank of the experimental vineyard Radmilovac, Serbia, *AgroReS*, 14: 41-47.  
<https://doi.org/10.63356/agrores.2025.005>
- Mirbod O., Yoder L., and Nuske S., 2016, Automated measurement of berry size in images, *IFAC-PapersOnLine*, 49(16): 79-84.  
<https://doi.org/10.1016/j.ifacol.2016.10.015>
- Mucalo A., Matić D., Morić-Španić A., and Čagalj M., 2024, Satellite solutions for precision viticulture: enhancing sustainability and efficiency in vineyard management, *Agronomy*, 14(8): 1862.  
<https://doi.org/10.3390/agronomy14081862>
- Roscher R., Herzog K., Kunkel A., Kicherer A., Töpfer R., and Förstner W., 2014, Automated image analysis framework for the high-throughput determination of grapevine berry sizes using conditional random fields, *Computers and Electronics in Agriculture*, 100: 148-158.  
<https://doi.org/10.1016/j.compag.2013.11.008>
- Sabir A., Kilinc S., and Sabir F., 2020, Qualitative and quantitative responses of early ripening table grape cultivars (*Vitis vinifera* L.) to pollination treatments under controlled growing condition, *Erwerbs-Obstbau*, 62: 75-80.  
<https://doi.org/10.1007/s10341-020-00499-6>
- Sharma S., Jr Munoz J.R., Torres-Lomas E., Lin J., Banayad H., Lupo Y., Nunez V., Gaspar A., Cantù D., and Diaz-Garcia L., 2025, Leveraging foundation models to dissect the genetic basis of cluster compactness and yield in grapevine, *Scientific Reports*, 16: 1434.  
<https://doi.org/10.1038/s41598-025-31531-y>
- Somogyi E., Kun Á., Lazar J., Bodor-Pesti P., and Sárdy D.A.N., 2021, Quantitative analysis of the berry size in grapevine cultivar 'Italia': digital image analysis of the grapevine, *Progress in Agricultural Engineering Sciences*, 17(S1): 53-60.  
<https://doi.org/10.1556/446.2021.30007>
- Thorat K.D., Upadhyay A., Samarth R.R., Machchhindra S.R., Jagtap M.A., Kushwaha K., Kesharwani P.K., Gaikwad P.S., Gawande D.N., and Somkuwar R.G., 2024, Genome-wide association analysis to identify genomic regions and predict candidate genes for bunch traits in grapes (*Vitis vinifera* L.), *Scientia Horticulturae*, 328: 112882.  
<https://doi.org/10.1016/j.scienta.2024.112882>
- Torres-Lomas E., Lado-Jimena J., Garcia-Zamora G., and Diaz-Garcia L., 2024, Segment anything for comprehensive analysis of grapevine cluster architecture and berry properties, *Plant Phenomics*, 6: 0202.  
<https://doi.org/10.34133/plantphenomics.0202>
- Trivedi M., Zhou Y., Moon J., Meyers J., Jiang Y., Lu G., and Vanden Heuvel J.E., 2023, A preliminary method for tracking in-season grapevine cluster closure using image segmentation and image thresholding, *Australian Journal of Grape and Wine Research*, 2023(2): 1-12.  
<https://doi.org/10.1155/2023/3923839>
- Upadhyaya P., Karkee M., Kshetri S., and Paudel A., 2023, Automated lag-phase detection in wine grapes using a mobile vision system, *Smart Agricultural Technology*, 6: 100381.  
<https://doi.org/10.1016/j.atech.2023.100381>
- Zhang Y.Y., Wang Y.J., Henke M., Carbonell-Bejerano P., Wang Z.M., Bert P.F., Wang Y., Li H.Y., Kong J.H., Fan P.G., Dai Z.W., and Liang Z.C., 2025, Integrating dense genotyping with high-throughput phenotyping empowers the genetic dissection of berry quality and resilience traits in grapevine, *Advanced Science*, 12(20): e2412587.  
<https://doi.org/10.1002/advs.202412587>

**Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Feature Review

## Open Access

## Yield-Related Traits in Cucumber (*Cucumis sativus*) and Their Variation and Formation

Wenjing Xu<sup>1</sup>, Mengting Luo<sup>2</sup> ✉<sup>1</sup> Jiande Riyue Agricultural Development Co., Ltd., Jiande, 311616, Zhejiang, China<sup>2</sup> Institute of Life Sciences, Jiyang College of Zhejiang A&F University, Zhuji, 311800, Zhejiang, China✉ Corresponding email: [mengting.luo@jicacat.org](mailto:mengting.luo@jicacat.org)Plant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0014](https://doi.org/10.5376/pgt.2026.17.0014)

Received: 10 May, 2026

Accepted: 12 Jun., 2026

Published: 25 Jun., 2026

**Copyright** © 2026 Xu and Luo, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Preferred citation for this article:**Xu W.J., and Luo M.T., 2026, Yield-related traits in cucumber (*Cucumis sativus*) and their variation and formation, Plant Gene and Trait, 17(3): 197-215 (doi: [10.5376/pgt.2026.17.0014](https://doi.org/10.5376/pgt.2026.17.0014))

**Abstract** This study explored the yield-related traits of cucumber and their variation and formation mechanisms. The characteristics and interrelationships of major yield components, including vegetative growth traits, reproductive growth traits, population structure, and stress-resistance traits, were systematically analyzed. The effects of natural germplasm resources, artificial selection, and molecular-level genetic variation on the formation of cucumber yield traits were also discussed. Furthermore, the physiological mechanisms underlying yield formation, including photosynthesis and assimilate accumulation, hormonal regulation, and environmental factor regulation, were elucidated. Advances in functional genes related to fruit morphology, parthenocarpy, organ size, and stress resistance, as well as signal transduction networks and multi-omics studies, were summarized. In addition, the roles of cultivation technologies, such as greenhouse environmental regulation, water and fertilizer management, and cultivation pattern optimization, in cucumber yield formation were reviewed. Combined with recent advances in marker-assisted selection, genomic selection, CRISPR/Cas gene editing, high-throughput phenotyping, and artificial intelligence technologies, the current challenges and future breeding directions for high-yield cucumber research were further discussed. This review provides a theoretical reference for breeding new cucumber varieties with high yield, superior quality, stable production, and stress resistance, as well as for optimizing efficient cultivation techniques.

**Keywords** Cucumber; Yield traits; Genetic variation; Molecular regulation; High-yield breeding

### 1 Introduction

Cucumber (*Cucumis sativus* L.) is one of the most widely cultivated and economically important vegetable crops in the Cucurbitaceae family worldwide, particularly in Asia and Europe, and is also an important vegetable crop in both protected horticulture and open-field cultivation systems. Cucumber is characterized by a short growth cycle, strong adaptability, and high yield potential. Its fruits are crisp, juicy, and rich in nutrients, giving the crop high market demand and economic value. In recent years, with the rapid development of protected agriculture and the modern horticultural industry, cucumber has become one of the largest vegetable crops cultivated under protected conditions in China, and China accounts for a dominant proportion of global cucumber production. Meanwhile, consumers have placed higher demands on cucumber commodity quality, nutritional quality, and year-round stable supply, making high yield, superior quality, stress resistance, and adaptability to mechanized production important objectives in modern cucumber breeding (Dey et al., 2023). However, under the increasing pressures of global climate change and resource limitations, abiotic stresses such as drought, salinity, high temperature, and low temperature, as well as biotic stresses including downy mildew and viral diseases, significantly affect cucumber growth, development, and yield stability (Serhiienko et al., 2025). Therefore, in-depth studies on yield-related traits and their formation mechanisms are of great significance for improving cucumber productivity and promoting the sustainable development of protected horticulture.

Cucumber yield is a typical complex quantitative trait whose formation is jointly regulated by genetic background, physiological metabolism, environmental conditions, and cultivation management practices. In general, cucumber yield is mainly determined by fruit number per plant, single fruit weight, and effective plant number per unit area, while agronomic traits such as fruit length, fruit diameter, female flower number, branch number, and vine length

are also closely associated with yield formation (Patidar et al., 2024). Correlation and path coefficient analyses have shown that fruit number, single fruit weight, and fruit size exert significant direct positive effects on yield per plant or yield per unit area, making them key factors influencing cucumber yield performance (Kaur et al., 2024; Lnu et al., 2025). In addition, the coordination between vegetative and reproductive growth directly affects yield formation. For example, plant height, leaf area, root activity, and photosynthetic capacity influence dry matter accumulation and assimilate transport, whereas female flower ratio, fruit set rate, and fruit enlargement rate determine yield formation capacity. Numerous studies have demonstrated that many cucumber yield-related traits possess high heritability and considerable genetic advance, indicating strong potential for genetic improvement and selection value (Negi et al., 2025).

However, cucumber yield-related traits also exhibit strong environmental sensitivity, with substantial differences observed among ecotypes, cultivation methods, and planting seasons (Lnu et al., 2025). Environmental factors such as temperature, light, water and fertilizer management, and cultivation conditions under protected or open-field systems can affect yield trait expression by regulating plant physiological and metabolic processes. Moreover, stresses such as high temperature, drought, salinity, and diseases interact complexly with genotypes, further influencing flowering, fruit set, and fruit development, thereby increasing the complexity of high-yield formation mechanisms in cucumber (Serhiienko et al., 2025). With the rapid development of molecular biology and omics technologies, studies on cucumber yield formation mechanisms have gradually expanded from traditional phenotypic observation and genetic analysis to genomics, transcriptomics, metabolomics, and epigenetic regulation. Studies have shown that cucumber yield formation is closely associated not only with physiological processes such as photosynthesis, hormonal regulation, carbon and nitrogen metabolism, and source-sink relationships, but also with the coordinated actions of multiple key genes and complex regulatory networks. For example, genes related to sex determination regulate female flower formation and fruit-setting ability, whereas fruit development-related genes directly affect fruit size, fruit shape, and single fruit weight (Dey et al., 2023). In recent years, technologies such as QTL mapping, candidate gene mining, genome-wide association studies (GWAS), genetic transformation, and CRISPR/Cas gene editing have been widely applied in studies of important agronomic traits in cucumber, providing new theoretical foundations and technical support for elucidating the genetic basis and molecular regulatory mechanisms of cucumber yield-related traits.

This study aims to explore cucumber yield-related traits and their variation and formation mechanisms by systematically reviewing research progress on genetic variation patterns, heritability, correlations, and gene effects of yield-related traits in different cucumber germplasm resources, landraces, breeding materials, and segregating populations. Due to differences in research materials, ecological environments, and evaluation criteria, systematic integration and unified understanding among different studies are still lacking. Considering the complexity of cucumber yield formation mechanisms and their importance in modern protected horticulture, this review focuses on summarizing the classification and characteristics of cucumber yield-related traits, patterns of genetic variation, physiological and molecular regulatory mechanisms, and the effects of cultivation techniques on yield formation. In addition, recent advances in multi-omics research and molecular breeding are discussed, and current research limitations and future development directions are analyzed and prospected. Through systematically summarizing relevant research achievements, this review aims to provide theoretical references for breeding new cucumber cultivars with high yield, superior quality, stable production, and stress resistance, as well as for optimizing efficient cultivation technologies, while also offering new insights into the genetic basis and regulatory mechanisms of complex quantitative traits in cucumber.

## **2 Yield-Related Traits in Cucumber**

### **2.1 Vegetative growth traits**

Vegetative growth traits in cucumber constitute an important foundation for determining plant vigor, canopy structure, and photosynthetic productivity. These traits mainly include vine length, internode length, branch number, leaf area, leaf number, root development, and photosynthetic capacity. Considerable phenotypic and genetic variation in these traits has been observed among different cucumber genotypes, landraces, and breeding materials, and some traits exhibit relatively high heritability, making them important indicators for high-yield

breeding and ideal plant architecture selection in cucumber (Negi et al., 2025). Previous studies have shown that traits such as vine length, branching ability, and leaf area contribute substantially in principal component analysis and cluster analysis and are important agronomic indicators for distinguishing different cucumber germplasm resources.

The vegetative growth process essentially reflects the plant's capacity to establish source-sink relationships and directly influences assimilate accumulation, transport, and subsequent fruit formation. In general, plants with strong vegetative vigor usually possess larger functional leaf areas and higher photosynthetic efficiency, thereby providing sufficient assimilates for female flower differentiation, fruit set, and fruit enlargement. Correlation and path coefficient analyses have demonstrated that traits such as vine length, leaf area, branch number, and leaf number are significantly positively correlated with yield per plant and yield per unit area and may indirectly promote yield formation by increasing fruit number and single fruit weight (Patidar et al., 2024; Lnu et al., 2025; Negi et al., 2025). However, excessive vegetative growth may also lead to excessive vine elongation, canopy overcrowding, and imbalance between vegetative and reproductive growth, thereby reducing fruit set rate and yield stability during later developmental stages.

Leaves and roots are the two core organs affecting vegetative growth efficiency in cucumber. Leaves serve as the primary site of photosynthesis, and their size, chlorophyll content, and photosynthetic rate directly determine canopy light-use efficiency and dry matter accumulation capacity. Roots influence sustained plant growth and stress adaptability by regulating water and mineral nutrient uptake. Under protected cultivation conditions, root vigor is closely associated with continuous fruiting ability, and well-developed root systems enhance tolerance to stresses such as drought, salinity, and high temperature. Therefore, coordinating aboveground canopy development with belowground root growth and maintaining balanced source-sink relationships are important physiological foundations for achieving high and stable cucumber yields.

## 2.2 Reproductive growth traits

Cucumber is a typical monoecious crop with separate male and female flowers, and its reproductive growth process is most directly associated with yield formation. Reproductive growth-related traits mainly include flowering time, node position of the first female flower, female-to-male flower ratio, number of female flowers, fruit set rate, fruit number, fruit length, fruit diameter, and single fruit weight. Among these traits, flowering time and first harvest time vary significantly among cucumber materials and exhibit moderate to high heritability in some studies, indicating considerable potential for improving early yield through selection for earliness (Negi et al., 2025). Studies have shown that total yield is negatively correlated with the periods from seedling emergence to flowering and from seedling emergence to fruiting, suggesting that earlier transition into the reproductive stage is beneficial for increasing cucumber yield (Serhiienko et al., 2025).

The capacity for female flower formation and fruit set are key determinants of cucumber fruiting potential. Generally, plants with a higher proportion of female flowers and lower node positions of the first female flower tend to exhibit stronger early fruiting ability. In modern cucumber breeding, gynoecious lines and parthenocarpic materials have been widely utilized in the development of high-yield hybrids. Gynoecious lines can significantly increase the number of female flowers, whereas parthenocarpic materials are capable of setting fruit without pollination, thereby improving yield stability under protected cultivation conditions (Dey et al., 2023; Kaur et al., 2024; Lnu et al., 2025). The genetic mechanisms underlying these traits are relatively complex, involving additive effects, dominant effects, and epistatic interactions among multiple genes. Therefore, integrated improvement strategies combining heterosis utilization and marker-assisted selection are generally required in breeding practice.

Fruit development traits constitute the core components of cucumber yield. Numerous studies have demonstrated that fruit number, fruit length, fruit diameter, and average single fruit weight are key determinants of yield per plant and yield per unit area (Patidar et al., 2024; Negi et al., 2025). These traits generally exhibit high phenotypic variation coefficients, heritability, and genetic advance, indicating substantial selection potential (Table 1) (Lnu et al., 2025). Meanwhile, plant hormones such as ethylene, auxin, and gibberellins play important regulatory roles in sex expression, fruit set, and fruit enlargement in cucumber. The elucidation of related functional genes and

regulatory networks has also provided important molecular foundations for understanding the mechanisms underlying high-yield formation in cucumber (Dey et al., 2023).

Table 1 Coefficient of variation, heritability and genetic advance of yield and quality traits (Adopted from Lnu et al., 2025)

Trait	Mean	CV (%)	GCV (%)	PCV (%)	H2 (b.s.) (%)	G.A (%)	G.A (as % mean)
G.	16.08	5.75	13.58	14.77	84.66	4.14	25.76
Fruit diameter	3.57	6.04	15.99	17.1	87.53	1.1	30.84
Fruit weight	157.67	9.88	27.45	29.13	88.81	84.03	53.29
Pistil length	3.74	4.38	11.79	12.58	87.8	0.85	22.76
Internodal length	10.19	3.07	20.86	21.09	97.84	4.33	42.5
Vine length	328.52	4.13	25.65	25.98	97.45	171.33	52.15
Number of female flowers per node	1.21	7.24	26.44	27.43	92.89	0.63	52.49
Number of fruits set per node	0.15	14.95	34.36	37.41	84.38	0.09	65
Number of fruits per plant	14.81	4.98	35.43	35.77	98.12	10.71	72.29
Fruit setting percentage	13.34	19.28	2.36	39.83	0.35	0.04	0.29
Fruit yield per plant	2.14	14.62	24.61	27.64	79.35	0.97	45.18
TSS	2.66	9.3	12.76	15.82	65.06	0.56	21.21
Flesh to seed cavity ratio	1.94	14.94	12.47	19.56	40.67	0.32	16.38
Water content	97.51	0.78	0.17	0.79	4.68	0.08	0.08
Vitamin C	3.08	9.24	16.85	19.19	77	0.94	30.45

### 2.3 Population structure and stress-resistance-related traits

Population structure is an important factor affecting light energy utilization efficiency and yield per unit area in cucumber populations, mainly involving plant architecture, planting density, spatial leaf distribution, and canopy ventilation and light transmission capacity. A reasonable population structure can improve the uniformity of light interception within the canopy and enhance photosynthetic efficiency, thereby promoting dry matter accumulation and fruit formation. In contrast, excessively high planting density or overly luxuriant plant architecture can intensify competition among plants, leading to canopy closure, premature senescence of lower leaves, and increased disease incidence. Under protected cultivation and high-density planting conditions, optimizing population structure, improving the field microclimate, and enhancing the canopy light environment are important agronomic measures for increasing continuous fruiting ability and yield per unit area in cucumber.

Different cucumber germplasm resources exhibit significant differences in vegetative growth vigor, fruit characteristics, environmental adaptability, and yield potential, and they often form distinct groups in cluster analyses (Serhiienko et al., 2025). In evaluations of landraces and breeding materials, researchers have identified several groups showing superior performance in fruit number, fruit size, and overall yield. These materials can serve as important resources for selecting high-yield parental lines, genetic mapping, and heterosis utilization. Such rich genetic diversity is not only beneficial for the improvement of high-yield traits but also provides important material foundations for studying the genetic mechanisms underlying complex quantitative traits in cucumber (Kaur et al., 2024).

Stress-resistance-related traits are important guarantees for achieving high and stable cucumber yield. Cucumber is sensitive to low temperature, high temperature, drought, and salt stress, and adverse environmental conditions can lead to reduced photosynthesis, abnormal floral organ development, and decreased fruit-setting rate, thereby significantly affecting yield and fruit quality. In recent years, traits such as drought tolerance, salt tolerance, and resistance to viral diseases have gradually been incorporated into evaluation systems for high-yield cucumber breeding. Quantitative genetic studies have shown that some stress-related traits and disease indices possess certain levels of heritability and may be genetically associated with flowering habits and fruit development traits, indicating the potential for the synergistic improvement of stress resistance and high yield. With the development of marker-assisted selection, genomic selection, and gene-editing technologies, important genes and QTLs related to low-temperature tolerance, salt tolerance, disease resistance, and fruit development have gradually been

identified, providing new technical support for breeding cucumber cultivars with high yield, multiple stress resistance, and adaptability to complex environmental conditions (Dey et al., 2023).

### 3 Genetic Variation of Cucumber Yield Traits

#### 3.1 Variation in natural germplasm resources

Cucumber germplasm resources contain abundant genetic variation and constitute an important genetic foundation for breeding programs targeting high yield, superior quality, and stress resistance. Owing to long-term natural evolution, ecological adaptation, and geographical differentiation, cucumber materials from different regions and ecotypes have developed significant differences in plant architecture, flowering habits, fruit morphology, fruiting ability, and environmental adaptability. Extensive germplasm evaluations have demonstrated broad natural phenotypic diversity in yield-related traits such as fruit length, fruit diameter, fruit number per plant, single fruit weight, and yield per plant in cucumber (Tadkal et al., 2024). This rich natural variation not only reflects the long-term ecological adaptation and cultivation history of cucumber but also provides abundant allelic resources for the genetic improvement of high-yield traits.

Landraces are important sources of natural variation in cucumber and represent core germplasm types for maintaining genetic diversity. Previous studies have shown that different landraces exhibit high coefficients of variation and broad-sense heritability for traits such as yield per plant, fruit number per plant, single fruit weight, and vine-related characteristics, indicating strong selection potential and substantial genetic control over these traits (Tadkal et al., 2024). Compared with modern commercial cultivars, traditional landraces may possess certain limitations in terms of fruit uniformity and adaptability to protected cultivation systems; however, they still exhibit important breeding value in stress adaptability, continuous fruiting ability, and yield stability. Therefore, landrace germplasm resources are not only important components of cucumber genetic diversity but also valuable materials for mining superior yield-related genes.

With the establishment of global germplasm repositories and genomic databases, the population genetic structure of natural variation in cucumber has gradually been elucidated (Figure 1). The 1 234 cucumber accessions conserved in the U.S. National Plant Germplasm System can be classified into three major genetic subpopulations, which differ significantly in geographical origin and horticultural type. Further genome-wide SNP analyses and core germplasm construction have revealed abundant and structured genetic diversity within cucumber germplasm resources, particularly among East Asian materials, which exhibit distinct characteristics in fruit length and fruit shape index (Lin et al., 2024). The extensive genetic variation present in natural germplasm resources provides important genetic sources for breeding new cucumber cultivars with high yield, superior quality, and stress resistance and also establishes a material foundation for elucidating the genetic basis of complex quantitative traits.

#### 3.2 Variation formed through artificial selection

Artificial selection has been an important driving force promoting the continuous optimization of cucumber yield traits and the differentiation of cultivation types. During long-term domestication, cultivation, and modern breeding processes, breeders have continuously selected materials with high yield, early maturity, high female flower ratio, strong continuous fruiting ability, and adaptability to protected cultivation, resulting in significant changes in plant architecture, flowering habits, fruit morphology, and yield components. Modern greenhouse cucumbers generally exhibit characteristics such as shorter internodes, higher degrees of femaleness, stable fruit set, and strong continuous fruiting ability, all of which are the result of long-term directional selection and genetic fixation. Artificial selection has not only altered the phenotypic range of yield-related traits in cucumber but also significantly affected the frequency distribution of related alleles and overall genome structure.

QTL mapping studies based on wild and cultivated cucumber populations have shown that multiple important loci associated with flowering time, fruit size, fruit number, and fruit weight per plant have undergone continuous selection during domestication. Some QTL regions exhibit clear signatures of domestication and breeding selection. Notably, certain alleles derived from wild or semi-wild materials still exert positive transgressive effects on fruit size and fruiting ability, indicating that wild germplasm resources retain abundant superior genetic

variation that has not yet been fully utilized. Therefore, the genetic complementarity between wild germplasm and modern cultivated varieties provides new directions for high-yield cucumber breeding.

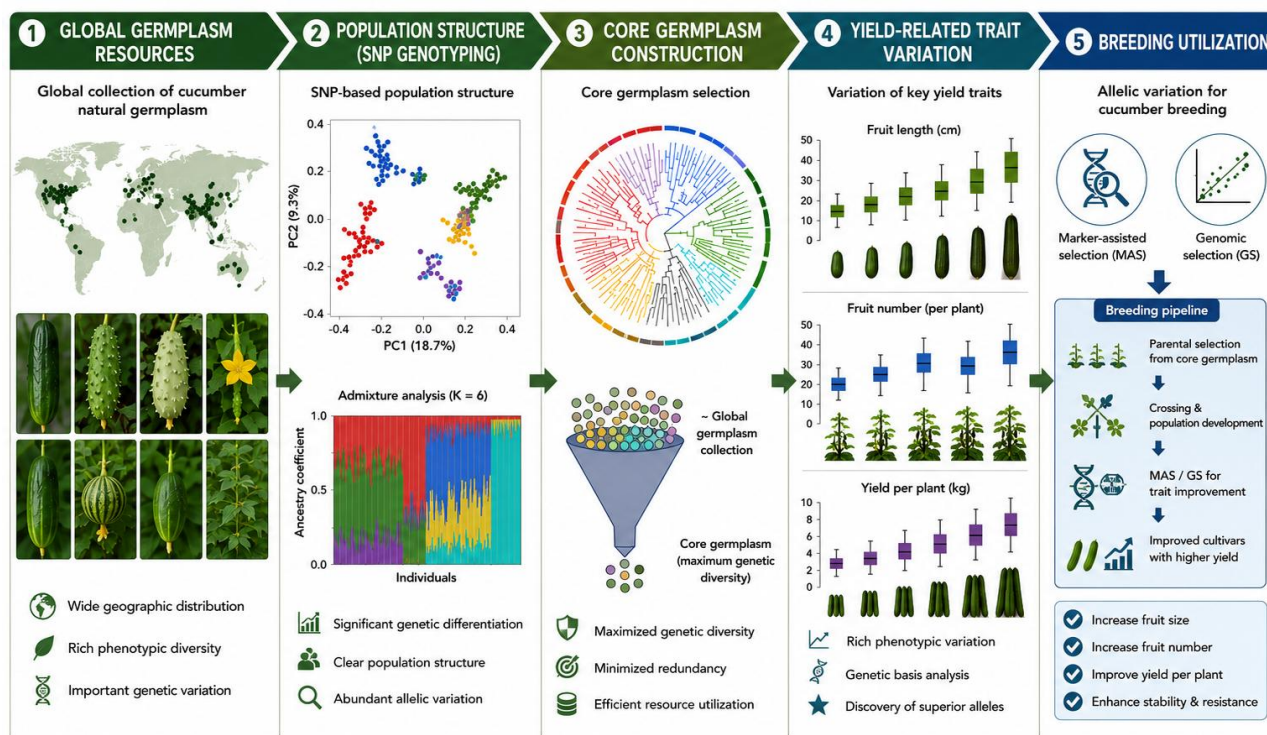


Figure 1 Genetic diversity and yield-related trait variation in cucumber natural germplasm resources

Clear signatures of artificial selection are also evident in modern breeding populations. Genome analyses of elite inbred lines from different breeding backgrounds have revealed significant differentiation in several important QTL regions associated with earliness, fruit size, and productivity, reflecting continuous reinforcement selection for high-yield-related traits during breeding processes. In addition, multi-environment evaluations and hybrid combination analyses have demonstrated substantial genetic variation among breeding materials, with some hybrid combinations exhibiting significantly higher total yield and fruit number than ordinary materials and even outperforming commercial control cultivars (El-Remaly and Shehata, 2023; Serhienko et al., 2025). In recent years, the utilization of heterosis, marker-assisted selection, and pyramiding of complex traits have further improved the efficiency of cucumber yield improvement (Shukla et al., 2025). Artificial selection has not only shaped the genetic variation patterns of modern cucumber yield traits but has also driven the continuous evolution of cucumber from traditional landraces to protected-cultivation-adapted and high-yield, high-quality cultivars.

### 3.3 Molecular-level genetic variation

With the development of genomics and high-throughput sequencing technologies, studies on cucumber yield traits have gradually progressed from traditional phenotypic and quantitative genetic analyses to the molecular level. Molecular variations such as single nucleotide polymorphisms (SNPs), insertions/deletions (InDels), copy number variations, and structural variations constitute important genetic bases underlying differences among cucumber materials in fruit size, flowering time, female flower ratio, fruiting ability, and stress resistance. Numerous studies have demonstrated that cucumber yield traits are typical complex quantitative traits generally regulated by multiple loci with small or moderate effects and jointly influenced by environmental factors and gene interactions.

Based on genome-wide SNP data and high-density genetic maps, researchers have identified important QTLs and significant SNP loci associated with fruit length, fruit diameter, fruit shape, and fruit weight in multiple populations. Genome-wide association studies (GWAS) have shown that many significant loci related to fruit size and fruit shape are clustered within known QTL regions, suggesting that cucumber yield-related traits possess relatively stable genetic foundations (Lin et al., 2024). Meanwhile, high-density SNP genetic maps constructed

using SLAF-seq technology have further improved QTL mapping resolution, and some major-effect loci can explain a relatively large proportion of phenotypic variation, thereby providing important foundations for candidate gene mining and marker-assisted selection. However, because of strong interactions among populations, environments, and cultivation conditions, some QTLs still exhibit environmental dependency and insufficient stability.

In addition to SNPs and InDels, structural variation and pan-genome studies have further revealed the deeper genetic basis underlying complex cucumber yield traits. Pan-genome analyses have identified large numbers of structural variations in cucumber, some of which are closely associated with important agronomic traits such as flowering time, fruit surface characteristics, and root development. Large chromosomal rearrangements and structural variations can directly influence the expression of key genes and may also regulate multiple yield components through linkage effects and pleiotropy. To date, numerous functional genes and QTLs related to fruit size, fruit number, and flowering time have been identified in cucumber, and molecular marker systems applicable to breeding are gradually being established. At the same time, advances in transcriptomics, metabolomics, and functional genomics, as well as the application of genetic transformation and CRISPR/Cas gene-editing technologies, have provided new technical approaches for elucidating the mechanisms underlying high-yield formation and for the targeted improvement of key yield traits in cucumber.

## **4 Physiological Mechanisms of Cucumber Yield Formation**

### **4.1 Photosynthesis and assimilate accumulation**

Cucumber yield formation is highly dependent on leaf photosynthetic carbon assimilation capacity and the efficiency of assimilate allocation to fruits. Through photosynthesis, leaves fix CO<sub>2</sub> and synthesize carbohydrates, thereby providing the material and energy sources required for vine growth, root development, female flower formation, and fruit enlargement. Therefore, photosynthesis constitutes the core physiological basis for dry matter accumulation and yield formation in cucumber. Studies have shown that increasing the activity of enzymes related to photosynthetic carbon metabolism can significantly enhance net photosynthetic rate and biomass accumulation in cucumber, indicating that improving photosynthetic biochemical capacity is an important approach for promoting greenhouse cucumber growth and increasing yield. In addition, cucumber fruit is a typical strong “sink” organ, and its continuous enlargement process is highly dependent on assimilate supply. Therefore, high-yield formation depends not only on the photosynthetic capacity of source organs but also on the competitive strength of sink organs and the coordination between source and sink.

The transport and distribution of assimilates in cucumber exhibit typical “source-sink-flow” regulatory characteristics. Functional leaves serve as the major “source” organs, fruits act as the primary “sink” organs, and phloem transport constitutes the pathway for assimilate translocation. Previous studies have demonstrated that cucumber fruit growth dynamics are closely related to assimilate supply capacity and sink activity, while temporary assimilate storage and redistribution can partially buffer imbalances between source and sink relationships. When vegetative and reproductive growth remain coordinated, assimilates are preferentially transported to fruits, thereby promoting fruit set and fruit enlargement. In contrast, excessive vegetative growth, premature leaf senescence, or reduced root vigor decreases assimilate transport efficiency, ultimately restricting fruit development and reducing yield. Therefore, maintaining balanced source-sink relationships is an important physiological basis for achieving high cucumber yield.

Environmental conditions can significantly influence photosynthesis and assimilate accumulation in cucumber. Increasing CO<sub>2</sub> concentration enhances carbon assimilation capacity, increases soluble sugar content, carbon-to-nitrogen ratio, and fruit dry matter accumulation, and promotes greater assimilate allocation to fruits, thereby improving fruit weight and yield. Root-zone temperature and nitrogen supply further affect photosynthetic acclimation and carbon-nitrogen balance. Increasing root-zone temperature and coordinating NO<sub>3</sub><sup>-</sup> supply can alleviate photosynthetic inhibition under elevated CO<sub>2</sub> conditions and improve leaf nitrogen content and photosynthetic efficiency. In addition, canopy structure and light environment are important factors influencing canopy light-use efficiency. Under low-light conditions in winter, appropriate leaf area and supplemental lighting

can improve canopy light interception and CO<sub>2</sub> assimilation efficiency, thereby increasing female flower number, fruit dry matter accumulation, and total yield (Koo et al., 2024).

#### 4.2 Hormonal regulatory mechanisms

Plant hormones are important endogenous signals regulating cucumber yield formation. By coordinating vegetative growth, reproductive development, and fruit formation processes, hormones directly influence cucumber fruiting ability and yield level. Cucumber is a typical monoecious crop with separate male and female flowers, and its female flower formation, fruit set stability, and fruit enlargement processes are highly dependent on hormonal balance. Hormones such as auxin, gibberellins (GA), ethylene, cytokinins, and abscisic acid (ABA) not only participate in sex differentiation and floral organ development but also influence fruit formation by regulating cell division, cell expansion, and assimilate transport (Baral et al., 2025). Therefore, hormonal regulation essentially constitutes an important physiological basis for “sink formation” and yield establishment in cucumber.

Ethylene is an important regulator of sex expression in cucumber and generally promotes female flower formation and increases the proportion of female flowers. Materials with stronger ethylene synthesis capacity often exhibit lower node positions of the first female flower and higher fruit-setting ability, thereby showing stronger early fruiting potential. Meanwhile, auxin and GA play key roles during fruit development. Auxin mainly participates in ovary development, cell division, and cell expansion, whereas GA regulates vine elongation, floral organ formation, and early fruit growth. Studies have shown that hormone signaling pathways mediated by Aux/IAA, ARF, and GA signaling components regulate cell proliferation and expansion during the early stages of fruit development, whereas ABA and ethylene levels gradually increase during fruit maturation (Baral et al., 2025). Dynamic changes in these hormonal signals collectively determine fruit formation efficiency and final yield performance.

Studies on cucumber parthenocarpy further demonstrate the important role of hormonal balance in maintaining stable fruit set and continuous fruiting. Materials with strong parthenocarpic ability generally exhibit higher cytokinin and GA levels but relatively lower ABA levels. Exogenous application of cytokinins, auxin, or GA<sub>4+7</sub> can induce parthenocarpic fruit formation in weakly parthenocarpic materials. These processes are accompanied by enhanced expression of cytokinin synthesis and response genes and reduced expression of ABA signaling receptors, indicating that higher levels of promotive hormones are beneficial for maintaining stable fruit set. In addition, hormones such as brassinosteroids, melatonin, and ethylene also participate in fruit formation regulation, while synergistic and antagonistic interactions among different hormonal pathways further increase the complexity of cucumber yield formation mechanisms.

#### 4.3 Environmental regulatory mechanisms

Cucumber yield formation is highly sensitive to environmental conditions. Factors such as temperature, light, CO<sub>2</sub> concentration, water availability, and root-zone environment can regulate yield formation by influencing photosynthesis, hormone metabolism, and reproductive development. Environmental factors not only determine plant assimilate production capacity but also affect female flower differentiation, fruit set, and fruit enlargement by regulating hormonal signaling pathways involving ethylene, auxin, GA, and ABA (Aparna et al., 2023). Under the background of climate change, stresses such as high temperature, drought, low light, and waterlogging often lead to reduced photosynthetic efficiency, imbalance between source and sink relationships, and restricted fruit development, thereby decreasing cucumber yield stability.

Temperature and light are key environmental factors affecting cucumber yield formation. Low-temperature conditions inhibit photosynthesis and root absorption capacity, impair floral organ development, and cause flower and fruit abortion, whereas high temperatures increase transpiration, reduce photosynthetic efficiency, and negatively affect gamete viability and fruit set rate (Aparna et al., 2023). Meanwhile, prolonged low-light conditions easily induce excessive elongation growth, reduce leaf functional capacity, and inhibit female flower differentiation, thereby restricting yield formation in greenhouse cucumber production. Supplemental lighting combined with CO<sub>2</sub> fertilization can effectively enhance photosynthetic source strength and increase female

flower number, fruit dry matter accumulation, and total yield (Koo et al., 2024). Furthermore, optimization of greenhouse light environments based on cucumber light response characteristics can further improve canopy light-use efficiency and greenhouse production efficiency.

Water availability and root-zone environment also significantly affect cucumber yield formation. Cucumber possesses a relatively shallow root system and is highly sensitive to changes in soil moisture. Water deficit inhibits photosynthesis and cell turgor, thereby restricting fruit enlargement, whereas excessive water accumulation can cause root hypoxia and disease occurrence, reducing CO<sub>2</sub> assimilation efficiency and PSII photochemical activity. Under protected cultivation conditions, coordinated management of CO<sub>2</sub> concentration, temperature, light, water, and fertilizer can effectively improve dry matter accumulation and fruit quality. Studies have shown that the combination of elevated CO<sub>2</sub> and appropriate temperature not only promotes the accumulation of sugars, organic acids, and amino acids but also enhances fruit heat tolerance under high-temperature conditions while maintaining yield stability (Du et al., 2024). Therefore, under climate change scenarios, combining precise environmental regulation in protected cultivation systems with the breeding of stress-resistant genotypes is an important strategy for improving cucumber yield stability and resource-use efficiency.

## **5 Molecular Mechanisms of Cucumber Yield Formation**

### **5.1 Yield-related functional genes**

Cucumber yield formation is a complex process jointly determined by multiple genes, loci, and regulatory pathways, involving vegetative growth, sex differentiation, floral organ formation, fruit set, fruit enlargement, organ size regulation, and stress responses. With the continuous advancement of cucumber genomics, fine QTL mapping, and functional genomics studies, an increasing number of functional genes and major-effect QTLs closely associated with yield component traits have been identified, providing an important foundation for elucidating the mechanisms underlying high-yield formation and for conducting molecular design breeding in cucumber. Yield-related functional genes in cucumber mainly affect final yield through the regulation of fruit morphology, organ size, source-sink relationships, and stress adaptability.

Fruit length, fruit diameter, fruit shape index, and single fruit weight are core traits constituting cucumber yield and are also the major focus of current functional gene studies. The major fruit shape QTL FS5.2 has been finely mapped to a 95.5-kb interval, and its candidate gene, CsCRC, encodes a YABBY transcription factor that influences fruit morphology by regulating cell length and cell layer number, resulting in the short and thick fruit phenotype of Xishuangbanna cucumber. CsTRM5, a homolog of tomato TRM5, controls the round-fruit locus FS2.1, and loss of its function alters the direction of cell division and weakens cell expansion, thereby producing spherical fruits (Xie et al., 2023). In addition, CsSUN, located at the FS1.2 locus, is a homolog of the SUN gene and contains a 161-bp deletion in round-fruit types, significantly affecting the fruit length-to-diameter ratio. These studies indicate that cell division orientation, cell layer number, and cell expansion processes are important cytological bases determining cucumber fruit morphology and single fruit weight.

In addition to fruit morphology-related genes, some functional genes can influence yield by regulating leaf size, photosynthetic source strength, organ development, and continuous fruiting ability. The R2R3-MYB transcription factor CsRAXs suppresses leaf size and continuous fruiting ability by promoting auxin glycosylation. Its triple mutants exhibit larger leaves, higher fruit number per plant, and higher yield, whereas overexpression lines display reduced leaf area and fruit number (Chen et al., 2024). The littleleaf (LL) gene encodes a WD40 protein that restricts cucumber organ size and is closely associated with the expression of organ size regulators and hormone-related genes (Xu et al., 2023). SF1 encodes a RING-type E3 ubiquitin ligase that regulates fruit elongation through ethylene-dependent cell division processes (Xie et al., 2023). To date, 81 simply inherited genes or major-effect QTLs and 322 QTLs associated with 42 quantitative traits have been systematically identified in cucumber, many of which are related to fruit size, fruit shape, fruit number, and other yield component traits, and diagnostic markers applicable to breeding have been developed. In addition, heat shock proteins, antioxidant enzymes, hormone signaling factors, and members of transcription factor families such as

MYB, NAC, WRKY, bHLH, and ERF/AP2 can maintain fruit-setting ability and yield stability by regulating photosynthetic stability, hormonal balance, and stress responses.

## 5.2 Signal transduction and regulatory networks

Cucumber yield formation depends not only on the effects of individual functional genes but also on the coordinated regulation of complex signal transduction and transcriptional regulatory networks. Plant hormone signaling serves as an important bridge linking functional genes, cellular development, and yield phenotypes. Hormones such as ethylene, auxin, gibberellins, cytokinins, and abscisic acid participate in sex differentiation, floral organ formation, parthenocarpy, fruit elongation, and organ size regulation through synergistic or antagonistic interactions. The dynamic balance among different hormonal pathways determines fruit set stability, fruit development efficiency, and final yield level in cucumber. Therefore, shifting from single-gene effects to integrated regulatory networks is an important direction for understanding the molecular mechanisms of cucumber yield formation.

Within the fruit morphology regulatory network, CsCRC acts as an important hub connecting transcriptional regulation, hormone signaling, and cellular development. FS5.2/CsCRC regulates auxin content and the expression of auxin-responsive genes, linking YABBY-mediated transcriptional regulation with auxin-regulated cell division and cell expansion processes. Meanwhile, gibberellin signaling also participates in CsCRC-dependent fruit elongation regulation. Further studies have shown that CsPIF1 can directly activate CsCRC transcription, whereas the YABBY protein CsINO and the bHLH protein CsSPT1 physically interact with CsCRC, thereby integrating light signaling, reproductive development, and auxin regulation into the fruit elongation process (Xie et al., 2023). Transcriptome analyses of near-isogenic lines with long and short fruits further revealed significant activation of microtubule-, cell cycle-, and hormone-related genes, while multiple transcription factor families, including MYB, bHLH, NAC, and ERF/AP2, were involved in fruit length regulation (Che et al., 2023; Xing et al., 2023).

Parthenocarpy is an important mechanism enabling greenhouse cucumber to maintain stable yield under insufficient pollination, low-light conditions, or environmental stresses, and its formation also depends on precise hormonal network regulation. Materials with strong parthenocarpic ability generally exhibit higher cytokinin and gibberellin levels and lower ABA levels, accompanied by upregulated expression of cytokinin biosynthesis and response genes as well as auxin signaling components. Exogenous application of cytokinins, auxin, or GA<sub>4+7</sub> can induce parthenocarpic fruit formation in weakly parthenocarpic materials. The GA biosynthesis gene CsGA20OX1 is considered a positive regulator of parthenocarpy, and both its expression level and related hormone contents are significantly increased in induced parthenocarpic fruits (Meng et al., 2026). In addition, LL-mediated organ size regulation is associated with auxin and cytokinin pathways, whereas CsTRM5-mediated fruit shape variation involves ABA accumulation and changes in the expression of genes related to cell wall formation and ABA signaling. Exogenous ABA can reduce fruit elongation by inhibiting cell expansion (Figure 2) (Xie et al., 2023; Xu et al., 2023). These findings indicate that cucumber yield formation results from the combined action of multi-level regulatory networks involving hormone signaling, transcriptional regulation, cell division, and cell expansion.

## 5.3 Multi-omics studies and molecular breeding

The development of multi-omics technologies has provided important tools for systematically elucidating the mechanisms underlying cucumber yield formation. Genomics can reveal genetic variation, structural variation, and candidate genomic regions among different materials; transcriptomics can analyze dynamic gene expression under different developmental stages and environmental conditions; and proteomics and metabolomics help clarify the relationships between yield formation and processes such as signal transduction, hormone metabolism, cell wall formation, sugar metabolism, and amino acid metabolism. Through integrated multi-omics analyses, researchers can construct molecular regulatory networks underlying high-yield formation in cucumber at multiple levels, including “gene-transcript-protein-metabolite-phenotype”.

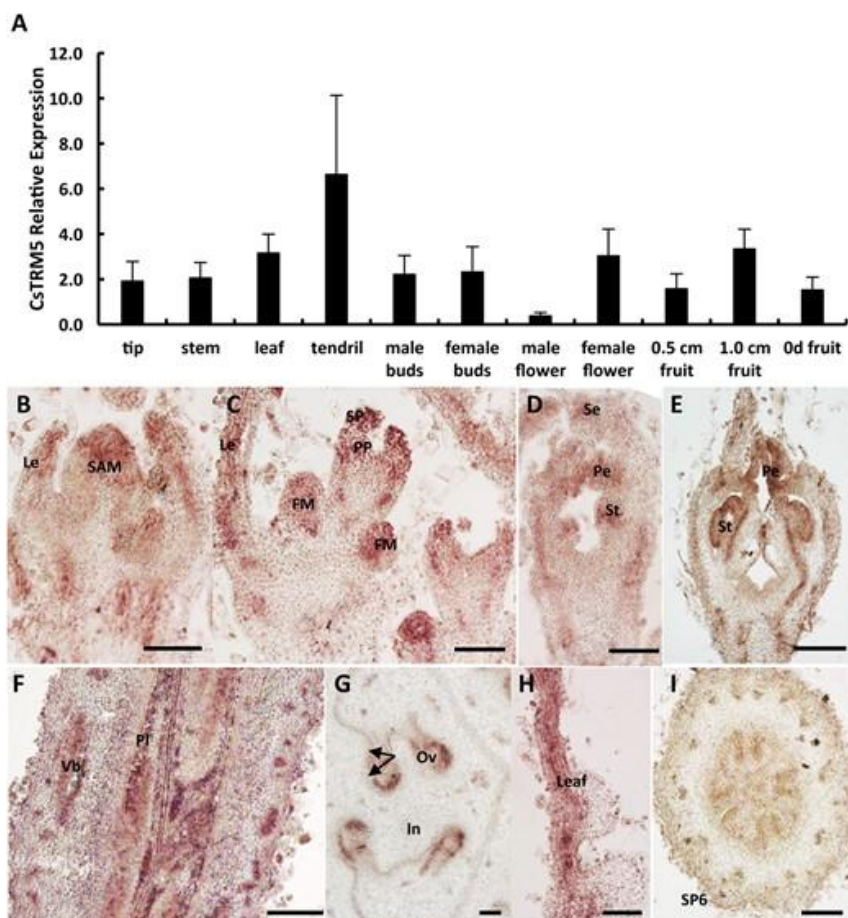


Figure 2 Expression pattern analysis of CsTRM5 in cucumber (Adopted from Xie et al., 2023)

Image caption: A Expression levels of CsTRM5 in different organs of 32X line detected by qRT-PCR. B-H In situ hybridization of CsTRM5 in cucumber shoot tip (B), leaf (H), and floral organs (C-G). I, Negative control of CsTRM5 sense probe in fruit cross-section. Le, leaf; SAM, shoot apical meristem; FM, floral meristem; SP, sepal primordium; PP, petal primordium; Se, sepal; Pe, petal; St, stamen; Vb, vascular bundle; Pl, placenta; Ov, ovule; In, integument. Scale bars: 100  $\mu$ m (Adopted from Xie et al., 2023)

In studies of fruit size and fruit shape, multi-omics approaches have demonstrated significant advantages. RNA-seq analysis of FS5.2 near-isogenic lines identified hundreds of differentially expressed genes mainly enriched in auxin biosynthesis and signaling, cell cycle regulation, and cell wall formation, thereby establishing the CsCRC-mediated regulatory network controlling fruit size and shape. Comparative transcriptome analyses of early fruits from materials with different fruit lengths identified 3,955 differentially expressed genes, including numerous microtubule-related genes, cell cycle genes, and members of various transcription factor families, providing genome-wide evidence for fruit length regulation. In addition, combined analyses using QTL-seq, bin-map resequencing, and RNA-seq identified multiple major-effect fruit length QTLs and differentially expressed genes related to auxin efflux carriers and hormone signaling, thereby providing important evidence for candidate gene mining (Che et al., 2023; Xing et al., 2023). In studies of parthenocarp and stable fruit set, integrated transcriptomic and metabolomic analyses linked the GA biosynthesis gene CsGA20OX1 and hormone pathways with GA-mediated parthenocarpic fruit development, indicating that multi-omics integration can effectively identify key regulators controlling stable fruit set and yield formation (Meng et al., 2026). Moreover, continuous omics studies in cucumber have identified numerous candidate genes associated with fruit length, locule formation, parthenocarp, and fruit quality traits, thereby providing theoretical support for marker-assisted selection, genomic selection, and molecular design breeding.

The development of molecular breeding technologies is promoting the transition of cucumber yield improvement from empirical selection to precision breeding. Marker-assisted selection (MAS) can rapidly identify superior materials using markers tightly linked to target traits; genomic selection (GS) can predict breeding values of

complex quantitative traits based on genome-wide markers; and CRISPR/Cas gene-editing technology enables precise modification of key functional genes. Currently, systematic gene and QTL databases, developed diagnostic markers, and continuously expanding functional gene studies are building molecular tool systems for pyramiding superior alleles associated with fruit size, organ size, parthenocarp, continuous fruiting ability, and stress resistance. In the future, integrated “pan-omics” platforms combining genomics, transcriptomics, proteomics, metabolomics, phenomics, machine learning, and gene editing will become an important direction for accelerating high-yield and stable-yield cucumber breeding (Roychowdhury et al., 2023).

## **6 Effects of Cultivation Techniques on Yield Formation**

### **6.1 Environmental regulation in protected cultivation**

Protected cultivation is an important production system for achieving high and stable cucumber yield. Its core advantage lies in the artificial regulation of key environmental factors, including temperature, light, humidity, CO<sub>2</sub> concentration, and root-zone conditions, thereby creating relatively suitable ecological conditions for vegetative growth, reproductive development, and fruit enlargement. Cucumber is a thermophilic crop; low temperatures inhibit root vigor, nutrient uptake, and photosynthetic metabolism, whereas high temperatures can reduce pollen viability, decrease fruit set rate, hinder fruit development, and accelerate plant senescence. Therefore, in protected cultivation systems, maintaining appropriate day and night temperatures through ventilation, shading, covering, heating, cooling, and misting is an important basis for sustaining continuous fruiting and high yield formation. Comparative studies of commercial greenhouses have shown that optimized heating, cooling, and semi-hydroponic systems can significantly increase winter cucumber yield per plant and improve chlorophyll fluorescence parameters, indicating that low-stress and near-optimal environments are beneficial for maintaining high photosynthetic activity and productivity. Microclimate monitoring in polyethylene greenhouses has also demonstrated that appropriate natural ventilation and mist cooling can effectively reduce greenhouse temperature and maintain cucumber growth within a more suitable temperature range.

Light conditions are another key limiting factor affecting yield formation in protected cucumber cultivation. During winter and spring greenhouse production, insufficient natural light often leads to excessive stem elongation, reduced leaf functionality, inadequate female flower differentiation, and slow fruit enlargement. Therefore, optimizing canopy structure, implementing proper pruning practices, supplementary lighting, and spectral regulation are important measures for improving light energy utilization efficiency in protected cucumber populations. Different covering materials and light-conversion films can influence photosynthesis and yield performance by altering greenhouse spectral composition, temperature, and vapor pressure deficit (VPD). Near-infrared reflective plastic films can reduce greenhouse temperature and VPD, enhance photosynthetic rate and transpiration, and improve fruit characteristics and yield. Red-orange light-enhancing films (RPO) can increase photosynthetically active radiation, promote cucumber yield and nutritional quality, and simultaneously affect auxin content and related gene expression in the fruit pedicel (Li et al., 2024). These studies indicate that greenhouse covering materials and light-environment regulation not only affect source strength formation but also further influence fruit set and fruit development processes.

With the development of intelligent protected agriculture, environmental regulation technologies based on sensors, crop models, and machine learning have gradually been applied in cucumber production. Yield prediction and greenhouse microclimate modeling studies have shown that total radiation and air temperature are important environmental factors determining weekly greenhouse cucumber yield (Hong et al., 2024). Integrating light intensity, temperature, and CO<sub>2</sub> concentration with photosynthetic response models can dynamically optimize light saturation points, photosynthetic rates, and environmental control strategies, thereby improving production stability under different climatic and radiation conditions (Bello et al., 2023). Therefore, environmental regulation in protected cucumber cultivation is transitioning from experience-based management to model-driven and intelligent decision-making systems, with the goal not merely of optimizing individual environmental factors but of achieving coordinated optimization of temperature, light, gas exchange, water, and nutrient conditions.

## **6.2 Water and fertilizer management techniques**

Water and nutrient supply constitute the important material basis for high cucumber yield formation. Cucumber has a relatively shallow root system and is sensitive to root-zone moisture and aeration conditions. Appropriate irrigation promotes root growth, nutrient uptake, photosynthesis, and fruit enlargement, whereas prolonged water deficit causes leaf wilting, reduced photosynthetic rate, flower and fruit abortion, and restricted fruit development. Excessive irrigation, however, can lead to root-zone hypoxia, nutrient leaching, and increased disease incidence. Nitrogen, phosphorus, and potassium are the primary nutrients required for cucumber growth and yield formation. Nitrogen affects leaf growth and photosynthetic capacity, phosphorus promotes root development and floral bud differentiation, and potassium participates in assimilate transport, fruit enlargement, and fruit quality formation. Micronutrients such as calcium, magnesium, and boron are also closely associated with cell wall stability, enzyme activity regulation, and fruit quality. Studies on multi-nutrient optimization have demonstrated that irrigation and nitrogen fertilizer significantly promote cucumber yield, whereas magnesium can improve reducing sugar and amino acid contents and alleviate nitrogen deficiency effects through interactions with nitrogen, thereby reducing nitrate accumulation (Li et al., 2023).

Integrated water and fertilizer management technologies can simultaneously deliver water and nutrients to the root zone through drip irrigation systems, enabling precise supply, improving water and fertilizer use efficiency, and reducing resource waste and environmental risks. Under arid climatic conditions, experiments combining different irrigation levels and nitrogen fertilizer rates have shown that moderate deficit drip irrigation combined with appropriate nitrogen application can achieve relatively high yield while improving fruit size and SPAD values, indicating that moderate water-saving practices do not necessarily reduce cucumber yield (Bello et al., 2023). Greenhouse soil cultivation studies have also shown that moderate irrigation combined with suitable nitrogen fertilizer application can achieve near-maximum yield, high water-use efficiency, and desirable fruit quality, whereas excessive nitrogen application may slightly increase yield but reduce nitrogen-use efficiency. Therefore, the key to cucumber water and fertilizer management lies not in simply increasing inputs but in achieving coordinated water and nitrogen supply according to plant demand, cultivation substrate, and environmental conditions.

In substrate cultivation, solar greenhouse systems, and high-input protected cultivation systems, the interaction between water and fertilizer exerts even greater influence on cucumber yield and resource-use efficiency. Studies on substrate bag cultivation have shown that high irrigation levels combined with standard nutrient solution application can achieve the highest yield, whereas appropriately reducing fertilizer application improves nitrogen-use efficiency. Meanwhile, water-use efficiency tends to decrease with increasing irrigation levels. Studies conducted in solar greenhouses on the North China Plain further demonstrated that sufficient water and fertilizer supply can maximize leaf area index, dry matter accumulation, and yield, whereas moderate reductions in water and fertilizer input have relatively small effects on yield but significantly improve water- and nitrogen-use efficiency and may even enhance fruit quality under certain conditions (Wang et al., 2025). In addition, aerated irrigation can alleviate root-zone hypoxia under high water and fertilizer conditions, thereby increasing cucumber yield and nitrogen-use efficiency. These findings indicate that efficient water and fertilizer management in cucumber production should shift from the concept of “high input for high yield” toward the coordinated improvement of yield, quality, and resource-use efficiency.

## **6.3 Optimization of cultivation patterns**

Optimization of cultivation patterns is an important strategy for improving canopy light-use efficiency, source-sink coordination, and yield per unit area in cucumber. It mainly includes planting density, row spacing configuration, pruning method, vine training method, pruning timing, and spatial arrangement. Traditional high-density planting systems can increase plant number per unit area, but excessive density often causes canopy overcrowding, poor ventilation and light penetration, premature senescence of lower leaves, and increased disease incidence, thereby restricting fruit set and fruit enlargement. Modern greenhouse cucumber production places greater emphasis on optimizing population structure by adjusting planting density, stem number, and spatial leaf distribution to achieve balance among yield per unit area, productivity per plant, and fruit quality. Greenhouse

experiments have shown that under natural light conditions, moderately reducing planting density does not reduce weekly or total yield per unit area but instead increases yield per plant, dry matter allocation to fruits, and nutritional quality while reducing nitrite content. However, in recirculating hydroponic systems with sufficient water and nutrient supply, higher planting density combined with double-stem pruning significantly increases total and marketable yield, indicating that the optimal density is not fixed but depends on greenhouse environment, cultivar type, and resource availability (Babatunde et al., 2023).

Pruning and vine training are important techniques for regulating canopy structure and assimilate allocation in protected cucumber cultivation. Appropriate pruning reduces the consumption of nonproductive branches and leaves, improves ventilation and light penetration, and promotes assimilate allocation to fruits. Greenhouse studies comparing different planting densities and pruning methods have shown that higher density and double-stem pruning increase total and marketable yield but reduce fruit diameter, reflecting the trade-off between fruit number and fruit size. Comparative studies of single-stem pruning and natural growth in different cucumber cultivars have also demonstrated that both cultivar type and pruning method significantly influence plant height, leaf area, biomass allocation, and production performance. Single-stem pruning promotes biomass allocation to reproductive organs and alters root-to-shoot ratio and tissue water content (Cheng et al., 2025). Therefore, pruning strategies should be optimized according to cultivar vigor, target yield, fruit size requirements, and cultivation environment rather than simply maximizing branch number or fruit number.

Row spacing configuration and spatial arrangement also affect cucumber canopy light environment, water and fertilizer supply efficiency, and convenience of field management. In solar greenhouses, wide-narrow row configurations combined with drip irrigation lines arranged in narrow rows while wide rows serve as operation channels can improve ventilation and light conditions in the middle and lower canopy while simultaneously accommodating efficient water and fertilizer supply and field management requirements (Wang et al., 2025). In addition, emerging cultivation systems such as substrate cultivation, soilless cultivation, vertical cultivation, and recirculating hydroponics can improve root-zone conditions and enhance land-use and resource-use efficiency, representing important directions for green and efficient cucumber production in protected cultivation systems. Overall, the core objective of cultivation pattern optimization is not to identify a single optimal parameter but to establish integrated cultivation schemes based on cultivar characteristics, greenhouse type, light and temperature conditions, water and fertilizer supply, and market demand in order to achieve coordinated improvement of high yield, superior quality, stable production, and efficient resource utilization.

## **7 Existing Problems and Future Development Trends**

### **7.1 Current problems in research**

Although significant progress has been made in recent years regarding cucumber yield-related traits and their formation mechanisms, the dissection of complex quantitative traits still faces considerable challenges. Cucumber yield is jointly influenced by multiple genes, regulatory networks, environmental factors, and genotype-by-environment interactions, exhibiting characteristics of polygenic control, small genetic effects, and strong environmental sensitivity. Numerous QTLs and functional genes have been reported, including 81 simply inherited genes or major-effect QTLs and 322 QTLs associated with 42 traits; however, these data remain fragmented, and QTL nomenclature systems, cross-population comparisons, and integration standards have not yet been unified. Only a limited number of loci have been successfully cloned, functionally validated, and widely applied in breeding practice. In addition, many identified QTLs explain only a small proportion of phenotypic variation, and insufficient attention has been paid to QTL-by-environment interactions, limiting the application of stable high-yield traits across different ecological regions and cultivation systems.

Insufficient exploitation and utilization of cucumber germplasm resources also represent important constraints on yield improvement. Although large numbers of landraces, wild relatives, and breeding materials are conserved worldwide, practical breeding still faces problems such as inadequate mining of superior alleles, narrow genetic backgrounds of parental lines, and repeated use of core germplasm. Long-term directional selection has promoted improvements in fruit quality, maturity, and adaptation to protected cultivation, but it may also reduce genetic

diversity in certain market types, thereby limiting further genetic gain in yield. For example, the yield improvement of processing cucumbers in the United States has shown little progress over the past 15 years despite long-term selection, suggesting that reliance solely on conventional selection may be insufficient to overcome yield plateaus.

The application of molecular breeding technologies to the improvement of complex yield traits still faces several bottlenecks. Because yield is generally controlled by multiple loci with small effects, the efficiency of marker-assisted selection (MAS) for complex yield traits remains limited, and the transferability of markers across different market types and genetic backgrounds is often inadequate. For example, background-selection SNP markers developed for Korean Baekdadagi-type cucumber exhibit reduced polymorphism in other materials, requiring the establishment of specialized marker systems. Even for major-effect traits, the lack of tightly linked and breeding-applicable markers can limit their practical use. Breeding for cucumber vein yellowing virus resistance was greatly constrained before the fine mapping of the CsCvy-1 gene and the development of KASP markers. Furthermore, candidate genes, QTLs, and regulatory mechanisms identified under laboratory or single-environment conditions still require validation across multiple locations, seasons, and cultivation systems to assess their stability and breeding value. This is particularly important under protected cultivation conditions, where temperature, light, CO<sub>2</sub> concentration, water and fertilizer supply, diseases, pests, and management practices interact strongly. Therefore, the translation of fundamental research findings into practical production systems remains a major challenge.

## **7.2 Future directions driven by emerging technologies**

With the rapid development of modern biotechnology, information technology, and intelligent agriculture, research on cucumber yield formation mechanisms is shifting from single-gene and single-trait analyses toward multi-omics, multi-environment, and data-driven systematic investigations. High-quality reference genomes, saturated SNP genetic maps, and genotyping platforms based on SLAF-seq, GBS, and resequencing technologies have provided important tools for high-density QTL mapping of fruit size, branching, flowering traits, parthenocarp, and other yield-related traits. Some identified loci can explain a substantial proportion of phenotypic variation and provide important foundations for fine mapping and gene cloning (Wang et al., 2024). In the future, greater emphasis should be placed on comparisons among different genetic backgrounds, cultivation environments, and market types to improve the stability and breeding applicability of key loci.

Multi-omics technologies will further enhance the dissection of complex yield traits in cucumber. Next-generation sequencing, transcriptomics, metabolomics, and proteomics can generate large amounts of molecular marker, expression, and metabolic regulatory data, thereby supporting genome-wide association studies (GWAS), haplotype mapping, candidate gene mining, and stress adaptation research. Future studies should strengthen the integration of genomic, transcriptomic, proteomic, metabolomic, phenomic, and environmental data to construct association models linking “genes-regulatory networks-physiological processes-yield phenotypes-cultivation environments.” Such approaches will promote the transition from single candidate-gene analysis to comprehensive dissection of complex regulatory networks.

Gene editing, high-throughput phenotyping, and artificial intelligence technologies will become major driving forces in high-yield cucumber breeding. CRISPR/Cas systems enable precise modification of key functional genes. For example, targeted knockout of the cucumber eIF4E gene has produced non-transgenic materials with broad-spectrum resistance to multiple RNA viruses without causing significant negative effects on yield, demonstrating that precise gene editing can rapidly achieve durable disease resistance and indirectly stabilize yield. In the future, gene editing should be integrated with MAS, genomic selection (GS), and multi-omics-based candidate gene mining to pyramid superior alleles related to high yield, superior quality, disease resistance, stress tolerance, and adaptation to protected cultivation systems (Bhat et al., 2023; Kumar et al., 2024). Meanwhile, intelligent phenotyping systems based on image recognition, automated sensors, three-dimensional reconstruction, unmanned platforms, and machine learning can continuously monitor plant architecture, leaf area, flowering dynamics, fruit development, and stress responses, thereby providing important support for predicting gene-trait

relationships, optimizing hybrid combinations, and designing ideal plant architectures (Bhat et al., 2023; Farooq et al., 2024; Zhu et al., 2024).

### 7.3 Prospects for high-yield cucumber breeding

Future high-yield cucumber breeding will increasingly rely on the deep integration of conventional breeding, molecular breeding, intelligent phenotyping, and digital decision-making technologies. Existing gene and QTL databases, together with gradually standardized nomenclature systems, provide important foundations for constructing genome-wide resource libraries of yield-related loci, plant architecture loci, sex expression loci, and disease-resistance QTLs, thereby facilitating the pyramiding of superior genes through MAS and marker-assisted backcrossing (Dhall et al., 2023). Studies on multi-pistillate flowering QTLs, the parthenocarp locus Parth2.1, sex expression, and fruit diameter indicate that multiple genetic factors affecting fruit number and fruit size can be effectively pyramided through molecular breeding approaches, enabling the development of ideal high-yield plant architectures adapted to different market types and production systems (Wang et al., 2024).

Future breeding objectives will shift from the simple pursuit of high yield toward the coordinated improvement of high yield, superior quality, stable production, stress resistance, and resource-use efficiency. As consumer demands for fruit quality, nutritional value, and food safety continue to increase, new cucumber cultivars should not only possess high yield potential but also exhibit excellent marketability, enhanced nutritional quality, strong disease and stress resistance, and adaptability to water-saving, fertilizer-saving, low-pesticide-input, and intelligent greenhouse management systems. Particularly under the background of climate change, breeding cultivars resistant to low temperature, low light, high temperature, salinity, and multiple diseases will become an important strategy for ensuring stable vegetable supply in protected horticulture systems (Mirzwa-Mróz et al., 2024).

Accelerated breeding technologies will substantially shorten the breeding cycle of new cucumber cultivars and improve selection efficiency for complex yield traits. Doubled haploid (DH) technology, rapid generation advancement, speed breeding in greenhouses, high-throughput phenotyping, and artificial intelligence-assisted selection can be integrated with GS and molecular design breeding to improve prediction accuracy and breeding efficiency for complex traits (Bhat et al., 2023). At the same time, multi-omics studies will further elucidate the regulatory networks underlying yield formation and stress adaptation and guide precise gene editing and molecular marker design (Mirzwa-Mróz et al., 2024). It can be anticipated that with the integrated development of genomics, phenomics, artificial intelligence, gene editing, and intelligent protected agriculture, high-yield cucumber breeding will gradually enter a new stage characterized by precision, intelligence, and sustainability.

## 8 Concluding Remarks

Cucumber is one of the most important vegetable crops worldwide, with an annual global production reaching tens of millions of tons. It has extremely high market demand in major producing countries such as China; therefore, improving and stabilizing cucumber yield is of great economic and food-supply significance. Cucumber yield is determined by multiple interacting traits, including earliness, sex expression, fruit number and size, and resistance to biotic and abiotic stresses. The genetic basis of these traits is being progressively elucidated through QTL mapping, mutant analysis, and molecular characterization. Owing to its relatively small genome, diploid structure, short life cycle, and abundant natural and artificial genetic variation, cucumber has become an ideal model crop for dissecting yield-related traits and rapidly applying research findings to breeding practice.

To date, systematic studies have identified at least 81 simply inherited genes or major-effect QTLs, as well as 322 QTLs associated with 42 quantitative traits, many of which are directly or indirectly related to cucumber yield, fruit size and shape, flowering time, sex expression, and stress resistance. Wild ancestral species and diverse landraces provide rich allelic resources for flowering traits, fruit size, and yield components. Some QTLs exhibit clear signatures of domestication and improvement selection; however, positive transgressive segregation can still occur when favorable wild alleles are introgressed into elite cultivated backgrounds. Fine mapping and cloning studies of major fruit-size QTLs, such as FS5.2/CsCRC, and parthenocarp QTLs, such as Parth2.1, have demonstrated that their regulatory mechanisms mainly involve auxin, gibberellin, and cytokinin signaling

pathways, as well as the regulation of cell division and expansion, thereby revealing how genetic variation is translated into different yield architectures. Meanwhile, studies on sex expression, multi-pistillate flowering, and parthenocarp have identified multiple key loci and marker-trait associations, further clarifying the genetic networks underlying increased female flower production and enhanced fruit set and their contributions to yield formation.

Future breeding for high-yield cucumber cultivars will increasingly rely on the integrated application of genomic resources, multi-omics data, and advanced selection technologies. Comprehensive gene/QTL databases, high-density SNP molecular markers, and background-selection marker systems provide powerful tools for the precise pyramiding of favorable alleles related to yield, quality, and stress resistance. Genomic selection (GS), genome-wide association studies (GWAS), and gene-editing technologies, combined with rapid generation advancement and efficient phenotyping approaches, are expected to accelerate the development of ideal cucumber plant types possessing desirable fruit size, high female flower ratio, multi-pistillate flowering ability, stable parthenocarp, and broad stress resistance. At the same time, continuous exploration of wild resources and underutilized germplasm, as well as further dissection of domestication-related QTLs and heterosis-associated loci, will broaden the allelic resource pool for high-yield and stable-yield cucumber breeding. Overall, the integration of conventional breeding approaches with genomics and molecular biology technologies has made cucumber an important crop in which high and stable yield can be achieved through the rational design and targeted pyramiding of yield-related traits.

### Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

### References

- Aparna A., Skarzyńska A., Płader W., and Pawełkowicz M.E., 2023, Impact of climate change on regulation of genes involved in sex determination and fruit production in cucumber, *Plants*, 12(14): 2651.  
<https://doi.org/10.3390/plants12142651>
- Babatunde A., Deborah R., Gan M., and Simon T., 2023, Effects of plant density and stem pruning on plant biomass yield and economic benefits in a low-cost gravel bed aquaponic system, *Journal of Applied Aquaculture*, 35(4): 837-863.  
<https://doi.org/10.1080/10454438.2022.2033664>
- Baral R., Vainer A., Melzer S., Hause B., and Panda S., 2025, “Bud to fruit”-hormonal interactions governing early fruit development, *Journal of Experimental Botany*, 76(22): 6657-6673.  
<https://doi.org/10.1093/jxb/eraf363>
- Bello A.S., Huda S., Chen Z., Khalid M.A., Alsafran M., and Ahmed T., 2023, Evaluation of nitrogen and water management strategies to optimize yield in open field cucumber (*Cucumis sativus* L.) production, *Horticulturae*, 9(12): 1336.  
<https://doi.org/10.3390/horticulturae9121336>
- Bhat J.A., Feng X., Mir Z.A., Raina A., and Siddique K.H.M., 2023, Recent advances in artificial intelligence, mechanistic models and speed breeding offer exciting opportunities for precise and accelerated genomics-assisted breeding, *Physiologia Plantarum*, 175(5): e13969.  
<https://doi.org/10.1111/ppl.13969>
- Che G., Song W., and Zhang X., 2023, Gene network associates with CsCRC regulating fruit elongation in cucumber, *Vegetable Research*, 3: 7.  
<https://doi.org/10.48130/vr-2023-0007>
- Chen J.C., Liu L., Chen G.X., Wang S.Y., Liu Y.Q., Zhang Z.Y., Li H.F., Wang L.M., Zhou Z.Y., Zhao J.Y., and Zhang X.L., 2024, CsRAXs negatively regulate leaf size and fruiting ability through auxin glycosylation in cucumber, *Journal of Integrative Plant Biology*, 66(5): 1024-1037.  
<https://doi.org/10.1111/jipb.13655>
- Cheng C., Dong C.Y., Wu L., Wu Y., Wang J.L., Gong Z.H., Feng L.P., Li Z.F., Yang F.Y., and Zheng S.H., 2025, Interaction effects of cucumber varieties and pruning methods across different growth stages, *Horticulturae*, 11(5): 464.  
<https://doi.org/10.3390/horticulturae11050464>
- Dey S.S., Sagar V., Kujur S.N., Priyanka N., Munshi A.D., Pandey S., and Behera T.K., 2023, Cucumber: breeding and genomics, *Vegetable Science*, 50(Special): 208-220.  
<https://doi.org/10.61180/vegsci.2023.v50.spl.07>
- Dhall R.K., Kaur H., Manchanda P., and Kaur E., 2023, Genetics and marker-assisted breeding for sex expression in cucumber, *Frontiers in Genetics*, 14: 1180083.  
<https://doi.org/10.3389/fgene.2023.1180083>

- Du X., Song Y., Pan L., and Cui S., 2025, Optimizing cucumber (*Cucumis sativus* L.) fruit metabolomics under elevated CO<sub>2</sub> and high-temperature stress in the greenhouse, *Horticulturae*, 11(1): 10.  
<https://doi.org/10.3390/horticulturae11010010>
- El-Remaly E.A., and Shehata A.I., 2023, Egyptian cucumber germplasm genetic diversity as an approach for developing novel hybrids under heat stress conditions, *Journal of Plant Production*, 14(1): 1-11.  
<https://doi.org/10.21608/jpp.2022.179569.1199>
- Farooq M., Gao S., Hassan M.A., Huang Z., Rasheed A., Hearne S., Prasanna B.M., Li X., and Li H., 2024, Artificial intelligence in plant breeding, *Trends in Genetics*, 40(10): 891-908.  
<https://doi.org/10.1016/j.tig.2024.07.001>
- Hong I., Yu J., Hwang S., and Kwack Y., 2024, Estimation of cucumber fruit yield cultivated under different light conditions in greenhouses, *Horticulturae*, 10(10): 1117.  
<https://doi.org/10.3390/horticulturae10101117>
- Kaur T., Dhali R.K., Manchanda P., and Kumari P., 2024, Utilizing gynoeocious cucumber (*Cucumis sativus* L.) inbreds to investigate inheritance of gynoeecium, fruit yield and its contrasting traits, *Genetika*, 56(2): 255-269.  
<https://doi.org/10.2298/genstr2402255k>
- Koo J., Hwang H., Hwang J., Park E., Yu J., Yun J., Hwang S., Choi H., and Hwang S., 2025, Supplemental lighting and CO<sub>2</sub> enrichment on the growth, fruit quality, and yield of cucumber, *Horticulture, Environment, and Biotechnology*, 66: 77-85.  
<https://doi.org/10.1007/s13580-024-00638-y>
- Kumar R.R., Das S., Choudhury B.U., Kumar A., Prakash N., Verma R., Chakraborti M., Devi A.G., Bhattacharjee B., Das R., Das B., Devi H.L., Das B., Rawat S., and Mishra V.K., 2024, Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing, *Biological Research*, 57: 80.  
<https://doi.org/10.1186/s40659-024-00562-6>
- Li J., Yang X., Zhang M., Li D., Jiang Y., Yao W., and Zhang Z., 2023, Yield, quality, and water and fertilizer partial productivity of cucumber as influenced by the interaction of water, nitrogen, and magnesium, *Agronomy*, 13(3): 772.  
<https://doi.org/10.3390/agronomy13030772>
- Li R., Gao Y., Cai B., Li G., Xue Z., Wang X., and Li Q., 2024, RPO film effectively promotes fruit quality and yield of cucumber through adjusting greenhouse environment and hormone contents, *BMC Plant Biology*, 24(1): 1250.  
<https://doi.org/10.1186/s12870-024-05946-0>
- Lin Y., Weng Y., Fei Z., and Grumet R., 2025, Mining the cucumber core collection: phenotypic and genetic characterization of morphological diversity for fruit quality characteristics, *Horticulture Research*, 12(3): uhae340.  
<https://doi.org/10.1093/hr/uhae340>
- Lnu R., Dhali R.K., Manchanda P., Kumari P., and Singathiya P., 2025, Study on characters associations and path coefficient analysis for yield and quality traits of parthenocarpic cucumber genotypes in poly-net house conditions, *Discover Plants*, 2: 15.  
<https://doi.org/10.1007/s44372-025-00103-9>
- Meng S., Wang K., Han Q., Xie Y., Li X., Song X., Jia J., Wang C., Yan L., Guo P., Zhang J., and Zhang J., 2026, Multi-omics analysis reveals the regulatory role of CsGA20OX1 in GA-mediated parthenocarp of cucumber, *Molecular Breeding*, 46: 23.  
<https://doi.org/10.1007/s11032-026-01635-y>
- Mirzwa-Mróz E., Zieniuk B., Yin Z., and Pawelkiewicz M.E., 2024, Genetic insights and molecular breeding approaches for downy mildew resistance in cucumber (*Cucumis sativus* L.): current progress and future prospects, *International Journal of Molecular Sciences*, 25(23): 12726.  
<https://doi.org/10.3390/ijms252312726>
- Negi C., Shah K.N., and Rana D.K., 2025, Evaluation of variability components and trait association in cucumber germplasm for yield contributing characters in Garhwal region of Uttarakhand, India, *Ecology, Environment and Conservation*, 31(Suppl. 2): S389-S394.  
<https://doi.org/10.53550/eecc.2025.v31i02s.067>
- Patidar P., Kumar P., and Sharma C., 2024, Study on correlation and path coefficient analysis for certain quantitative traits in cucumber (*Cucumis sativus* L.), *International Journal of Advanced Biochemistry Research*, 8(9): 361-364.  
<https://doi.org/10.33545/26174693.2024.v8.i9e.2172>
- Roychowdhury R., Das S.P., Gupta A., Parihar P., Chandrasekhar K., Sarker U., Kumar A., Ramrao D.P., and Sudhakar C., 2023, Multi-omics pipeline and omics-integration approach to decipher plant's abiotic stress tolerance responses, *Genes*, 14(6): 1281.  
<https://doi.org/10.3390/genes14061281>
- Serhiienko O., Solodovnyk L., Harbovska T., Nemchenko S., Radchenko L., and Alsina I., 2025, Factors influencing the variability of the traits "yield," "earliness," and their components in breeding genotypes of cucumber in open field conditions, *Vegetable and Melon Growing*, (76): 6-13.  
<https://doi.org/10.32717/0131-0062-2024-76-6-13>
- Shukla H., Upadhyay D.K., Jha A., Paswan S.K., Singh A.K., and Yadav R.K., 2025, Genetic evaluation for variability, heritability and genetic advance in cucumber (*Cucumis sativus* L.) genotypes, *Journal of Advances in Biology and Biotechnology*, 28(6): 2500.  
<https://doi.org/10.9734/jabb/2025/v28i62500>
- Tadkal R., Rajasree V., Swarnapriya R., Senthil N., and Raveendran M., 2024, Exploring genetic variability, heritability, and genetic advance in growth and yield characteristics of cucumber (*Cucumis sativus*), *International Journal of Advanced Biochemistry Research*, 8(7): 1595.  
<https://doi.org/10.33545/26174693.2024.v8.i7sh.1595>

- Wang C., Qiang X., Wang K., Li H., Zhang X., Liu S., and Gong X., 2025, Optimization of water and nitrogen application rates for synergistic improvement of yield and quality in solar greenhouse cucumber production on the North China Plain, *Plants*, 14(9): 1285.  
<https://doi.org/10.3390/plants14091285>
- Wang W., Xu Z., Qian L., Hang S., Niu Y., Shen C., Wei Y., and Liu B., 2024, Genetic mapping and validation of QTL controlling fruit diameter in cucumber, *BMC Plant Biology*, 24(1): 1271.  
<https://doi.org/10.1186/s12870-024-06000-9>
- Xie Y., Liu X., Sun C., Song X., Li X., Cui H., Guo J., Liu L., Ying A., Zhang Z., Zhu X., Yan L., and Zhang X., 2023, CsTRM5 regulates fruit shape via mediating cell division direction and cell expansion in cucumber, *Horticulture Research*, 10: uhad007.  
<https://doi.org/10.1093/hr/uhad007>
- Xing Y., Cao Y., Yang Y., Wang F., Xin S., and Zhu W., 2023, QTL mapping and transcriptomic analysis of fruit length in cucumber, *Frontiers in Plant Science*, 14: 1208675.  
<https://doi.org/10.3389/fpls.2023.1208675>
- Xu N., Fang X., Xie K., Cheng S., Wang Y., Yang S., Zhu H., Sun S., Weng Y., and Yang L., 2023, Transcriptional and phytohormone regulatory network involved in LITTLELEAF-mediated organ size development in cucumber (*Cucumis sativus*), *Scientia Horticulturae*, 321: 112294.  
<https://doi.org/10.1016/j.scienta.2023.112294>
- Zhu W., Li W., Zhang H., and Li L., 2025, Big data and artificial intelligence-aided crop breeding: progress and prospects, *Journal of Integrative Plant Biology*, 67: 722-739.  
<https://doi.org/10.1111/jipb.13791>



---



#### **Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

## Review and Progress

## Open Access

**Fruit Quality Traits and Cultivation Practices in Chinese Bayberry (*Myrica rubra*): Current Status and Research Progress**Yaya Jin<sup>1,2</sup> <sup>1</sup> Taizhou Manman Agricultural Development Co., Ltd., Linhai, 317000, Zhejiang, China<sup>2</sup> Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China Corresponding email: [1029281612@qq.com](mailto:1029281612@qq.com)Plant Gene and Trait, 2026, Vol.17, No.3 doi: [10.5376/pgt.2026.17.0015](https://doi.org/10.5376/pgt.2026.17.0015)

Received: 16 May, 2026

Accepted: 20 Jun., 2026

Published: 30 Jun., 2026

**Copyright** © 2026 Jin, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.**Preferred citation for this article:**Jin Y.Y., 2026, Fruit quality traits and cultivation practices in Chinese bayberry (*Myrica rubra*): current status and research progress, Plant Gene and Trait, 17(3): 216-234 (doi: [10.5376/pgt.2026.17.0015](https://doi.org/10.5376/pgt.2026.17.0015))

**Abstract** This study explored the current research status and progress of fruit quality traits and cultivation measures in Chinese bayberry (*Myrica rubra* Sieb. et Zucc), systematically analyzing advances in external quality, internal quality, and quality evaluation systems, and reviewing the major factors affecting fruit quality formation and their regulation technologies based on recent domestic and international studies. As an important subtropical fruit tree species in China, Chinese bayberry possesses high nutritional value, economic value, and potential for functional product development. Its fruits are rich in sugars, organic acids, vitamin C, anthocyanins, polyphenols, and volatile aroma compounds. Studies have shown that genetic background, ecological environment, and fruit growth and development are the major factors influencing fruit quality formation, among which light, temperature, water conditions, and maturity stage significantly affect sugar-acid metabolism, anthocyanin accumulation, and flavor formation. Appropriate water and fertilizer management, tree and flower-fruit regulation, protected cultivation, and green ecological cultivation techniques can effectively improve fruit size, coloration, sugar-acid ratio, and functional nutrient contents, thereby enhancing fruit commercial quality and stability. In recent years, the application of electronic nose, electronic tongue, Vis/NIR spectroscopy, hyperspectral imaging, and intelligent detection technologies based on machine learning and deep learning in Chinese bayberry quality evaluation has continuously expanded, providing new approaches for rapid, non-destructive, and accurate quality assessment. Although significant progress has been achieved in Chinese bayberry quality research, there are still limitations in understanding the mechanisms of quality formation, establishing standardized cultivation regulation systems, and developing unified quality evaluation standards. Overall, strengthening research on high-quality cultivar breeding, precision and intelligent cultivation, and green high-quality production systems will contribute to the high-quality and sustainable development of the Chinese bayberry industry.

**Keywords** Chinese bayberry; Fruit quality; Cultivation regulation; Quality evaluation; Intelligent detection**1 Introduction**

Chinese bayberry (*Myrica rubra* Sieb. et Zucc) is one of the most important subtropical evergreen fruit trees native to China, originating from southern China and parts of East Asia, with a cultivation and domestication history of more than 2 000 years (Gao et al., 2024). At present, Chinese bayberry is mainly distributed in southern regions of China, including Zhejiang, Jiangsu, Fujian, Guangdong, Guangxi, and Yunnan provinces, among which Zhejiang and Jiangsu are the major commercial cultivation areas (Ren et al., 2021). As the country with the largest cultivation area and production of Chinese bayberry in the world, China's bayberry industry plays an important role in promoting mountainous economic development, increasing farmers' income, and advancing rural revitalization (Mo et al., 2024). In recent years, with the improvement of living standards and increasing awareness of healthy diets, market demand for high-quality specialty fruits and functional foods has continued to grow, further promoting the rapid development of the Chinese bayberry industry. However, Chinese bayberry fruits ripen rapidly, are highly perishable, and are extremely sensitive to temperature and storage conditions. Their shelf life at room temperature is usually only a few days, which greatly limits long-distance transportation and industrial utilization.

Chinese bayberry fruits are characterized by their bright color, unique flavor, and high nutritional and health-promoting value. Studies have shown that the fruits are rich in sugars, organic acids, vitamin C, phenolic compounds, flavonoids, and anthocyanins, among which cyanidin-3-O-glucoside not only contributes to the

attractive reddish-purple appearance but also exhibits strong antioxidant, anti-inflammatory, and anti-diabetic activities (Li et al., 2023). In addition, by-products such as leaves, kernels, and pomace also contain abundant bioactive compounds that can be used for the development of functional foods and natural antioxidant products (Mo et al., 2024). Fruit quality of Chinese bayberry includes both external quality traits, such as fruit size, color, and firmness, and internal quality traits, such as sugar-acid ratio, aroma volatiles, and functional nutritional components. Among them, sucrose is the major soluble sugar in ripe fruits, while citric acid is the predominant organic acid, and together they determine the characteristic sweet-sour flavor of Chinese bayberry.

At present, research on the formation and regulation mechanisms of Chinese bayberry fruit quality has been continuously advancing worldwide. Existing studies have demonstrated that genetic background, ecological environmental conditions, and cultivation management practices significantly affect fruit quality. Different cultivars exhibit obvious differences in fruit size, coloration, sugar-acid content, and accumulation of bioactive compounds, while sugars, organic acids, phenolics, and volatile substances change rapidly during fruit ripening. Ecological factors such as light, temperature, water availability, and soil conditions influence nutrient accumulation and flavor formation, whereas proper fertilization, water management, flower and fruit thinning, protected cultivation, and LED supplemental lighting can effectively improve both external and internal fruit quality (Tang et al., 2025). In addition, technologies such as preharvest melatonin treatment, ozonated water treatment, and low-temperature storage have shown favorable effects in delaying fruit senescence and extending shelf life (Chen et al., 2024).

This study aims to explore the current research status and progress of fruit quality traits and cultivation regulation measures in Chinese bayberry. With the development of molecular biology and omics technologies, important progress has been achieved in understanding the genetic and molecular mechanisms underlying fruit quality formation. Transcription factor families such as MYB and WRKY play important roles in regulating anthocyanin biosynthesis and flavonoid metabolism. Meanwhile, the construction of high-density genetic maps and multi-omics databases has provided new theoretical foundations and technical support for the analysis of fruit quality-related traits and molecular-assisted breeding. Although related studies have increased rapidly in recent years, comprehensive research integrating industrial development status, fruit quality formation mechanisms, cultivation management practices, and postharvest regulation technologies remains relatively limited. Therefore, this study further discusses external quality, internal quality, and quality evaluation systems of Chinese bayberry, and systematically analyzes the effects of cultivar, environment, and cultivation management on fruit quality formation, with the aim of providing theoretical references for high-quality cultivation, fruit quality improvement, and sustainable development of the Chinese bayberry industry.

## **2 Research on Fruit Quality Traits of Chinese Bayberry**

### **2.1 External quality traits**

The external quality of Chinese bayberry fruit is an important basis for consumers to evaluate commercial value and is also a key target affecting market competitiveness, fruit grading, and breeding selection. It mainly includes fruit size, fruit shape, peel color, fruit surface integrity, and ripening uniformity (Zhang et al., 2022; Xue et al., 2024). Among these traits, single-fruit weight, longitudinal and transverse diameters, and fruit shape index directly influence commercial grade and market price, and large-fruited bayberries are generally more favored by consumers. Significant differences in fruit size and shape exist among different cultivars and germplasm resources. For example, large-fruited cultivars such as ‘Dongkui’ possess high commercial value, whereas some local cultivars mainly produce medium- or small-sized fruits. Large-scale phenotypic analyses have shown that fruit size, fruit shape, and related appearance traits in Chinese bayberry generally exhibit continuous distributions with large coefficients of variation, suggesting that these traits may be quantitatively inherited and jointly influenced by genetic background, tree nutritional status, crop load, and cultivation environment (Zhang et al., 2024).

Peel color is one of the most recognizable external quality traits of Chinese bayberry and is closely associated with fruit maturity and anthocyanin accumulation. Fruit color can gradually transition from white and pink to red, dark red, and nearly black, and this color gradient is mainly determined by differences in anthocyanin content and

composition (Xue et al., 2024). Dark-colored fruits usually contain higher anthocyanin concentrations and exhibit stronger antioxidant capacity, whereas white or yellow types contain low or undetectable levels of cyanidin-3-O-glucoside, resulting in lighter coloration. At the molecular level, MYB transcription factors and their gene clusters are considered important genetic regulators of fruit color variation in Chinese bayberry. MrMYB1 and related MYB/QTL regions are significantly associated with anthocyanin biosynthesis and fruit pigmentation (Cao et al., 2021; Zhang et al., 2024).

In addition to genetic factors, light, temperature, and water-fertilizer management also affect fruit coloration. Appropriate light conditions promote anthocyanin biosynthesis and deepen peel color, whereas prolonged rainy weather, dense canopies, or insufficient ventilation and light penetration may result in uneven coloration, dull appearance, and reduced commercial quality. Fruit surface integrity is also an important component of appearance evaluation. Fruit cracking, mechanical injury, and pest or disease damage not only reduce commercial value but also accelerate postharvest decay and impair storage and transportation performance. Therefore, research on Chinese bayberry external quality is gradually shifting from simple trait description to comprehensive quantitative evaluation systems based on color parameters, image recognition, fruit shape indices, and fruit uniformity (Zhang et al., 2024).

## **2.2 Internal quality traits**

The internal quality of Chinese bayberry fruit mainly includes sugars, organic acids, vitamin C, phenolic compounds, flavonoids, anthocyanins, volatile aroma compounds, flesh texture, and antioxidant capacity, which are the core factors determining fruit flavor, nutritional value, and functional properties (Zhang et al., 2022). Among these, soluble solids content, titratable acidity, and sugar-acid ratio are important indicators for evaluating fresh-eating quality. Most studies have shown that sucrose is the predominant soluble sugar in ripe Chinese bayberry fruits, followed by fructose and glucose, while citric acid is the major organic acid. During fruit development and ripening, total soluble solids, sugars, and anthocyanin contents generally increase gradually, whereas titratable acidity decreases or stabilizes, resulting in the characteristic sweet-sour flavor balance of mature Chinese bayberry fruits.

Chinese bayberry is rich in various functional bioactive compounds, particularly anthocyanins, polyphenols, and flavonoids. Cyanidin-3-O-glucoside is the major anthocyanin component in Chinese bayberry and usually accounts for a high proportion of total anthocyanins, serving as an important material basis for the dark coloration and antioxidant capacity of the fruit. Significant differences in total phenolics, total flavonoids, total anthocyanins, and antioxidant capacity have been observed among cultivars, with black or dark-red fruits generally showing higher levels than pink or white types. Antioxidant evaluation methods such as DPPH, FRAP, ABTS, PSC, and CAA have demonstrated that the antioxidant capacity of Chinese bayberry is usually significantly positively correlated with total phenolic and flavonoid contents (Xia et al., 2021).

Flesh texture and aroma composition are also important aspects of internal quality evaluation. High-quality Chinese bayberry fruits are typically characterized by tender and juicy flesh, low fiber content, balanced sweetness and acidity, and rich aroma. The volatile aroma compounds of Chinese bayberry mainly include terpenoids such as  $\alpha$ -pinene,  $\beta$ -caryophyllene, and D-limonene, as well as aldehydes and esters. Significant differences in volatile composition exist among cultivars and ripening stages, resulting in sensory characteristics such as pine-like, woody, grassy, and overripe aromas. During postharvest storage, temperature and ethylene treatments significantly affect sugar-acid balance, volatile release, firmness retention, and off-flavor formation. Moderate low-temperature storage helps delay quality deterioration, whereas higher temperatures may accelerate flavor degradation and fruit decay (Gao et al., 2024; Saeed et al., 2024).

## **2.3 Quality evaluation index system**

With the continuous development of the Chinese bayberry industry, establishing a scientific, systematic, and quantifiable quality evaluation system has become increasingly important for improving fruit standardization, commercialization, and industrial competitiveness. At present, Chinese bayberry quality evaluation mainly includes sensory evaluation, physicochemical measurements, functional quality analysis, postharvest stability

evaluation, and molecular marker-assisted evaluation (Zhang et al., 2022; Zhang et al., 2024). Sensory evaluation mainly relies on comprehensive assessment of fruit size, color, aroma, sweetness and acidity, taste, and texture, which can directly reflect consumer experience. However, because sensory evaluation is easily influenced by subjective factors, it is usually combined with objective physicochemical indicators for comprehensive analysis. Physicochemical measurements are currently the most widely used methods and mainly include soluble solids, titratable acidity, sugar-acid ratio, vitamin C, total phenolics, total flavonoids, anthocyanins, fruit firmness, and color parameters. Among these, soluble solids and sugar-acid ratio are commonly used to evaluate fresh-eating quality, whereas anthocyanins, polyphenols, and antioxidant capacity are mainly used to evaluate functional quality.

In recent years, technologies such as high-performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS), electronic nose, electronic tongue, near-infrared detection, and image recognition have gradually been applied to Chinese bayberry quality evaluation, improving the accuracy and efficiency of quality detection (Gao et al., 2024). In statistical analysis, multivariate methods such as principal component analysis, cluster analysis, and correlation analysis have been widely used for dimensionality reduction and cultivar classification of quality data. By integrating indices such as sugar-acid ratio, anthocyanin content, total phenolic level, color parameters, and aroma compounds, different Chinese bayberry cultivars can be classified into sweet type, balanced sweet-sour type, dark high-anthocyanin type, and highly aromatic type (Zhang et al., 2022). In addition, in postharvest storage and processing studies, indicators such as decay rate, fruit firmness, titratable acidity, phenolic compounds, antioxidant enzyme activity, and volatile markers have gradually been incorporated into quality evaluation systems to assess shelf-life stability and flavor changes (Saeed et al., 2024).

At the molecular level, studies based on telomere-to-telomere (T2T) reference genomes and genome-wide association analysis have linked multiple standardized quality traits with specific SNP loci and candidate genes, providing an important foundation for molecular marker-assisted selection of superior external and internal quality traits in Chinese bayberry (Zhang et al., 2024). Meanwhile, analyses of the MYB transcription factor family and its regulatory networks have provided candidate gene resources for improving fruit color, flavonoid biosynthesis, and functional quality (Cao et al., 2021; Xue et al., 2024). Although a unified industrial quality evaluation standard has not yet been established, integrating sensory evaluation, physicochemical detection, functional activity, postharvest stability, and molecular markers has become an important trend in constructing comprehensive quality evaluation systems and promoting targeted quality improvement in Chinese bayberry.

### **3 Major Factors Affecting Fruit Quality of Chinese Bayberry**

#### **3.1 Variety factors**

Cultivar is the fundamental factor determining the formation of fruit quality in Chinese bayberry. Significant differences exist among cultivars in fruit size, coloration, sugar and acid contents, aroma characteristics, nutritional composition, and accumulation of functional compounds. A systematic evaluation of 173 Chinese bayberry germplasm accessions demonstrated extensive phenotypic variation in 29 quality traits, including fruit color, size, sugars, organic acids, and amino acids. Most traits exhibited continuous distributions, indicating that Chinese bayberry quality traits possess a complex quantitative genetic basis (Zhang et al., 2024). Common commercial cultivars currently include ‘Dongkui’, ‘Biqizhong’, ‘Dingao’, and ‘Wandao’. Among them, ‘Dongkui’ is favored by the market because of its large fruit size, bright color, and excellent commercial quality, whereas ‘Biqizhong’ is characterized by rich flavor, balanced sugar-acid ratio, and high anthocyanin content in dark-colored fruits (Zhang et al., 2024).

Differences in genetic background among cultivars directly affect fruit color, ripening time, anthocyanin accumulation capacity, and antioxidant activity. Chinese bayberry germplasm resources include white, pink, red, and black-purple types, and these color types differ significantly in anthocyanin content, antioxidant capacity, and processing suitability (Xue et al., 2024). Dark-colored cultivars generally possess higher cyanidin-3-O-glucoside content and stronger antioxidant activity, making them more suitable for fresh consumption, high-anthocyanin products, and functional food development. In addition, different cultivars show distinct postharvest ethylene

release patterns, sugar-acid dynamic changes, and shelf-life performance, indicating that cultivar selection not only affects on-tree fruit quality formation but also determines postharvest quality retention capacity (Saeed et al., 2024).

At the molecular level, differences in Chinese bayberry fruit color are mainly controlled by anthocyanin biosynthesis-related genes and their allelic variations. Studies have shown that the MYB tandem gene cluster, especially MrMYB1.1-MrMYB1.3 and MrMYB2, is closely associated with the fruit color gradient ranging from white ‘Shuijing’ to dark-red ‘Biqizhong’. Functional MrMYB1.1 and MrMYB1.3 alleles can activate anthocyanin biosynthesis-related genes, whereas a single-base deletion in MrMYB1.1 may result in gene inactivation and the formation of a white-fruit phenotype (Xue et al., 2024). Furthermore, telomere-to-telomere (T2T) reference genome and genome-wide association study (GWAS) analyses identified a significant SNP cluster related to fruit color on chromosome 6, containing MYB genes and candidate MLP-like protein genes, thereby providing important molecular marker resources for peel and flesh color improvement (Zhang et al., 2024).

### 3.2 Environmental factors

Ecological environmental conditions play important roles in the formation of Chinese bayberry fruit quality. Factors such as light, temperature, rainfall, humidity, and soil conditions all influence fruit development, coloration, sugar-acid metabolism, and the accumulation of functional compounds. Chinese bayberry is a typical subtropical fruit tree that grows best under warm, humid, well-drained, and slightly acidic environmental conditions. Appropriate temperatures are beneficial for fruit enlargement, sugar accumulation, and anthocyanin formation, whereas high- or low-temperature stress may lead to poor fruit development, uneven coloration, flesh softening, and flavor deterioration. Diurnal temperature differences during the fruit ripening stage usually favor the accumulation of sugars and aroma compounds and are therefore important microclimatic factors affecting fresh-eating quality.

Light is a key environmental factor affecting both external and internal quality of Chinese bayberry fruit. Studies using bagging materials with different light transmittance rates demonstrated that shading during early fruit development significantly reduced fruit coloration and multiple quality indices. Under non-light-transmitting bag treatments, sucrose, glucose, fructose, organic acids, total flavonoids, vitamin C, and total anthocyanin contents all decreased significantly, whereas titratable acidity increased. Combined transcriptomic and metabolomic analyses further showed that weak light conditions suppressed the expression of genes related to flavonoid and anthocyanin biosynthesis and reduced the accumulation of key anthocyanin metabolites. In contrast, bagging materials with better light transmittance promoted fruit coloration and improved overall fruit quality (Yang et al., 2025). In addition, MrMYB1 expression and anthocyanin biosynthesis are highly sensitive to light, and shading through bagging can inhibit fruit coloration and the expression of anthocyanin pathway genes.

Temperature and postharvest storage conditions also significantly affect the stability of Chinese bayberry fruit quality. Studies on different storage temperatures showed that low-temperature conditions of 0 °C-4 °C can slow down fruit firmness decline, sugar-acid degradation, and volatile compound changes, thereby better maintaining fruit flavor, texture, and shelf quality (Figure 1) (Saeed et al., 2024). Under higher temperatures, ethylene release and respiratory metabolism increase, accelerating fruit softening, color changes, and flavor deterioration, while off-flavor volatile compounds such as ethanol, benzaldehyde, and octanoic acid gradually accumulate (Gao et al., 2024). Room-temperature storage also causes rapid decreases in acidity and fluctuations in sugar-acid balance, significantly shortening the marketable shelf life of fresh fruits.

### 3.3 Growth and development factors

The growth and developmental status of Chinese bayberry trees directly affects fruit quality formation, among which tree age, tree vigor, crop load, and nutrient distribution are important influencing factors. Generally, Chinese bayberry trees in the full-bearing stage exhibit relatively stable fruiting ability, with better fruit size, sugar accumulation, and ripening uniformity, whereas young or aging trees are more likely to produce uneven fruit size and unstable quality. Excessive tree vigor may result in an imbalance between vegetative growth and reproductive growth, thereby affecting fruit coloration and sugar accumulation, while weak tree vigor reduces leaf

photosynthetic capacity and is unfavorable for fruit enlargement and metabolite accumulation. In addition, appropriate crop load helps maintain tree nutritional balance and improves fruit size, sugar-acid balance, and ripening consistency, whereas excessive fruit load intensifies nutrient competition among fruits, leading to smaller fruits, lower soluble solids, and uneven ripening. Therefore, practices such as flower and fruit thinning, proper pruning, and canopy regulation are commonly applied in production to adjust tree load and improve fruit commercial quality and quality stability.



Figure 1 Color development in five cultivars and cross-sections of Chinese bayberry fruits stored at different temperatures after harvest (Adopted from Saeed et al., 2024)

From the perspective of fruit development, the ripening stage is the key factor determining dynamic changes in Chinese bayberry fruit quality. During fruit ripening, chlorophyll and titratable acidity gradually decrease, whereas sugars, anthocyanins, and volatile compounds continuously accumulate, ultimately resulting in mature fruits characterized by softness, juiciness, balanced sweet-sour taste, and rich aroma. Comparisons among mature-green, pink, red-ripe, and fully ripe fruits have shown that total soluble solids, sugar composition, and total anthocyanin content usually reach relatively high levels at the fully ripe stage. In contrast, some antioxidant indices such as total phenolics and DPPH, FRAP, and ABTS activities may decline during ripening progression, and immature fruits sometimes exhibit stronger antioxidant capacity (Wu et al., 2018). Volatile compound composition also changes markedly during ripening. Immature fruits are usually dominated by citrus-like terpene aromas, whereas fully ripe fruits exhibit grassy, herbal, and cucumber-like aroma characteristics.

At the molecular and metabolic levels, Chinese bayberry fruit ripening is accompanied by extensive transcriptional reprogramming. RNA-Seq and EST studies have shown that genes related to anthocyanin biosynthesis are globally upregulated during ripening, while pathways associated with sugar-acid metabolism, energy metabolism, and cell wall modification undergo significant changes. Factors such as sucrose phosphate synthase and vacuolar ATPase subunits may participate in sugar accumulation and fruit quality formation. In addition, fruit maturity also affects postharvest quality trajectories. Immature and mature fruits usually exhibit obvious climacteric respiration and ethylene peaks, whereas fully ripe fruits may not display typical climacteric behavior. Under room-temperature conditions, soluble solids, titratable acidity, and organic acids decrease rapidly in fruits at all maturity stages (Wu et al., 2018). Therefore, fruit growth and development not only determine on-tree quality formation but also influence postharvest softening, aroma changes, and shelf life. Proper determination of harvest maturity is thus a crucial technical factor balancing fresh-eating quality, transportation performance, and processing suitability.

## 4 Regulatory Effects of Cultivation Measures on Fruit Quality

### 4.1 Water and fertilizer management measures

Water and fertilizer management is an important cultivation practice affecting the formation of Chinese bayberry fruit quality and is directly related to fruit enlargement, sugar accumulation, sugar-acid balance, and yield stability. Rational fertilization provides the nutritional basis for tree growth and fruit development, among which nitrogen, phosphorus, and potassium have particularly significant effects on fruit quality. Appropriate nitrogen application promotes branch and leaf growth as well as photosynthesis; however, long-term excessive chemical fertilizer input, especially excessive nitrogen fertilization, can easily result in excessive tree vigor, decreased fruit sugar content, delayed ripening, intensified soil acidification, and organic carbon loss, which are unfavorable for the sustainable production of Chinese bayberry orchards (Hong et al., 2023). In contrast, moderately reducing nitrogen and phosphorus application can improve soil quality without significantly decreasing yield and fruit quality, indicating that balanced fertilization is more beneficial for stable production of high-quality Chinese bayberry than simply increasing fertilizer input (Chen et al., 2025).

Potassium fertilizer generally promotes sugar accumulation, peel coloration, and flavor formation, whereas phosphorus fertilizer is closely associated with root development and flower bud differentiation. Specialized compound fertilizers developed for Chinese bayberry can improve soil organic matter and nutrient content, increase leaf biomass and chlorophyll content, and enhance fruit total sugar, reducing sugar, and soluble solids contents. Among different application rates, approximately 8 kg per tree applied in two split applications showed favorable effects (Wu et al., 2021). Foliar nutrient regulation also has considerable application potential. Amino acid foliar fertilizer treatment on ‘Dingao’ Chinese bayberry significantly increased soluble solids, total sugars, sugar-acid ratio, and soluble solids/titratable acidity ratio, while reducing total acidity and improving postharvest water loss and decay rates, indicating its ability to simultaneously improve eating quality and storage tolerance.

Micronutrients and water management also influence Chinese bayberry fruit quality. Boron deficiency can cause small leaves, bud dieback, and reduced flowering and fruiting, whereas soil boron application or foliar boron spraying can improve fruit set, yield, single-fruit quality, and sugar-acid ratio. Combined application of boron fertilizer and paclobutrazol can also alleviate alternate bearing. In addition, intercropping ryegrass in Chinese bayberry orchards can improve rhizosphere soil properties and microbial environments, increase fruit sugar, vitamin C, and flavonoid contents, and reduce acidity, thereby providing a new technical approach for ecological water and fertilizer management (Li et al., 2023). Regarding water management, adequate water supply should be ensured during the fruit enlargement stage, whereas excessive rainfall or drastic soil moisture fluctuations during the ripening stage can easily lead to fruit cracking, sugar dilution, and increased decay. Therefore, precise irrigation and drainage management should be implemented according to different fruit developmental stages.

### 4.2 Tree, flower, and fruit management

Tree structure and flower-fruit management can influence Chinese bayberry fruit quality by regulating canopy microenvironment, source-sink relationships, and reproductive load. Proper training and pruning can improve canopy ventilation and light penetration, enhance leaf photosynthetic efficiency, and promote nutrient accumulation and uniform fruit coloration. Chinese bayberry trees generally have large canopies, and if not pruned for long periods, dense canopy closure can occur, resulting in insufficient inner-canopy light, increased pest and disease incidence, and uneven fruit coloration. Therefore, orchard management commonly includes thinning overly dense branches, vigorous shoots, and weak or diseased branches to optimize canopy structure and improve fruit commercial quality and ripening consistency.

Flower and fruit management is an important approach for regulating nutrient distribution and improving fruit quality. Excessive fruit load intensifies nutrient competition among fruits, leading to reduced single-fruit weight and insufficient sugar accumulation. Moderate flower and fruit thinning can reduce ineffective nutrient consumption and direct more assimilates toward retained fruits, thereby improving fruit size, coloration, soluble solids content, and sugar-acid ratio. The combined regulation of paclobutrazol and boron fertilizer has also been used to balance vegetative and reproductive growth. Spraying 100-200 mg·L<sup>-1</sup> paclobutrazol during the

full-bloom stage can increase fruit weight, soluble solids content, and sugar-acid ratio while reducing total acidity; however, excessively high concentrations may significantly reduce fruit set, and therefore the application rate must be carefully controlled.

Management practices directly affecting the local fruit environment can also improve fruit quality. Insect- and rain-proof nets not only reduce pest damage and fruit cracking risks but also increase fruit diameter, single-fruit weight, edible rate, and sugar-acid ratio, showing better overall effects than open-field cultivation or insecticide application alone. These nets can also alter the bacterial community structure on fruit surfaces, reducing microbial groups associated with sugar consumption and disease, thereby favoring fruit growth and sugar accumulation (Yu et al., 2021). In addition, fruit bagging treatments can affect fruit quality by regulating light conditions, temperature, humidity, and pest pressure around the fruit. Using opaque bags during the young fruit stage reduces sugars, total flavonoids, vitamin C, and total anthocyanin contents while increasing acidity, resulting in poorer coloration and reduced fruit quality (Figure 2) (Yang et al., 2025). Therefore, Chinese bayberry bagging should balance light transmittance, ventilation, and protective effects.

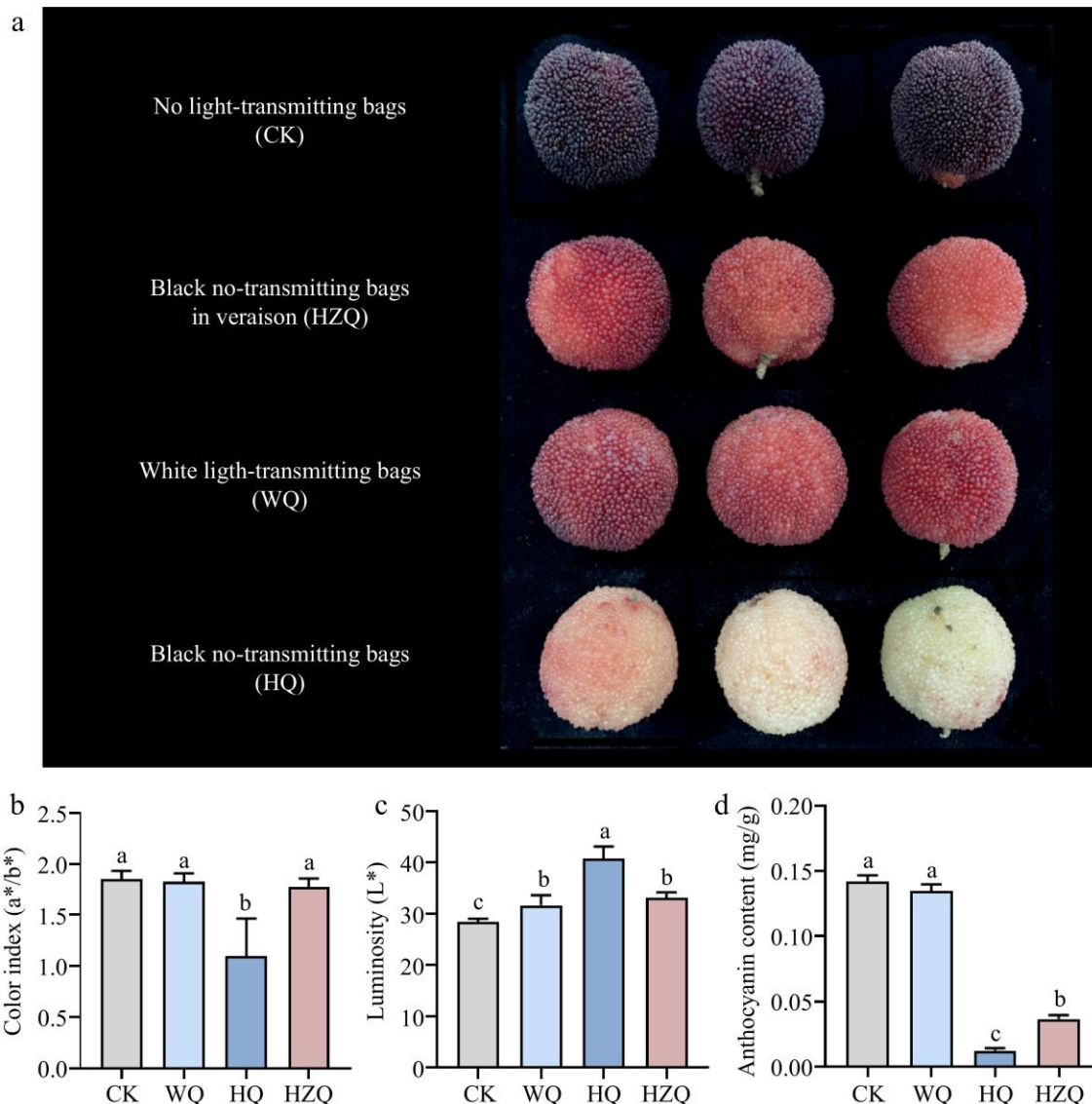


Figure 2 (a), (b) Effect of fruit color, (c) luminosity, and (d) anthocyanin in mature Chinese bayberry after treatment with different light-transmitting bags (Adopted from Yang et al., 2025)

Image caption: Bars represent the mean and standard error of three independently transformed biological replicates (n = 3). Different letters indicate statistically significant difference in one-way ANOVA analysis ( $p < 0.05$ ); HQ: Blank-no-transmitting bag treatment; WQ: white-light treatment (Adopted from Yang et al., 2025)

### 4.3 Facility cultivation and green cultivation techniques

Facility cultivation and protected cultivation provide effective approaches for regulating temperature, light, and rainfall during critical periods of Chinese bayberry fruit development. Rain-shelter cultivation, greenhouse cultivation, insect- and rain-proof netting, and mulching cultivation can improve orchard microclimate and reduce the negative effects of continuous rainfall during the ripening period. Particularly during the fruit ripening stage, rainy weather often causes fruit cracking, decay, and disease occurrence, whereas rain-shelter facilities can reduce water accumulation on fruit surfaces and disease pressure, thereby improving fruit integrity, marketability, and postharvest storability.

Greenhouse cultivation can avoid the adverse effects of rainfall during harvest and improve single-fruit weight, fruit size, soluble solids content, and sugar-acid ratio, showing superior fruit quality compared with open-field cultivation. The physiological basis for these improvements may be associated with enhanced sucrose phosphate synthase activity and reduced acid invertase activity under greenhouse conditions, thereby promoting sucrose accumulation and improving sugar-acid quality (Wu et al., 2021). In controlled facility environments, LED supplemental lighting technology can further precisely regulate light intensity and spectral composition. For example, optimized LED supplemental lighting significantly increased fruit weight, soluble solids, and vitamin C content while reducing organic acid content in ‘Heitan’ Chinese bayberry. ‘Dongkui’ showed relatively weaker responses but still exhibited certain quality improvements, indicating cultivar-specific differences in light environment regulation requirements (Tang et al., 2025).

Green ecological cultivation technologies are important directions for the high-quality development of the Chinese bayberry industry. Their core objective is to maintain and improve fruit quality while reducing chemical inputs and improving orchard ecological environments. Intercropping ryegrass in Chinese bayberry orchards can improve soil physicochemical properties, rhizosphere microbial communities, and metabolic environments, thereby increasing fruit sugar, vitamin C, and flavonoid contents while reducing acidity (Li et al., 2023). Insect- and rain-proof nets, as green facility technologies, can simultaneously achieve pest control, rainfall exclusion, and microclimate regulation, thereby improving fruit size, edible rate, and economic benefits while reducing pesticide dependence (Yu et al., 2021). In addition, grafting ‘Biqizhong’ onto North American bayberry (*Morella cerifera*) rootstocks enables plants to maintain healthy growth under saline-alkaline soil conditions and increases fruit sucrose and citric acid contents, providing new possibilities for high-quality Chinese bayberry production in saline-alkaline areas (Saeed et al., 2023).

## 5 Research Progress in Detection and Evaluation Technologies for Chinese Bayberry Quality

### 5.1 Traditional detection methods

Traditional quality detection methods are an important foundation for evaluating Chinese bayberry fruit quality and are still widely used in scientific research, cultivar comparison, postharvest quality analysis, and commercial grading. Sensory evaluation is one of the most common traditional methods and mainly assesses overall fruit quality based on fruit color, size, shape, aroma, sweetness and acidity, flesh texture, juiciness, and overall acceptability. This method is highly intuitive and can directly reflect consumer perception of fresh fruit quality, thus playing an important role in fruit grading, market evaluation, and flavor description. Descriptive sensory analysis combined with physicochemical indices has been used to classify sweetness, acidity, and aroma characteristics among different Chinese bayberry cultivars, thereby providing a basis for selecting superior fresh-eating cultivars.

Physicochemical measurement is the core component of the traditional quality evaluation system and mainly includes soluble solids, titratable acidity, sugar-acid ratio, vitamin C, total phenolics, anthocyanins, fruit firmness, sugars, organic acids, and color difference parameters (Xuan et al., 2022). Among these, soluble solids are commonly used to reflect sugar content, whereas titratable acidity is used to evaluate acidity; together, they determine flavor balance in Chinese bayberry fruit. Vitamin C, polyphenols, and anthocyanins are often analyzed to evaluate nutritional and functional quality. Conventional analytical methods, such as refractometry for Brix determination, acid-base titration for titratable acidity, spectrophotometry for total phenolics and anthocyanins,

and high-performance liquid chromatography (HPLC) for sugars and organic acids, have provided important data for studies on fruit quality formation.

Traditional sensory evaluation systems and standardized analytical methods remain important references in Chinese bayberry quality research. In studies on Chinese bayberry juice, evaluators trained according to Chinese GB/T and ISO standards used quantitative descriptive analysis to establish flavor descriptors such as sourness, sweetness, bitterness, and astringency. Combined with analyses of nine organic acids, three sugars, and total polyphenol content, partial least squares analysis was used to establish relationships between sensory flavor characteristics and key chemical compounds. In addition, manual sensory evaluation can also be used to determine thresholds for “off-flavor” formation during storage, such as identifying the transition point from normal flavor to deteriorated flavor under different storage temperatures (Gao et al., 2024). However, traditional methods generally suffer from limitations including sample destructiveness, time-consuming procedures, high labor intensity, and strong subjective variability, making them insufficient for the modern Chinese bayberry industry’s demand for high-throughput, online, and rapid grading technologies.

## 5.2 Modern detection technologies

With the development of sensing technologies, spectral analysis, and chemometrics, modern detection technologies have been increasingly applied in Chinese bayberry quality research. In particular, non-destructive optical detection, electronic sensing, and volatile fingerprint analysis have provided faster, more objective, and data-driven approaches for quality evaluation. Color difference analysis objectively evaluates fruit color changes through comprehensive color parameters such as L, a, and b values. Digital image analysis, combined with image acquisition and computer processing, can rapidly analyze fruit size, shape, coloration, and surface defects, thereby improving the standardization of external quality evaluation.

Visible/near-infrared (Vis/NIR) spectroscopy is one of the earliest and most mature non-destructive technologies used in Chinese bayberry detection. Reflectance spectra within the 325-1075 nm range have been used to establish partial least squares models for rapid prediction of acidity and pH in intact Chinese bayberry fruit, showing high model correlations. Near-infrared transmittance spectroscopy has also been applied to predict titratable acidity, malic acid, and citric acid contents in different cultivars, demonstrating good application potential under temperature-controlled conditions. In recent years, hyperspectral imaging has further integrated spectral and spatial image information, allowing simultaneous prediction of single-fruit weight and soluble solids content, even for packaged fruit. After feature selection, PLS modeling, and model transfer correction, prediction stability across batches can be significantly improved (Yuan et al., 2025).

Electronic sensing technologies rapidly identify flavor quality by simulating human olfactory and gustatory systems. Electronic noses use multi-gas sensor arrays to monitor volatile flavor changes during storage. Combined with sensory evaluation and GC-MS analysis, they can identify stages of off-flavor formation and key volatile markers. Studies have shown that electronic nose response patterns change regularly with prolonged storage time and increasing temperature, allowing identification of off-flavor fruit stored for more than 2 days at 20 °C or more than 7 days at 10 °C, which is associated with the accumulation of ethanol, benzaldehyde, octanoic acid, and other volatiles (Gao et al., 2024). By combining electronic noses with stochastic resonance signal processing and regression models, indicators such as firmness, color, pH, total soluble solids, and reducing sugars can also be predicted, demonstrating potential for low-cost and rapid overall quality evaluation.

Electronic tongues are mainly used to analyze taste characteristics such as sweetness, sourness, bitterness, and astringency. In Chinese bayberry juice analysis, electronic tongues combined with sensor arrays and discriminant analysis effectively differentiated juices from different origins and cultivars, while correlating electronic tongue signals with organic acids, sugars, total polyphenols, and sensory scores. In addition, HS-GC-IMS technology can establish volatile fingerprint profiles during storage of Chinese bayberry juice and NFC products, identifying off-flavor markers such as ethanol and ethyl acetate, thereby providing a new approach for monitoring processing quality and early warning of flavor deterioration (Xuan et al., 2022). Low-cost portable Vis/NIR devices have also

been applied for non-destructive detection of sugar and acidity in Chinese bayberry fruit, and their portability makes them more suitable for small-scale growers and field grading applications (Wang et al., 2023).

### **5.3 Intelligent evaluation technologies**

In recent years, the rapid development of artificial intelligence, big data, machine vision, and the Internet of Things (IoT) has promoted the transformation of Chinese bayberry quality evaluation from traditional single-point detection to intelligent, real-time, and systematic approaches. Intelligent recognition technologies based on machine vision can automatically collect fruit images and use machine learning or deep learning algorithms to identify fruit size, color, defects, and maturity, thereby enabling rapid grading. Compared with traditional manual sorting, intelligent recognition technologies have advantages including fast detection speed, unified standards, traceability, and suitability for online deployment. In orchard production, machine vision can also be used to monitor fruit ripening processes and canopy growth status, thereby supporting timely harvesting and precision management (Knott et al., 2022).

In sensor data modeling, machine learning has become an important tool for constructing Chinese bayberry quality detection models. Electronic nose data combined with stochastic resonance signal processing and multivariate regression models can be used to predict firmness, pH, color, soluble solids, and reducing sugars with high prediction accuracy. In postharvest flavor evaluation, artificial neural network models can verify the classification performance of electronic noses in distinguishing normal-flavor and off-flavor fruits under different storage temperatures (Gao et al., 2024). In spectral detection, algorithms such as partial least squares analysis, principal component analysis, and artificial neural networks are often combined with Vis/NIR or hyperspectral data for acidity prediction, cultivar identification, and cross-batch model transfer (Yuan et al., 2025). These technologies improve detection efficiency and data processing capability, providing technical support for rapid and non-destructive quality evaluation of Chinese bayberry fruit.

Image-based deep learning technologies have also begun to be applied to maturity and quality detection in Chinese bayberry. Zheng et al. (2025) proposed a cascaded framework combining the lightweight instance segmentation model SOLOv2-Light and multi-feature regression for field maturity detection of Chinese bayberry fruit (Figure 3). After segmenting individual fruits, the system integrated deep semantic features, LAB color information, and multi-scale texture features, while using a/b values obtained by colorimeters as maturity labels, achieving high segmentation performance and low maturity prediction error. When maturity was divided into three levels, classification accuracy reached 95.5%. From the perspective of industrial applications, future intelligent quality evaluation of Chinese bayberry is expected to develop toward multi-source data fusion and online detection. In orchards, machine vision and IoT technologies can be combined to monitor fruit maturity, temperature, humidity, and soil moisture; in sorting facilities, hyperspectral imaging and Vis/NIR technologies can be used for automatic grading; and in processing plants, electronic noses, electronic tongues, and HS-GC-IMS can be used to monitor flavor stability (Dhiman et al., 2022; Apostolopoulos et al., 2023). With the development of lightweight models, portable sensors, and cloud platforms, Chinese bayberry quality detection is expected to achieve real-time application from laboratories to orchards, storage systems, and processing lines (Hassan et al., 2025; Júnior et al., 2025).

## **6 Current Problems in Existing Research**

### **6.1 Insufficient research on the mechanisms of quality formation**

In recent years, increasing attention has been paid to the mechanisms underlying Chinese bayberry fruit quality formation, and technologies such as genomics, transcriptomics, and metabolomics have gradually been applied to the analysis of quality traits. However, overall, the mechanisms governing fruit quality formation still lack systematic, in-depth, and experimentally validated theoretical explanations. At present, a high-quality telomere-to-telomere (T2T) reference genome and genome-wide association studies (GWAS) have linked 29 quality traits with 1 937 SNP loci and 1 039 candidate genes, and identified an important MYB/MLP regulatory locus associated with fruit color on chromosome 6 (Zhang et al., 2024). Nevertheless, for complex quality traits

such as fruit texture, flavor balance, aroma compound accumulation, storability, and postharvest softening, studies on key allelic variations, regulatory networks, and functional validation remain insufficient.

### Simple Scene



### Multi-objective Stacking Scene



Figure 3 Prediction performance of the regression model (Adopted from Zheng et al., 2025)

Image caption: A simple scenario and a complex scenario with multiple stacked fruits. The predicted maturity percentage is shown above each fruit. Each color represents a distinct instance mask (Adopted from Zheng et al., 2025)

Existing studies mainly focus on a limited number of quality indicators, including sugars, acidity, anthocyanins, and flavonoids, whereas research on terpene metabolism, cell wall metabolism, antioxidant systems, hormone signaling, and volatile aroma formation is still inadequate (Yang et al., 2025). Transcriptomic studies during fruit ripening have shown that anthocyanin biosynthesis-related genes are generally upregulated during fruit maturation, and significant changes also occur in sugar-acid metabolism and energy metabolism pathways. However, many candidate genes remain at the stage of correlation analysis, lacking further validation through gene editing, transient expression, genetic transformation, or population genetics approaches (Sun et al., 2024).

In addition, Chinese bayberry fruit quality formation is the result of interactions among genetic factors, environmental conditions, developmental stages, and cultivation practices, yet most current studies still focus on single factors or a few cultivars. For example, studies on light exposure and fruit bagging have demonstrated that weak light conditions reduce the accumulation of sugars, organic acids, vitamin C, flavonoids, and anthocyanins, and identified HY5 and LDOX as possible regulators involved in light-induced coloration (Yang et al., 2025). However, the interactions among light signaling, hormone signaling, transcription factor networks, and metabolic pathways remain unclear. Furthermore, most studies are concentrated on a few major cultivars such as ‘Biqi’ and ‘Dongkui’, while investigations involving large germplasm populations, multiple ecological regions, and different cultivation systems are still relatively limited (Zhang et al., 2024). Therefore, current quality regulation in Chinese bayberry production still relies heavily on empirical cultivation practices rather than precise regulation based on mechanistic understanding.

## 6.2 Incomplete cultivation regulation technology system

Although cultivation and management technologies for Chinese bayberry have continuously developed, there is still a lack of standardized, systematic, and region-specific regulation systems aimed at fruit quality improvement. Most existing studies focus on individual production stages or single technologies, such as water and fertilizer management, canopy regulation, bagging, insect- and rain-proof nets, LED supplemental lighting, and postharvest preservation, whereas studies on the synergistic effects, application boundaries, and long-term ecological impacts of these measures remain limited. For example, insect- and rain-proof nets can effectively control fruit flies, improve fruit size, edible rate, and sugar-acid ratio, and optimize fruit-surface microbial communities, but related studies are still mainly restricted to single cultivars and regions, and continuous evaluation of their long-term effects on tree vigor, soil microorganisms, and orchard ecosystems is lacking (Yu et al., 2021).

Facility cultivation and precision light environment regulation also exhibit strong cultivar dependence and insufficient standardization. LED supplemental lighting can significantly improve fruit size, soluble solids, and vitamin C content in the cultivar ‘Black Charcoal’, but its effects are relatively limited in cultivars such as ‘Dongkui’, indicating that different cultivars respond differently to light intensity, spectral composition, and lighting periods. Therefore, cultivar-, region-, and developmental stage-specific light regulation systems still need to be established (Tang et al., 2025). Similarly, technologies such as water and fertilizer regulation, flower and fruit thinning, canopy pruning, and harvest maturity management may produce different effects under different ecological regions and cultivars, but systematic comparisons and standardized technical protocols are currently lacking.

The postharvest preservation technology system also requires further improvement. Technologies such as ultrasonic treatment combined with slightly acidic electrolyzed water, ozone water treatment, low-temperature storage, and optimized temperature management can reduce microbial populations, delay fruit softening, and maintain sugar-acid balance and phenolic contents. However, most studies are still limited to experimental conditions or small-scale validation and have not yet formed integrated systems linked with harvest maturity, transportation damage control, packaging materials, cold-chain logistics, and market circulation periods (Suo et al., 2023; Gao et al., 2024). Moreover, current research mainly focuses on short-term quality or shelf-life indicators, while comprehensive evaluations regarding economic cost, carbon footprint, low-residue production, farmer operability, and industrial promotion models remain insufficient. Therefore, future cultivation regulation of Chinese bayberry should shift from optimization of individual technologies toward integrated “cultivar-environment-tree-facility-postharvest” management systems.

## 6.3 Lack of unified quality evaluation standards

At present, Chinese bayberry quality evaluation has not yet formed unified and widely applicable industrial standards, which to some extent restricts the standardization, branding, and commercialization of the industry. Current quality evaluation systems have gradually expanded from traditional sugar-acid indices to multidimensional indicators including external appearance, phenolic compounds, flavonoids, anthocyanins, antioxidant activity, texture, volatile compounds, sensory evaluation, and postharvest storability (Gao et al., 2024; Saeed et al., 2025; Yang et al., 2025). However, different studies, production regions, and enterprises often adopt different detection methods, indicator combinations, and evaluation thresholds, resulting in poor comparability among research results.

For example, studies on Chinese bayberry juice processing indicate that sugar-acid ratio, total sugar, and titratable acidity are key factors influencing sensory preference and can be used to screen suitable processing cultivars. However, these evaluation criteria have not yet been effectively linked with fresh fruit commercial grades, processing grades, or national and industry standards. Flavor studies can classify cultivars according to aroma-active compounds and sensory characteristics, but corresponding grading systems applicable to market circulation, brand construction, and trade evaluation have not yet been established. In postharvest quality research, methods such as electronic nose off-flavor classification, artificial neural network models, volatile markers, and

texture analysis possess important scientific value, but they have not yet been transformed into practical industrial evaluation procedures and standard thresholds (Suo et al., 2023; Gao et al., 2024).

Furthermore, unified relationships among genetic markers, laboratory detection indices, and market grades for fresh and processed products are still lacking. Although GWAS studies have identified 29 quantifiable quality traits and related loci, these molecular indicators have not yet been systematically translated into breeding selection standards, cultivar certification indices, or commercial grading criteria (Zhang et al., 2024). Meanwhile, different consumer groups exhibit varying preferences for Chinese bayberry quality; some focus more on sweetness and taste, whereas others emphasize nutritional functionality, safety, or processing suitability. Therefore, future studies should establish multidimensional comprehensive evaluation systems covering appearance quality, flavor quality, nutritional and functional quality, safety quality, postharvest quality, and processing quality, while promoting the integration of non-destructive detection, intelligent evaluation, and molecular marker data with industrial standards.

## **7 Future Development Directions**

### **7.1 Breeding of high-quality new varieties**

With increasing consumer demands for Chinese bayberry fruit quality, the breeding of high-quality new cultivars will become a core direction for the future development of the industry. Current market demands have gradually shifted from focusing solely on fruit size and yield to comprehensive traits such as flavor quality, nutritional value, functional components, storability, transportation tolerance, and green safety. Therefore, breeding objectives for Chinese bayberry should shift from traditional high-yield orientation toward coordinated improvement of high quality, efficiency, multifunctionality, and adaptability. By developing new cultivars with balanced sugar-acid ratios, attractive coloration, tender flesh, and high contents of anthocyanins and flavonoids, the market competitiveness and industrial value of Chinese bayberry can be further enhanced (Zhang et al., 2024; Saeed et al., 2025).

In recent years, genomics and multi-omics technologies have provided an important foundation for molecular design breeding in Chinese bayberry. Telomere-to-telomere reference genomes and GWAS analyses of 173 germplasm accessions have identified 1,937 SNP loci and 1,039 candidate genes associated with 28 fruit quality traits, among which MYB and MLP-like gene regions on chromosome 6 are closely related to fruit coloration and anthocyanin accumulation (Zhang et al., 2024). The Chinese bayberry database integrates multi-omics information including genomes, transcriptomes, molecular markers, phenotypes, and fruit images, thereby providing a platform for quality trait marker development, candidate gene screening, and computer-assisted breeding (Ren et al., 2021). In addition, high-density SNP genetic maps and QTL mapping studies have laid the foundation for marker-assisted selection of traits related to tree growth, leaf characteristics, and yield performance (Zhang et al., 2021).

These molecular research achievements have gradually begun to integrate with breeding practice. For example, the new hybrid line ‘BD-107’, developed from a cross between ‘Biqi’ and ‘Dongkui’, showed superior fruit firmness, sugar content, and vitamin C content compared with both parents, and also contained richer terpene and flavonoid compounds, demonstrating the potential of hybrid breeding for improving flavor, texture, and functional quality (Saeed et al., 2025). In addition, studies on developmental regulatory genes such as MrSPL4 suggest that these genes may influence vegetative growth and flowering time, thereby providing new strategies for breeding early-maturing, highly adaptable, and facility-suitable cultivars (Zhang et al., 2022). In the future, breeding of Chinese bayberry cultivars should further focus on the coordinated improvement of flavor quality, nutritional functionality, storability, pest and disease resistance, stress tolerance, maturity regulation, and processing suitability.

### **7.2 Precision and intelligent cultivation**

Precision and intelligent cultivation represent important trends in the development of modern fruit industries and are also key approaches for improving fruit quality stability and production efficiency in Chinese bayberry. Traditional Chinese bayberry cultivation mainly relies on empirical management, which often results in

uncertainties in water and fertilizer supply, pest control, canopy regulation, and harvest judgment, thereby causing resource waste and quality fluctuations. In the future, technologies such as sensor monitoring, automated control, machine vision, and artificial intelligence models should be used to monitor soil moisture, nutrient status, canopy light conditions, tree vigor, and fruit maturity in real time, enabling the transition from experience-based management to data-driven management (Sharma and Shivandu, 2024; Soussi et al., 2024).

Facility cultivation and controlled-environment technologies will become important directions for precision cultivation of Chinese bayberry. In protected cultivation systems, LED supplemental lighting can improve fruit quality through precise regulation of light intensity and spectral composition. Studies have shown that optimized LED lighting can increase fruit weight, soluble solids, and vitamin C content while reducing organic acid content, especially in the cultivar ‘Black Charcoal’, indicating that genotype-specific “light recipe” management has considerable application potential (Tang et al., 2025). Insect- and rain-proof nets can reduce pest damage and fruit cracking, improve fruit size, sugar-acid ratio, and economic returns, and simultaneously reduce the risks associated with sugar-consuming and pathogenic microorganisms by regulating fruit-surface microbial communities, thereby achieving coordinated regulation of microenvironment, microbial populations, and fruit quality (Yu et al., 2021). Intercropping with ryegrass can improve rhizosphere nutrient conditions, microbial structure, and metabolite composition, while increasing fruit sugars, vitamin C, and flavonoid contents and reducing acidity, thus providing an ecological model for region-specific precision water and fertilizer management (Li et al., 2023).

Intelligent sensing and robotic technologies will also play increasingly important roles in Chinese bayberry orchard management. A Chinese bayberry fruit recognition model based on an improved YOLOv7-tiny network and attention mechanisms achieved a fruit detection recall rate of 97.6% under natural conditions, while maintaining a lightweight architecture suitable for deployment on mobile harvesting robots (Zheng et al., 2025). In addition, lightweight instance segmentation models combined with multi-feature regression can be used for field fruit maturity recognition, thereby providing real-time decision support for timely harvesting and selective picking. In the future, considering the complex canopy structure of Chinese bayberry trees, severe fruit occlusion, concentrated harvesting periods, and the susceptibility of fruits to mechanical damage, research should focus on the development of maturity recognition, intelligent harvesting, non-destructive sorting, and integrated postharvest cold-chain systems.

### 7.3 Construction of green and high-quality production systems

The construction of green and high-quality production systems is an important direction for achieving sustainable development in the Chinese bayberry industry. Although traditional high-input cultivation systems have improved yield to some extent, excessive application of chemical fertilizers and pesticides may lead to soil degradation, increased ecological pressure, and declines in fruit quality. Future Chinese bayberry production should place greater emphasis on resource-use efficiency, ecological environmental protection, fruit safety, and farmer profitability, thereby promoting the transformation of the industry from a purely high-yield orientation toward green, high-quality, efficient, and sustainable development (Liu et al., 2020; Sharma and Shivandu, 2024).

The establishment of green production systems requires greater adoption of organic fertilizer substitution, biological control, ecological regulation, and orchard biodiversity management. Ryegrass intercropping can improve soil structure and rhizosphere microecological environments while increasing fruit sugars, vitamin C, and flavonoid contents and reducing acidity, thereby achieving both ecological benefits and quality improvement simultaneously (Li et al., 2023). Insect- and rain-proof net technologies not only improve yield and quality but also reduce pesticide use and pest contamination, making them more compatible with green and low-residue production systems (Yu et al., 2021). Therefore, future green production of Chinese bayberry should focus on developing integrated technical systems combining “organic fertilizer substitution-orchard grass cover-biological control-rain shelter cultivation-precision water and fertilizer management-intelligent monitoring”.

Future green and high-quality production systems should also be integrated with postharvest preservation, cold-chain logistics, deep processing, and brand development. Chinese bayberry possesses high nutritional and

health-promoting value and shows considerable potential in functional foods, juices, wines, vinegars, dried fruit, and bioactive compound extraction (Zhang et al., 2022). Therefore, further studies on harvesting, preservation, storage, transportation, and processing technologies are needed to extend the industrial chain and increase product added value. Meanwhile, the establishment of unified green production standards, quality evaluation systems, and regional branding systems could enhance consumer trust and promote the transformation of the Chinese bayberry industry from a fresh-fruit-oriented market toward a diversified model integrating “fresh consumption + processing + functional products + ecological branding.”

## 8 Concluding Remarks

In recent years, research on Chinese bayberry fruit quality has gradually evolved from the early basic description of ripening changes to multi-omics analyses of the genetic and biochemical regulatory mechanisms underlying fruit quality formation. Early EST and RNA-Seq studies identified a large number of differentially expressed genes during fruit ripening, revealing coordinated upregulation of anthocyanin biosynthesis-related genes and significant alterations in sugar and organic acid metabolism, which together determine fruit color and flavor quality. With the development of high-quality genome assembly technologies, including the recent telomere-to-telomere (T2T) reference genome and earlier draft genomes, researchers have further elucidated antioxidant, flavonoid, and terpenoid metabolic pathways and identified key expanded gene families and candidate regulatory factors associated with antioxidant capacity and flavor traits. Genome-wide association studies (GWAS) based on large germplasm populations linked 29 measurable fruit quality traits to more than 1,000 genes and identified important MYB/MLP loci closely associated with fruit color formation. At the same time, studies of the MYB family revealed key transcription factors regulating metabolic flux between anthocyanins and flavonols. In addition, the development of molecular marker technologies such as SSR, EST-SSR, SNP, and InDel markers, together with the establishment of multi-omics databases, has further promoted germplasm identification, genetic diversity analysis, and molecular breeding research in Chinese bayberry, thereby providing a more robust platform for fruit quality improvement.

In addition to genetic research, cultivation management and postharvest handling practices have gradually been recognized as important approaches for regulating Chinese bayberry fruit quality. Field studies have shown that insect- and rain-proof nets can not only effectively control major pests but also improve fruit size, edible rate, and sugar-acid ratio while optimizing fruit-surface microbial community structure, thereby significantly enhancing commercial fruit quality and yield. Protected cultivation and LED supplemental lighting technologies can precisely regulate orchard microenvironments and fruit coloration parameters, significantly increasing fruit weight, soluble solids, and vitamin C content while reducing acidity in some sensitive cultivars, although responses differ considerably among cultivars. Postharvest studies have demonstrated that ethylene regulation and low-temperature storage conditions play crucial roles in maintaining fruit quality. Cold storage at approximately 4°C can slow firmness decline, stabilize sugar-acid changes, and extend shelf life, indicating that storage environment is an important factor influencing consumer acceptance. In addition, studies on flesh compartment development and hormonal regulation suggest that cultivation practices affecting hormonal balance, canopy structure, and source-sink relationships interact with intrinsic fruit developmental programs and jointly determine final fruit texture and juiciness. Overall, optimized cultivation and postharvest management technologies tailored to different cultivars are equally important as genetic background for achieving stable production of high-quality Chinese bayberry fruit.

Looking forward, the integration of genomics, physiology, and modern agricultural technologies will provide broad prospects for upgrading the Chinese bayberry industry. Applications of high-resolution genomes, GWAS signals, and high-density molecular markers make marker-assisted breeding and even genomic selection possible, helping overcome the long juvenile period of Chinese bayberry and accelerating the breeding of superior cultivars with improved flavor, coloration, antioxidant capacity, and environmental adaptability. At the same time, abundant local germplasm resources and regional genetic diversity provide important foundations for exploiting heterosis and selecting superior hybrid combinations. In the production sector, integrating insect- and rain-proof nets, precise light management, and low-input green cultivation systems with optimized postharvest handling

technologies can further improve fruit safety and added value while reducing postharvest losses. Moreover, studies on Chinese bayberry fruits, kernels, and processing by-products have demonstrated broad application potential in functional foods and nutraceutical products, particularly due to their strong antioxidant and antidiabetic activities, which may significantly extend the industrial value chain of Chinese bayberry. In the future, further efforts are still needed to clarify the mechanisms underlying fruit quality formation, establish region- and cultivar-specific cultivation systems, and develop unified quality evaluation and grading standards linking genetic markers, laboratory detection indices, and market grades, thereby promoting the high-quality and sustainable development of the Chinese bayberry industry.

## Acknowledgments

The author acknowledges GenBreed Publisher for providing editorial assistance.

## Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

- Apostolopoulos I.D., Tzani M.A., and Aznaouridis S.I., 2023, A general machine learning model for assessing fruit quality using deep image features, *AI*, 4(4): 812-830.  
<https://doi.org/10.3390/ai4040041>
- Cao Y.L., Jia H.M., Xing M.Y., Jin R., Grierson D., Gao Z.S., Sun C.D., Chen K.S., Xu C.J., and Li X., 2021, Genome-wide analysis of *MYB* gene family in Chinese bayberry (*Morella rubra*) and identification of members regulating flavonoid biosynthesis, *Frontiers in Plant Science*, 12: 691384.  
<https://doi.org/10.3389/fpls.2021.691384>
- Chen Y.C., Xiang L., Li F., Chang Y.J., Yu H.G., Zhang J., and Xie Z.L., 2025, The appropriate reduction of nitrogen fertilization enhances soil quality without compromising fruit yield and quality in a bayberry orchard, *Polish Journal of Environmental Studies*, 35(3): 3537-3549.  
<https://doi.org/10.15244/pjoes/204242>
- Dhiman B., Kumar Y., and Kumar M., 2022, Fruit quality evaluation using machine learning techniques: review, motivation and future perspectives, *Multimedia Tools and Applications*, 81: 16255-16277.  
<https://doi.org/10.1007/s11042-022-12652-2>
- Gao J.P., Zheng X.A., Jiang A.Z., Rong J., Yue W., Cao J.P., and Sun C.D., 2024, Characterization of flavor quality deterioration of postharvest Chinese bayberry (*Myrica rubra* cv. Dongkui) at different storage temperatures, *Journal of Food Composition and Analysis*, 130: 106146.  
<https://doi.org/10.1016/j.jfca.2024.106146>
- Hassan E., Ghazalah S., El-Rashidy N., El-Hafeez T.A., and Shams M.Y., 2025, Sustainable deep vision systems for date fruit quality assessment using attention-enhanced deep learning models, *Frontiers in Plant Science*, 16: 1521508.  
<https://doi.org/10.3389/fpls.2025.1521508>
- Hong L.D., Yao Y.L., Lei C.T., Hong C.L., Zhu W.J., Zhu F.X., Wang W.P., Lu T., and Qi X.J., 2023, Declined symptoms in *Myrica rubra*: the influence of soil acidification and rhizosphere microbial communities, *Scientia Horticulturae*, 311: 111892.  
<https://doi.org/10.1016/j.scienta.2023.111892>
- Júnior M.S., Santos R., De Azevedo Sales L., Vargas R., Deltsidis A., and De Oliveira L.F., 2025, Image-based and ML-driven analysis for assessing blueberry fruit quality, *Heliyon*, 11: e42288.  
<https://doi.org/10.1016/j.heliyon.2025.e42288>
- Knott M., Pérez-Cruz F., and Defraeye T., 2023, Facilitated machine learning for image-based fruit quality assessment, *Journal of Food Engineering*, 345: 111401.  
<https://doi.org/10.1016/j.jfoodeng.2022.111401>
- Li C.X., Li G., Qi X.J., Yu Z.P., Abdallah Y.A.Y., Ogunyemi S.O., Zhang S.W., Ren H.Y., Mohany M., Al-Rejaie S.S., Li B., and Liu E.M., 2023, The effects of accompanying ryegrass on bayberry trees by change of soil property, rhizosphere microbial community structure, and metabolites, *Plants*, 12(21): 3669.  
<https://doi.org/10.3390/plants12213669>
- Liu Y., Sun D., Wang H., Wang X., Yu G., and Zhao X., 2020, An evaluation of China's agricultural green production: 1978-2017, *Journal of Cleaner Production*, 243: 118483.  
<https://doi.org/10.1016/j.jclepro.2019.118483>
- Mo J.L., Rashwan A.K., Osman A.I., Eletmany M.R., and Chen W., 2024, Potential of Chinese bayberry (*Myrica rubra* Sieb. et Zucc.) fruit, kernel, and pomace as promising functional ingredients for the development of food products: a comprehensive review, *Food and Bioprocess Technology*, 17: 3506-3524.  
<https://doi.org/10.1007/s11947-023-03313-9>
- Ren H.Y., He Y.H., Qi X.J., Zheng X.L., Zhang S.W., Yu Z.P., and Hu F.R., 2021, The bayberry database: a multiomic database for *Myrica rubra*, an important fruit tree with medicinal value, *BMC Plant Biology*, 21: 452.  
<https://doi.org/10.1186/s12870-021-03232-x>

- Saeed M., Elsadek M.F., Chen Z., Zhao L., Wang G., Zhou C., Sun D., Gao Z., and Jiao Y., 2025, Enhancing the terpenoid and flavonoid profiles and fruit quality in an elite Chinese bayberry line through hybridization, *Food Chemistry*, 479: 143784.  
<https://doi.org/10.1016/j.foodchem.2025.143784>
- Saeed M., Zhao H., Chen Z., Ju P., Wang G., Zhou C., Jia H., Zhu C., Jia H., Jiao Y., Gao Z., and Zhao L., 2023, Wax bayberry is a suitable rootstock for Chinese red bayberry cultivated in saline-alkali soil, *Scientia Horticulturae*, 321: 112463.  
<https://doi.org/10.1016/j.scienta.2023.112463>
- Saeed M., Zhao L., Rashwan A.K., Osman A.I., Chen Z., Wang G., Zhou C., Tu T., Alabd A., Jiao Y., and Gao Z., 2024, Ethylene-induced postharvest changes in five Chinese bayberry cultivars affecting the fruit ripening and shelf life, *Horticulturae*, 10(11): 1144.  
<https://doi.org/10.3390/horticulturae10111144>
- Sharma K., and Shivandu S.K., 2024, Integrating artificial intelligence and internet of things (IoT) for enhanced crop monitoring and management in precision agriculture, *Sensors International*, 5: 100292.  
<https://doi.org/10.1016/j.sintl.2024.100292>
- Soussi A., Zero E., Sacile R., Trincherio D., and Fossa M., 2024, Smart sensors and smart data for precision agriculture: a review, *Sensors*, 24(8): 2647.  
<https://doi.org/10.3390/s24082647>
- Sun L., Zhang S.W., Yu Z.P., Zheng X.L., Liang S.M., Ren H.Y., and Qi X.J., 2024, Transcription-associated metabolomic analysis reveals the mechanism of fruit ripening during the development of Chinese bayberry, *International Journal of Molecular Sciences*, 25(16): 8654.  
<https://doi.org/10.3390/ijms25168654>
- Suo K., Zhang Y., Feng Y.B., Yang Z.F., Zhou C.S., Chen W., and Wang J.C., 2023, Ultrasonic synergistic slightly acidic electrolyzed water processing to improve postharvest storage quality of Chinese bayberry, *Ultrasonics Sonochemistry*, 101: 106668.  
<https://doi.org/10.1016/j.ultsonch.2023.106668>
- Tang N., Hao C.C., and Qiu R., 2025, Enhancement of growth and quality of Chinese bayberry using LED supplemental lighting, *Phyton-International Journal of Experimental Botany*, 94(8): 2551-2562.  
<https://doi.org/10.32604/phyton.2025.070556>
- Wang J., Wu W., Tian S., He Y., Huang Y., Wang F., and Zhang Y., 2023, Non-destructive determination of bayberry sugar and acidity by hyperspectral remote sensing of Si-sensor and low-cost portable instrument development, *Sensors*, 23(24): 9822.  
<https://doi.org/10.3390/s23249822>
- Wu B., Zhang C., Gao Y., Zheng W., and Xu K., 2021, Changes in sugar accumulation and related enzyme activities of red bayberry (*Myrica rubra*) in greenhouse cultivation, *Horticulturae*, 7(11): 429.  
<https://doi.org/10.3390/horticulturae7110429>
- Wu D., Cheng H., Chen J., Ye X., and Liu Y., 2019, Characteristics changes of Chinese bayberry (*Myrica rubra*) during different growth stages, *Journal of Food Science and Technology*, 56: 654-662.  
<https://doi.org/10.1007/s13197-018-3520-4>
- Xia W., Gong E., Lin Y., Li T., Lian F., Zheng B., and Liu R.H., 2021, Comparison of phytochemical profiles, antioxidant and antiproliferative activities in Chinese bayberry (*Myrica rubra* Sieb. et Zucc.) fruits, *Journal of Food Science*, 86(10): 4691-4703.  
<https://doi.org/10.1111/1750-3841.15899>
- Xuan X., Sun R., Zhang X., Cui Y., Lin X., Sun Y., Deng W., Liao X., and Ling J., 2022, Novel application of HS-GC-IMS with PCA for characteristic fingerprints and flavor compound variations in NFC Chinese bayberry (*Myrica rubra*) juice during storage, *LWT*, 167: 113882.  
<https://doi.org/10.1016/j.lwt.2022.113882>
- Xue L., Liu X., Wang W., Huang D., Ren C., Huang X., Yin X., Lin-Wang K., Allan A.C., Chen K., and Xu C., 2024, MYB transcription factors encoded by diversified tandem gene clusters cause varied *Morella rubra* fruit color, *Plant Physiology*, 195(1): 598-616.  
<https://doi.org/10.1093/plphys/kiad063>
- Yang H., Sun L., Qi Y., Li Z., Lei K., Cheng F., Wu Y., Ying Z., Lei Y., Ahmed T., Yu Z., Qi X., and Zhang S., 2025, Integrated transcriptomic and metabolomic analysis reveals light-induced modulation of anthocyanin biosynthesis in Chinese bayberry (*Myrica rubra*), *Fruit Research*, 5: e015.  
<https://doi.org/10.48130/frures-0025-0004>
- Yu H.Y., Tian S.K., Huang Q.B., Chen J.Z., Wu Y.P., Wang R.Z., and Lu L.L., 2021, An insect- and rain-proof net raises the production and quality of Chinese bayberry by preventing damage from insects and altering bacterial communities, *Frontiers in Plant Science*, 12: 732012.  
<https://doi.org/10.3389/fpls.2021.732012>
- Yuan L., Fu X., Zuo X., Jiang Q., Ji H., Chen X., Jiang C., Xie Z., and Chen X., 2025, Non-destructive assessment of bayberry quality using hyperspectral imaging analysis: from individual to template-packaged product via model transfer, *Food Chemistry*, 497: 146965.  
<https://doi.org/10.1016/j.foodchem.2025.146965>
- Zhang S., Yu Z., Qi X., Wang Z., Zheng Y., Ren H., Liang S., and Zheng X., 2021, Construction of a high-density genetic map and identification of leaf trait-related QTLs in Chinese bayberry (*Myrica rubra*), *Frontiers in Plant Science*, 12: 675855.  
<https://doi.org/10.3389/fpls.2021.675855>
- Zhang S.W., Yu Z.P., Sun L., Liang S.M., Xu F., Li S.J., Zheng X.L., Yan L.J., Huang Y.H., Qi X.J., and Ren H.Y., 2024, T2T reference genome assembly and genome-wide association study reveal the genetic basis of Chinese bayberry fruit quality, *Horticulture Research*, 11(3): uhae033.  
<https://doi.org/10.1093/hr/uhae033>

- Zhang S.W., Yu Z.P., Sun L., Ren H.Y., Zheng X.L., Liang S.M., and Qi X.J., 2022, An overview of the nutritional value, health properties, and future challenges of Chinese bayberry, PeerJ, 10: e13070.  
<https://doi.org/10.7717/peerj.13070>
- Zheng H., Sun L., Wang Y., Yang H., and Zhang S.W., 2025, Image-based detection of Chinese bayberry (*Myrica rubra*) maturity using cascaded instance segmentation and multi-feature regression, Horticulturae, 11(10): 1166.  
<https://doi.org/10.3390/horticulturae11101166>



#### **Disclaimer/Publisher's Note**

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---





GenBreed Publisher®

## Reasons to publish in GenBreed Publisher *An Online Publishing Platform*

- ★ Peer review quickly and professionally
- ☆ Publish online immediately upon acceptance
- ★ Deposit permanently and track easily
- ☆ Access free and open around the world
- ★ Disseminate multilingual available

Submit your manuscript at: <http://genbreedpublisher.com/>

