Scientia Horticulturae 358 (2026) 114702

Contents lists available at ScienceDirect

Scientia Horticulturae

journal homepage: www.elsevier.com/locate/scihorti

Research Paper

Integration of phenotype characterization and DNA barcoding for comprehensive genetic diversity assessment in *Hydrangea* germplasm

Yufei Zhang [a,b,1], Bo Jiang [b,1,*], Jiayue Zhang [b], Beier Zhou [b], Biao Li [c], Yue Wu [b], Yichen Li [d], Ye Jiang [b], Huanghui Lu [e], Erxu Pi [a,*]

[a] College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China
[b] College of Biology and Food Engineering, Suzhou University of Technology, Suzhou 215500, China
[c] Management Office of Changshu National Agricultural Science & Technology Park, Suzhou, 215500, China
[d] Changshu Agricultural Technology Extension Center, Suzhou, 215500, China
[e] Zhimali Forestry Workstation, Guangze County, Fujian 354101, China

ARTICLE INFO

Key words:
Hydrangea
DNA barcoding
DUS testing
Genetic diversity
Phenotypic variation
Variety identification

ABSTRACT

Hydrangea has become a focus of horticultural research owing to its exceptional ornamental value. However, analysis of genetic diversity in *Hydrangea* using DNA barcode markers is limited. This study systematically investigated the genetic diversity and phylogenetic relationships among *Hydrangea* varieties using a combination of distinctness, uniformity, and stability (DUS) phenotypic analysis and DNA barcode markers. Diversity analysis, correlation analysis, and principal component analysis (PCA) were conducted on 24 phenotypic traits for 50 accessions. The coefficient of variation for the DUS traits ranged from 12.37% to 132.34%. The simpson diversity index indicated that floral organ traits showed the most significant variation and highlighted the substantial genetic diversity within the *Hydrangea* population. PCA identified 11 core traits explaining 81.97% of the total variation, demonstrating their utility as primary evaluation indicators. Compared to *rbc*L, *psb*A–*trn*H, and their combinations, *mat*K exhibited superior discriminatory potential, making it a more effective specific DNA barcode marker; consistent with this, amplification of 267 DNA sequences from 89 samples revealed that *mat*K and *psb*A-*trn*H displayed the highest mutation rates and optimal identification efficiency. PCA, phylogenetic reconstruction, and the Mantel test identified *mat*K and *psb*A–*trn*H as optimal markers for cultivar identification, which showed significant correlations with phenotypic traits. Haplotype analysis of the three DNA fragments revealed 13 haplotypes. The haplotype comprised three dominant and 10 unique haplotypes. The results revealed substantial genetic diversity among the tested cultivars and the utility of multiple genetic markers for variety differentiation. The molecular marker and genetic diversity information provide crucial insights useful for future *Hydrangea* breeding programs.

1. Introduction

Hydrangeas are widely cherished in the global horticultural market for their diverse flower colors and shapes. *Hydrangea* cultivars are extremely popular as ornamental flowering shrubs. Among the approximately 70–80 *Hydrangea* species, *Hydrangea macrophylla* (Thunb.) Ser. is a deciduous, semi-evergreen shrub with large flower clusters. *H. macrophylla* is the most economically important species with hundreds of horticultural and cut-flower varieties in cultivation, and has long been a cornerstone of germplasm resource utilization (Liao et al., 2025; Wu et al., 2023). According to the 2019 USDA Horticultural Census, *Hydrangeas* sold in U.S. nurseries were valued at $106,862,000, ranking them as the third most commercially valuable deciduous shrub (USDA, 2019), highlighting their immense economic value in the horticultural industry. Despite their commercial importance, interspecific phylogenetic relationships and the evolutionary significance of interspecific hybridization and gene flow among *Hydrangea* species remain controversial (Huang et al., 2025).

* Corresponding authors.
E-mail addresses: wwlyl885522@163.com (Y. Zhang), jiang19810@163.com (B. Jiang), 18852985436@163.com (J. Zhang), z2383890456@163.com (B. Zhou), 13812800899@139.com (B. Li), 13776049480@163.com (Y. Wu), 1127473446@qq.com (Y. Li), 3091621503@qq.com (Y. Jiang), 32849909@qq.com (H. Lu), pierzaixian001@163.com (E. Pi).
[1] These authors contributed equally.

https://doi.org/10.1016/j.scienta.2026.114702
Received 24 December 2025; Received in revised form 12 February 2026; Accepted 16 February 2026
Available online 20 February 2026

Modern *Hydrangea* breeding faces dual challenges. First, more than 90% of existing cultivars originate from inbreeding within *H. macrophylla*, resulting in a narrow genetic foundation and limited genetic gains for stress resistance traits. DNA barcoding is a powerful tool for species identification and biodiversity conservation, particularly in plant species that are genetically diverse but taxonomically understudied (Hartvig et al., 2015; Qin et al., 2022; Rather et al., 2023). The continuous selection of new cultivars, coupled with widespread habitat conversion, is accelerating the depletion of wild gene pools. Therefore, the conservation of existing genetic diversity across populations and cultivars is crucial. Conservation practitioners must identify and prioritize the long-term persistence and adaptability of germplasm resources with unique genetic value, which may involve establishing gene banks, designating protected areas, or implementing conservation breeding programs (van Oosterhout et al., 2025). As Ellegren and Galtier (2016) emphasized, conservation of plant species with unique genetic traits maintains biodiversity and provides valuable genetic resources for breeding research. Furthermore, aligning cultivar genetics with specific growth conditions and application requirements optimizes resource utilization, thereby improving production efficiency and practical benefits. Previous studies (Rinehart et al., 2006; Wu and Alexander, 2020) validate this perspective, stressing the critical importance of genetic diversity when selecting *Hydrangea* varieties for different geographical regions and horticultural applications.

The guidelines for distinctiveness, uniformity, and stability (DUS) testing of *Hydrangeas*, established by the International Union for Plant Variety Protection, provide a systematic and standardized research methodology for evaluating morphological traits of *Hydrangea* varieties through phenotypic analysis (UPOV 2020). These guidelines serve as a vital tool for variety identification, classification, and protection (Yang et al., 2021; Zhu et al., 2023). Chen et al. (2023) applied DUS testing to impatiens to conduct a statistical analysis of quantitative traits, to validate the applicability of the UPOV guidelines, and to propose recommendations for the addition of new group traits. With the advent of the big data era, the rapid expansion in data generation and collection sources has led to a dramatic increase in data volume, explosive growth in feature dimensions, and expansion of label space, posing significant challenges to traditional learning paradigms (Dabbu et al., 2022; Zheng et al., 2024). DUS testing has shown considerable potential in phenotypic diversity analysis of germplasm resources in multiple species (He et al., 2022; Tian et al., 2024; Zheng et al., 2025). Tang et al. (2023) conducted genome resequencing on 125 *Hydrangea* accessions. Using high-throughput genotyping data, the authors identified single-nucleotide polymorphism (SNP) loci that were significantly associated with *Hydrangea* inflorescence types, thereby establishing genotype–phenotype correlations. This molecular-level evidence supports species identification and breeding efforts. UPOV and multinational testing institutions have widely applied DUS technology in constructing germplasm reference collections for various crops (Kapoor et al., 2025; Zheng et al., 2025), enabling efficient identification of new variety traits and protection of variety rights through parallel field phenotyping and molecular marker genotyping. Wang et al. (2022) assessed trends in artificial selection and genetic diversity in wheat varieties based on a DUS trait analysis.

DNA barcoding technology, which utilizes specific loci such as *mat*K, *rbc*L, and the *psb*A–*trn*H intergenic spacer, has become a robust molecular tool for studying genetic diversity in *Hydrangea*. These gene loci, characterized by conserved regions and variable intercalated sequences, are widely used in plant molecular research for genetic discrimination of species and cultivars. Variation in the nucleotide sequence reflects genetic mutations and temporal differentiation patterns. The *mat*K gene, located in the chloroplast genome, is informative in phylogenetic and taxonomic studies (Muino et al., 2024), and thus may reveal evolutionary relationships among cultivars. Comparison of *mat*K sequences across cultivars enables researchers to determine genetic distances and infer phylogenetic positions. For instance, Wilson (2004) conducted a phylogenetic analysis of all subgenera and 46 species of *Iridium* (excluding three specific groups) using *mat*K and *trn*K intron sequences. In a comparative and phylogenetic analysis of the chloroplast genome of *Paeonia* species, Wu et al. (2021) identified *mat*K as the most suitable gene among 14 high-transmission DNA regions for phylogenetic analysis, with results consistent with those from full chloroplast genome studies. The *rbc*L gene, also present in the chloroplast genome, has been widely utilized in plant systematics research (Clegg, 1993; Kang et al., 2017; Yong et al., 2024). Using the closely related genera of the Hydrangeaceae as an example, the *rbc*L gene has assisted with the resolution of species relationships within *Hydrangea* and of relationships with closely related genera. Based on combined data from chloroplast genes *mat*K and *rbc*L, Hufford et al. (2001) used a combined *mat*K–*rbc*L sequence dataset to analyze the phylogenetic relationships among 42 species representing 10 genera in the Hydrangeaceae. Their analysis supported the division of Hydrangeaceae into two major subfamilies: Jamesioideae (*Jamesia* and *Fendlera*) and Hydrangeoideae (all remaining genera). Although the resolution may be slightly inferior to other genetic loci used for discriminating closely related species, the two regions are sufficiently informative for understanding the genetic framework of *Hydrangea*. Previous studies have analyzed the highly variable *psb*A–*trn*H intergenic spacer in plant species (Pang et al., 2012; Zhang et al., 2025). This region is crucial for distinguishing closely related cultivars within the *Syringa* genus (Yao et al., 2022). The *psb*A–*trn*H region is among the fastest-evolving non-coding regions in the chloroplast genome, and comprises a relatively conserved 3' untranslated region (UTR) of the *psb*A gene and a highly variable non-transcriptional spacer sequence (Kress and Erickson, 2007; Shaw et al., 2007). This region is widely used in plant DNA barcoding studies owing to its sequence variation (including SNPs and insertion/deletion events) to achieve efficient interspecific and even population-level discrimination (UPOV, 2009). Hirota et al. (2022) confirmed that *psb*A–*trn*H sequences can effectively distinguish multiple morphologically similar *Hydrangea* varieties, including an accession identified as *H. serrata*, demonstrating its potential application for variety identification and assessment of genetic diversity. However, the prevailing research on *Hydrangeas* predominantly employs internal transcribed spacer (ITS), *trn*L–*trn*F, or single *mat*K fragments, without systematic evaluation of the utility of the combined *mat*K+*rbc*L+*psb*A–*trn*H loci in *Hydrangea*.

Phenotypic traits are environmentally sensitive, which may lead to misidentification or misinterpretation (Kroon et al., 2025; Sommer, 2020). For instance, the floral color of *H. macrophylla* varies with soil nutrient concentrations, temperature, and humidity, and thus identification solely on this trait is unreliable. In contrast, DNA barcoding provides a stable, reliable genetic fingerprint that is unaffected by environmental factors (Abdi et al., 2024; Antil et al., 2023). Integration of phenotypic and molecular data enables a more robust analysis of genetic diversity and phylogenetic relationships among *Hydrangea* species. Multiple studies support the value of such an approach. Shen et al. (2023) successfully distinguished 16 rhododendron varieties by combining *mat*K+*psb*A–*trn*H barcoding with leaf epidermal micromorphological traits. Mohanty et al. (2023) analyzed phenotypic traits and Kim-*mat*K+ITS2 sequences for wild relatives of eggplant, demonstrating that this DNA barcode achieved 100% molecular differentiation among closely related species with minor morphological differences. The integration of DNA barcoding (or more broadly, DNA fingerprinting) with traditional phenotypic DUS testing is emerging as a powerful approach for plant variety identification, conservation, and genetic resource research (Antil et al., 2023; Chac and Thinh, 2023; Liberty et al., 2025). This integrated approach aims to overcome the limitations of traditional DUS testing, such as prolonged growth cycles and environmental sensitivity, and provide a more efficient and accurate technical solution for plant biodiversity and germplasm conservation. This combination offers complementary advantages: it addresses the limitations of exclusively phenotypic-trait DUS testing in distinguishing highly similar varieties and overcomes the challenge that variation in DNA barcodes

may fail to capture subtle phenotypic variation.

2. Materials and methods

2.1. Plant materials, DNA extraction

A total of 50 *Hydrangea* germplasm resources (accessions *Hydr*V01 to *Hydr*V50) were collected for this study. These materials, provided by the Germplasm Resource Bank of Suzhou Institute of Technology, were used for phenotypic and molecular data analysis (Table 1). The plants were maintained in a controlled greenhouse at the Suzhou Institute of Technology with the following parameters: temperature 20–25 °C, relative humidity 60%–70%, natural light supplemented with artificial lighting to provide a 12-h photoperiod. Plants were irrigated daily using an automatic drip irrigation system to maintain consistent soil moisture. Fertilization was performed weekly using a balanced NPK fertilizer (20:20:20) applied at a rate of 2 g L^{-1}. All plants were potted in the same commercial potting mix (peat:perlite:vermiculite, 3:1:1, v/v/v) and arranged in a completely randomized design to avoid positional effects. The accessions comprised 44 *H. macrophylla* cultivars, two *H. arborescens* cultivars, two hybrid *Hydrangea* cultivars ('Runaway Bride' and 'French Bolero'), and one *H. paniculata* cultivar.

The evaluation was conducted using a completely randomized block design, with all materials maintained under standardized management conditions at the institute's *Hydrangea* breeding base, where they exhibited good health status. In plant DNA barcoding analysis, an unbalanced biological replication sampling strategy was employed for 50 target cultivars in this study. This sampling strategy effectively ensured data quality and methodological rigor. In molecular identification of germplasm resources, replicate analysis is crucial for controlling technical errors and ensuring data reliability (Hollingsworth et al., 2011). For groups with confirmed stable diagnostic sequences or those maintained through clonal propagation, a single sample can provide sufficient information for effective identification (Hempel et al., 2018). Researchers conducted DNA barcoding analysis on 89 plant samples, with sequence accession numbers listed in Table S1. The DNA barcoding primers were synthesized by Tsingke Biotechnology Co., Ltd. (Table 2). All samples were processed as follows: 0.5 grams of leaves were homogenized with liquid nitrogen, followed by genomic DNA extraction using the DNA Quick Plant System (TIANGEN BIOTECH Co., Ltd., Beijing, China) (Sun et al., 2024).

2.2. Assessment of phenotypic traits

The description and classification of plant traits adhered to the "Guidelines for testing the specificity, consistency and stability of new plant varieties - Hydrangea"(Chinese Standard LY/T 3397-2024) (National Forestry and Grassland Administration, 2024). The trait table was adapted based on the actual expression observed in the experimental materials. Following established guidelines, this study conducted a two-year morphological analysis of 24 DUS traits across 50 *Hydrangea* varieties. Researchers selected 10 plants from each candidate variety for morphological evaluation, using average values as the basis for data analysis. A total of 24 key traits encompassing plant, leaf, inflorescence, and other characteristics were ultimately selected (see Table 2 for details, including UPOV codes and descriptions). All traits were rated using the internationally recognized 9-point scale (UPOV, TG/1/3). When the plants reached maturity in their second year, 24 phenotypic traits were recorded for 10 plants randomly selected for each variety. The phenotypic traits consisted of seven qualitative and 17 quantitative traits. The qualitative traits comprised: Plant type; Stem flattening; Stem anthocyanin coloration in leaf axils (for green-stemmed varieties); Stem coat; Leaf: Coats on the back; Petiole color; and Sterile flower number of sepals. The quantitative traits comprised: Number of lenticels in stem; Stem lenticel size; Leaf length; Leaf width; Leaf tip length; Leaf margin serration; Leaf anthocyanin coloration; Leaf adaxial

Table 1
Summary information on the *Hydrangea* accessions used in the study.

No.	Species	Varieties	Sample name	DNA barcoding	
				Number of individuals	Material number
1	*H. Macrophylla*	HL"An inch of wind"	*Hydr*V01	3	R1-R3
2	*H. Macrophylla*	HL"Lei Yin"	*Hydr*V02	3	R4-R6
3	*H. Macrophylla*	HL"Lala"	*Hydr*V03	3	R7-R9
4	*H. Macrophylla*	HL"Charlotte"	*Hydr*V04	3	R10-R12
5	*H. Macrophylla*	HL"Lagoon"	*Hydr*V05	3	R13-R15
6	*H. Macrophylla*	HL"Aisha"	*Hydr*V06	3	R16-R18
7	*H. Macrophylla*	HL"Dancing snow"	*Hydr*V07	3	R19-R21
8	*H. Arborescens*	HS"Pink elf"	*Hydr*V08	3	R22-R24
9	*H. Macrophylla*	HL"Da Lai"	*Hydr*V09	3	R25-R27
10	*H. Macrophylla*	HL"Wing"	*Hydr*V10	3	R28-R30
11	*H. Macrophylla*	HL"Lej"	*Hydr*V11	3	R31-R33
12	*H. Macrophylla*	HL"Miss saori"	*Hydr*V12	3	R34-R36
13	*H. Macrophylla*	HL"Xian Jian"	*Hydr*V13	3	R37-R39
14	*H. Macrophylla*	HL"Cocktail"	*Hydr*V14	3	R40-R42
15	*H. Macrophylla*	HL"Bon jour"	*Hydr*V15	3	R43-R45
16	*H. Macrophylla*	HL"Elizabeth"	*Hydr*V16	3	R46-R48
17	*H. Macrophylla*	HL"ZhuZi Hongbaoshi"	*Hydr*V17	3	R49-R51
18	*H. Macrophylla*	HL"Hua Chui Xue"	*Hydr*V18	3	R52-R54
19	*H. Macrophylla*	HL"KAMO HINAMATSURI"	*Hydr*V19	3	R55-R57
20	*H. Macrophylla*	HL"Mokana"	*Hydr*V20	2	R58-R59
21	*H. Macrophylla*	HL"Hopcorn"	*Hydr*V21	1	B1
22	*H. Macrophylla*	HL"Ying Ban"	*Hydr*V22	1	B2
23	*H. Macrophylla*	HL"Luxor"	*Hydr*V23	1	B3
24	*H. Macrophylla*	HL"Twist-n-Shout"	*Hydr*V24	1	B4
25	*H. Macrophylla*	HL"Angelica"	*Hydr*V25	1	B5
26	*H. Macrophylla*	HL"Endless Summer 31b"	*Hydr*V26	1	B6
27	*H. Macrophylla*	HL"Summer Crush"	*Hydr*V27	1	B7
28	*H. Macrophylla*	HL"Bloom Struck"	*Hydr*V28	1	B8
29	*H. Macrophylla*	HL"Fukuoka"	*Hydr*V29	1	B9
30	*H. paniculata*	HP"Little lime"	*Hydr*V30	1	B10
31	*H. Macrophylla*	HL"Keiko"	*Hydr*V31	1	B11
32	*H. Arborescens*	HS"Ruby Annabelle "	*Hydr*V32	1	B12
33	*H. Macrophylla*	HL"Hua Bao"	*Hydr*V33	1	B13
34	*H. Macrophylla*	HL"Mei Xiao Xiao Ting"	*Hydr*V34	1	B14
35	*H. Macrophylla*	HL"Lei Wang"	*Hydr*V35	1	B15
36	*H. Macrophylla*	HL"Sharon"	*Hydr*V36	1	B16
37	*Hydr cv. Runaway Bride*	HL × HW"Runaway Bride"	*Hydr*V37	1	B17
38	*H.Macrophylla*	HL"Karen"	*Hydr*V38	1	B18
39	*H. Paniculata*	HP"limelight"	*Hydr*V39	1	B19
40	*H. Macrophylla*	HL"Sang Ni"	*Hydr*V40	1	B20
41	*H. Macrophylla*	HL"Mei Yun"	*Hydr*V41	1	B21
42	*H. Macrophylla*	HL"Yu Dian"	*Hydr*V42	1	B22
43	*Hydr cv. Runaway Bride*	HL × HW"French Bolero"	*Hydr*V43	1	B23
44	*H. Macrophylla*	HL"Hun Li Hua Shu"	*Hydr*V44	1	B24
45	*H. Macrophylla*	HL"Xiao Ti Qin"	*Hydr*V45	1	B25
46	*H. Macrophylla*	HL"Liu Li"	*Hydr*V46	1	B26
47	*H. Macrophylla*	HL"Mu Zhi"	*Hydr*V47	1	B27
48	*H. Macrophylla*	HL"Ling Bo"	*Hydr*V48	1	B28
49	*H. Macrophylla*	HL"Double Dutch"	*Hydr*V49	1	B29
50	*H. Macrophylla*	HL"Perfect"	*Hydr*V50	1	B30

Abbreviations: HL, *Hydrangea macrophylla*; HP, *Hydrangea paniculata*; HS, *Hydrangea arborescens*; HW, *Hydrangea serrata*.

Table 2
Character states for 24 phenotypic traits.

Trait (QL/QN)		Description of characteristics	Rating level (1-9) represents the code and specific description
Qualitative	Plant	Plant type	1:Climbing, 2:Non-climbing
	Stem	Stem flattening	1: Absent, 9: Present
	Stem	Stem: coat	1: Absent, 9: Present
	Leaf	Leaf: lobes	1: Absent, 9: Present
	Leaf	Leaf: Polycolor	1: Absent, 9: Present
	Leaf	Leaf: Coats on the back	1: Absent, 9: Present
	Petioles	Petioles: color	1: Green; 2: Red; 3: Greenish-brown; 4: Black
	Sterile flowers	Sterile flowers: number of sepals	1:3/4; 2:4 3: 4/5; 4: 5/6; 5: ≥7
Quantitative	Stem	Stem: number of skin pores	1: None or very little; 2: Little; 3: Middle; 4: Many; 5: Numerous
	Stem	Stem: cutaneous pore size	1: Small; 2: Intermediate; 3: Big
	Leaf	Leaf: Length	3: Short; 5:Middle; 7: Long
	Leaf	Leaf: width	3: Thin; 5:Middle; 7: Wide
	Leaf	Leaf: Tip length	1: None or short; 2: Middle; 3: Long
	Leaf	Leaf: The degree of serration of the leaf margin	1: Weak; 2: Weaker; 3: Intermediate; 4: Strong; 5: Stronger
	Leaf	Leaf: Anthocyanin coloring	1: Weak; 2: Weaker; 3: Intermediate; 4: Strong; 5: Stronger
	Leaf	Leaf: Glossiness of the upper surface	1: Weak; 2: Intermediate; 3: Strong
	Leaf	Leaf: Bubble degree	1:Weak; 2:Weaker; 3: Intermediate; 4:Strong; 5:Stronger
	Inflorescence	Inflorescence: height	3: Short; 5: Middle; 7: Long
	Inflorescence	Inflorescence: Width	3: Thin; 5: Middle; 7: Wide
	Inflorescence	Inflorescence: Pregnant flowers are distinct	1: Weak; 2: Intermediate; 3: Strong
	Sterile flowers	Sterile flowers: sepals are in a raw posture	1: Upright; 2: Semi-upright; 3: Horizontal
	Sterile flowers	Sterile flowers: sepal size	3: Small; 5: Intermediate; 7: Big
	Sterile flowers	Sterile flowers: sepal folds	1: Weak; 2: Intermediate; 3: Strong
	Sterile flowers	Sterile flowers: notched margin of the sepal	1: Neither; 2: Part of it exists.; 3: Both
	Flowering period	Make the flowering period	3: early; 5: Intermediate; 7: Late

surface glossy; Leaf: Bubble degree; Inflorescence height; Inflorescence width; Distinct fertile flowers in inflorescence; Sterile flowers: sepal folds; Sterile flower sepal size; Sterile flower sepal folds; Flowering period; and Sterile flower notched margin of sepal. The traits were grouped into Plant Type, Stem (five traits), Leaf (six traits), Inflorescence (four traits), and Sterile flower accessory characteristics (eight traits). The partial trait states are illustrated in Fig. 1. The raw data for the 24 DUS traits are presented in Table S2.

The inter-rater reliability was evaluated using Fleiss' Kappa coefficient based on a random 10% sample of the phenotypic data ($n = 120$). Prior to the formal assessment, all three assessors underwent standardized training, including study guides and calibration exercises using standard variety images, until consensus was reached on scoring criteria for all traits. Three professionally trained assessors independently recorded the phenotypic expression state of each trait on a 1–9 scale without knowledge of the variety identities. Fleiss' Kappa (κ) is computed using the following formula (McHugh, 2012). (Eq. 1)

$$K = (P_o - P_e) / (1 - P_e) \quad (1)$$

P_o represents the observed proportion of agreement, and P_e denotes the proportion of agreement expected by chance. The associated κ range is expressed as the 95% confidence interval (95% CI), which is estimated based on the standard error (SE) of κ to assess the reliability of the measurement. The inter-rater reliability was calculated using the irr package for R software (version 4.3.1) (Gamer et al., 2010). According to the evaluation criteria proposed by Landis and Koch (Landis and Koch, 1977), only the traits with Kappa value> 0.60 (indicating "high consistency" or above) were retained for further analysis, while the rest were excluded (Table 3 and Table S4).

2.3. PCR amplification and sequencing

The DNA barcoding primers were synthesized by Tsingke Biotechnology Co., Ltd. (Beijing, China). Primers and references for the PCR protocol are provided in Table 4. PCR amplifications were performed targeting three chloroplast DNA regions (*mat*K, *rbc*L, and *psb*A–*trn*H). The PCR amplification for the *rbc*L, *mat*K and *psb*A–*trn*H regions was performed as follows: The PCR reaction mixture contained 2 × Phanta Flash Master Mix (Dye Plus) (Nanjing Vazyme Biotech Co., Ltd, Luo et al., 2022), 0.5 μL of each primer, 2 μL of template DNA, and 7 μL of ddH_2O.

The thermal cycling conditions were as follows: an initial denaturation at 98°C for 30 seconds, followed by 35 cycles of denaturation at 98°C for 10 seconds, annealing at 56°C for 5 seconds, and extension at 72°C for 5 seconds, with a final extension at 72°C for 1 minute. The PCR reaction was performed using the Sensoquest LabCycler gradient PCR system manufactured by SENSO, Germany. The PCR products were separated by electrophoresis on a 1% agarose gel at 135 V for 20 minutes. Samples showing no or weak bands were subjected to repeated PCR amplification under identical conditions. The purified PCR products were sequenced commercially by Tsingke Biotechnology Co., Ltd. (Beijing, China) using Sanger sequencing on an ABI 3730XL platform.

2.4. Sequence data processing

The raw sequencing quality was evaluated using FastQC (Andrews, 2014), with low-quality bases removed with Trimmomatic version 0.39 (Bolger et al., 2014) based on a Phred score < 20. Multiple sequence alignment was performed using MAFFT version 7.526 (Katoh and Standley, 2013) with default parameters. The aligned sequences were manually edited to trim unaligned ends and regions with excessive gaps to ensure consistent matrix length. Sequence concatenation was conducted in Geneious Prime version 2024.0.5 (Kearse et al., 2012), generating seven independent DNA barcode datasets for comparative analysis. These datasets comprised three individual plastid regions (*rbc*L, *mat*K, and *psb*A–*trn*H) and all four possible combinations thereof (*rbc*L+*mat*K, *rbc*L+*psb*A–*trn*H, *mat*K+*psb*A–*trn*H, and the three-marker combination *rbc*L+*mat*K+*psb*A–*trn*H). DNA barcoding analysis was conducted on the 89 plant samples, with sequence accession numbers listed in Table S1.

2.5. Phylogenetic and genetic distance cluster analysis

Geneious Prime software was used to calculate the genetic distances among the three DNA fragments (Kearse et al., 2012). We selected the Kimura two-parameter model to analyze the frequency distribution of the intraspecific and interspecific distances, thereby identifying overlapping regions in the genetic distances between species. Different combinations of the three fragments were then evaluated for their distribution in the barcoding region of *Hydrangea* based on distance

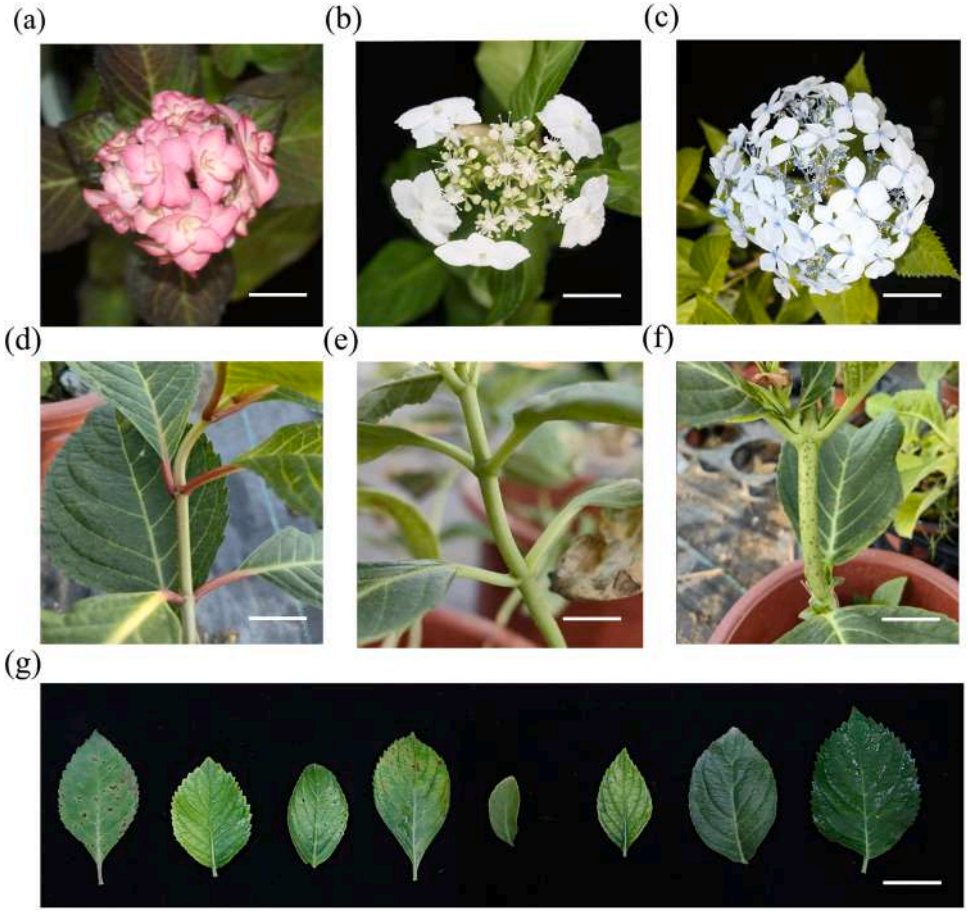

Fig. 1. Stem, leaf, and floral traits during peak blooming. (a) *H. macrophylla* 'Miss Saori', spherical inflorescence with semi-upright growth; (b) *H. macrophylla* 'Da Lai', flat-topped type with prominent central fertile flowers and horizontal inflorescence; (c) *H. macrophylla* 'Hua Chui Xue', spherical inflorescence with moderately prominent fertile flowers; (d) *H. macrophylla* 'Endless Summer-Blooming', notable anthocyanin pigmentation in leaf axils; (e) hybrid 'Runaway Bride', no anthocyanin pigmentation in leaf axils, with small and sparse stem lenticels; (f) *H. macrophylla* 'Elizabeth', numerous small lenticels on stems; (g) representative leaves of varieties (from left to right): 'Twist-n-Shout', 'Sang Ni', 'Hun Li Hua Shu', 'Ying Ban', 'Runaway Bride', 'Perfect', 'Angelica', and 'Liu Li'. Bar = 2 cm.

Table 3
Fleiss' Kappa coefficient evaluated on a random 10% sample of the phenotypic data ($n = 120$).

κ Value range	Consistency interpretation
$\kappa = 1.000$	Perfect
$0.8 \leq \kappa < 1$	Almost perfect
$0.6 \leq \kappa < 0.8$	Medium to high consistency
$\kappa < 0.6$	Eliminate or retrain
κ Value range	Consistency interpretation

Table 4
Primers for the three chloroplast DNA regions analyzed in the study.

ID	Name of the primer	Quotation marks(5′to3′)	bp	References
1	rbcLb-sF	AGACCTTTTTGAAGAAGGTTCTGT	24	Dong et al.
2	rbcLb-sR	TCGGTCAGAGCAGGCATATGCCA	23	(2014)
3	matK-472F	CCCRTYCATCTGGAAATCTTGGTTC	25	Yu et al.
4	matK-1248R	GCTRTRATAATGAGAAAGATTTCTGC	26	(2011)
5	psbA-3(F)	GTTATGCATGAACGTAATGCTC	22	Sang et al.
6	trnH-3(R)	CGCGCATGGTGGATTCACAATCC	23	(1997)

magnitude. IBM SPSS Statistics 27 software (IBM Corporation, 2020) was used to conduct Wilcoxon rank sum tests to assess the significance of differences in variation among the fragment sequences.

Preprocessed DNA sequences were converted into a high-dimensional numerical feature matrix via 3-mer frequency extraction (LaPierre et al., 2019). After standardization, Principal Component Analysis (PCA) was performed to extract the first three principal components and calculate their variance explained ratios. For seven DNA fragment combinations (single fragments: comprising the single fragments *rbc*L, *mat*K, *psb*A–*trn*H; and combinations of fragments: *rbc*L+*mat*K, *rbc*L+*psb*A–*trn*H, *mat*K+*psb*A–*trn*H, and *rbc*L+*mat*K+*psb*A–*trn*H), the principal component eigenvalues and total cumulative variance explained were statistically summarized.

The maximum likelihood (ML) method was used for phylogenetic reconstruction, utilizing the ModelFinder feature in IQ-TREE v3.0.1 (Kalyaanamoorthy et al., 2017) to select the optimal nucleotide substitution model for each dataset based on comparison of Bayesian information criterion values. The ML trees were constructed using IQ-TREE. To assess the support for each node in the tree, we conducted an UltraFast Bootstrap analysis with 1000 replicates. To validate the ML tree topology, Bayesian inference analysis was performed with MrBayes v3.2.7 (Ronquist et al., 2012). Tree adjustment for display purposes was performed with the ITOL online tool (Letunic and Bork, 2024).

2.6. Phenotype data processing and statistical analysis

The Gower distance between each trait was calculated and a distance matrix for the 24 phenotypic traits was generated. A principal coordinate analysis (PCoA) and hierarchical clustering analysis of the distance

matrix were performed to evaluate cultivar similarity based on the phenotypic traits (Tiburtini et al., 2025). For trait association analysis, a Spearman rank correlation analysis was applied for the ordinal traits, the Kruskal–Wallis test was used for ordinal and categorical traits, and Cramer's *V* coefficient was calculated for categorical traits. Raw data were standardized (Z-score normalization) and analyzed using the PCA module in Python 3.13 (Python Software Foundation, 2024). The first three principal components were retained for three-dimensional visualization, as they collectively captured the majority of total variance. Trait loadings were examined to identify the most influential characters. Core traits for DUS detection were identified based on the PCA results. Pairwise comparison matrices were constructed in Excel 2016 to determine DNA barcode genotypes or DUS traits that differed between cultivars. Statistical analyses were performed using Python 3.13, primarily using the scikit-learn (Pedregosa et al., 2011).

2.7. Association analysis of phenotypes and molecular markers

Genetic distance matrices were constructed using pairwise Fixation Index (F_{ST} values, a measure of population genetic differentiation. F_{ST} estimates were calculated for each DNA barcode fragment from aligned sequence data using the method of Weir and Cockerham (Weir and Cockerham, 1984) as implemented in MEGA11 (Tamura et al., 2021). The resulting pairwise F_{ST} values between all cultivar pairs were organized into symmetric matrices.

To analyze the association patterns between the phenotypic and molecular data, we employed the Mantel test (9,999 permutations) to evaluate the significance of correlations between the phenotypic distance matrix (based on the Bray–Curtis algorithm) and molecular genetic distance matrix (based on F_{ST} values) ($\alpha = 0.01$). For all statistical tests (including Wilcoxon and Mantel tests), multiple testing correction was performed using the Benjamini–Hochberg procedure to control the False Discovery Rate (FDR). The raw p-values from all tests were adjusted to generate q-values, with significance declared at $q < 0.05$. This correction was applied uniformly using the multipletests function in the StatsModels 0.14.5 libraries (Seabold & Perktold., 2010).

2.8. SNP analysis, haplotype identification and phenotypic association

Using DnaSP 6.0 software (Rozas et al., 2017), unique DNA sequences were classified into distinct haplotypes. Unique DNA sequences were identified through multiple sequence alignment using ClustalW (Thompson et al., 1994) with default parameters. Sequences with identical alignment patterns across all polymorphic sites were classified into distinct haplotypes. Haplotype diversity (Hd) and nucleotide diversity (π) were calculated using DnaSP with the following parameters: gap handling set to "pairwise deletion" for missing data, and transitions/transversions weighted equally. SNP markers with minor allele frequency < 0.05 or deletion rate $> 20\%$ were excluded to ensure robustness of downstream analyses. Analysis of variance (ANOVA) was used to examine the association between haplotypes and phenotypic traits. All ANOVA analyses were conducted using Python 3.13 with the SciPy 1.16.1 (Virtanen et al., 2020) and StatsModels 0.14.5 libraries (Seabold & Perktold., 2010).

For SNP-phenotype association analysis, a linear mixed model (LMM) was employed. The significance threshold for association was set at a corrected p-value < 0.05. Multiple testing correction was performed using the false discovery rate (FDR) method (Benjamini–Hochberg procedure). All analyses were conducted in Python 3.13, utilizing the following key libraries and versions: scikit-learn (1.7.1), Pandas (1.4.2), NumPy (2.3.2), SciPy (1.16.1), Matplotlib (3.10.5)(Hunter., 2007), Seaborn (0.13.2)(Waskom., 2021), and StatsModels (0.14.5)(Seabold & Perktold., 2010). Figures were generated and visualized using Matplotlib (3.10.5).

3. Results

3.1. Hydrangea phenotypic diversity analysis

This study strictly followed the guidelines of the "Test Guide for Specificity, Consistency, and Stability Testing of New Plant Varieties-*Hydrangea*" (LY/T 3397-2024) to collect phenotypic data from all test varieties. Quantitative traits (such as flower diameter and leaf length) were evaluated using a standardized 1-9 rating system. The 1-point and 9-point scales represent extreme phenotypic traits within known cultivar populations. Their core purpose is to quantify and rank continuously variable traits (e.g., leaf length, flower count) for objective, standardized differentiation among cultivars. The standard includes 24 basic traits covering various plant organs of *Hydrangeas*, including plant type, stems, leaves, inflorescences, sterile flowers, and flowering period (Table S2). Notably, approximately half of the DUS test traits are related to floral characteristics, indicating their significant value in *Hydrangea* variety identification. Based on surveys of three experimenters, we retained 24 distinctive and representative traits from *Hydrangeas* to statistically analyze the morphology of 50 varieties, creating box-and-line charts for these traits (Fig. 2).

All 24 DUS traits showed notable variation, with the coefficient of variation (CV) ranging from 12.37% to 132.34% (mean 48.83 %) (Table 5). Analysis of the 24 phenotypic traits in all samples demonstrated significant differences in trait variation frequencies. The Simpson diversity index revealed significant phenotypic diversity in commercially important traits (flower color and inflorescence morphology) among the *Hydrangea* varieties, whereas certain stable traits (leaf abaxial surface hairs and leaf pigmentation) showed convergence. Anthocyanin pigmentation in the petiole, which is associated with flower color and shade tolerance, may serve as an indirect selection indicator in breeding. Stem pubescence was mostly observed in *H. paniculata* and *H. arborescens*. The size and degree of folding of the sterile sepals jointly determine the ornamental value of "double-petaled" *Hydrangeas*, and are central to the appeal of such varieties in the cut-flower market. These traits show strong potential for improvement.

Four traits, including plant type, leaf coats on the back, Leaf anthocyanin coloring, Petioles: color exhibited a standard deviation (SD) < 0.72 and a standard deviation of inter-traits (SDI) < 0.15. Low genetic differentiation may be attributed to the combined selection pressure for erect growth and green stems in commercial cultivars. Traits with a high SDI value typically showed a high CV, although "stem anthocyanins" was an exception, indicating an extreme distribution rather than uniform variation. These results suggest that several varieties exhibited extreme phenotypes, which may be useful for constructing specialized flower-color screening populations.

3.2. Principal component analysis of phenotypic traits

To evaluate the phenotypic similarity among the 50 cultivars, we integrated information from all quantitative and qualitative traits using Gower distance to calculate pairwise distances between the cultivars. The PCA scatterplot based on this distance matrix revealed the clustering patterns and primary differentiation of the cultivars in multidimensional space (Fig. 3a). Hierarchical cluster analysis further validated the group structure (Fig. 3b), revealing the relationships among the Gower distance-based cultivar clusters and highlighting the major branches or clusters. Hierarchical clustering using Ward's method divided the cultivars into four or five clusters. Different clusters represented groups of cultivars with similar characteristics, where cultivars within the same cluster exhibited higher genetic or phenotypic similarity, while those placed in different clusters showed lower genetic or phenotypic similarity.

To simplify the field DUS testing, the core DUS traits required screening. We performed a PCA on the 24 DUS traits across the 50

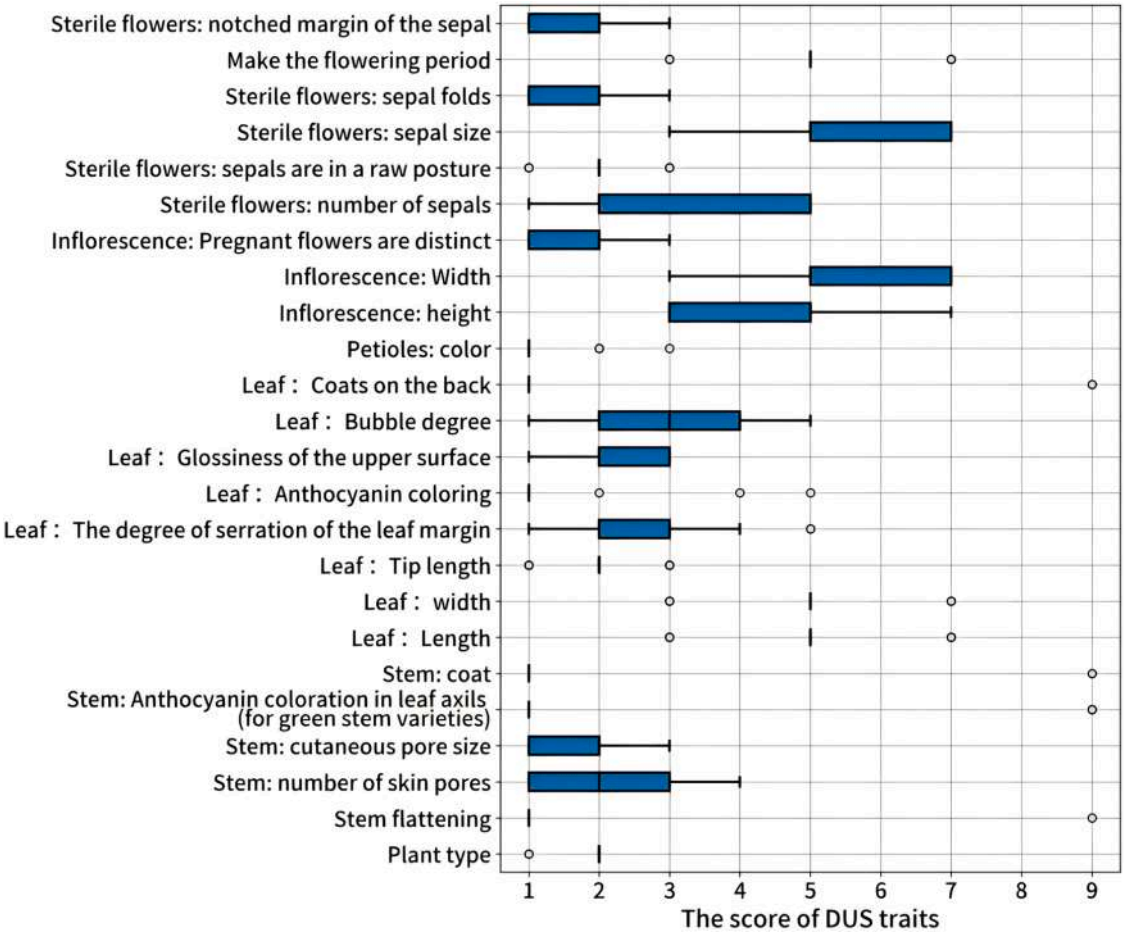

Fig. 2. Box plot/histogram illustrating the distribution pattern and range in variation for the 24 phenotypic traits in the 50 *Hydrangea* cultivars. The thick black line within each box indicates the median for each trait. The length of the box, the range of the upper and lower whiskers, and potential outliers indicate the dispersion and distribution characteristics of the data for each trait.

Table 5
Genetic diversity index for ordered and unordered phenotypic traits for the 50 *Hydrangea* cultivars.

Trait (QL/QN)		Minimum	Maximum	Standard deviation	Simpson diversity index	CV/%	GDI
Orderly shapes	Stem: number of skin pores	1.00	4.00	1.03	0.72	48.69	0.74
	Stem: cutaneous pore size	1.00	3.00	0.84	0.59	49.72	0.49
	Leaf: Length	3.00	7.00	1.23	0.54	24.45	0.67
	Leaf: width	3.00	7.00	1.20	0.52	23.57	0.63
	Leaf: Tip length	1.00	3.00	0.69	0.61	34.64	0.43
	Leaf: The degree of serration of the leaf margin	1.00	5.00	1.13	0.73	39.92	0.83
	Leaf: Anthocyanin coloring	1.00	5.00	0.70	0.12	60.62	0.08
	Leaf: Glossiness of the upper surface	1.00	3.00	0.69	0.61	30.44	0.42
	Leaf: Bubble degree	1.00	5.00	1.26	0.77	44.68	0.97
	Inflorescence: height	3.00	7.00	1.13	0.52	25.54	0.59
	Inflorescence: Width	3.00	7.00	1.23	0.56	21.79	0.69
	Inflorescence: Pregnant flowers are distinct	1.00	3.00	0.78	0.53	50.65	0.41
	Sterile flowers: number of sepals	1.00	5.00	1.47	0.57	38.08	0.84
	Sterile flowers: sepal size	3.00	7.00	1.39	0.61	24.96	0.85
	Make the flowering period	3.00	7.00	1.06	0.44	20.77	0.47
Disordered traits	Plant type	1.00	2.00	0.24	0.11	12.24	0.03
	Stem flattening	1.00	9.00	1.12	0.04	96.55	0.04
	Stem: Anthocyanin coloration in leaf axils (for green stem varieties)	1.00	9.00	3.31	0.34	120.07	1.14
	Stem: coat	1.00	9.00	2.17	0.15	132.34	0.32
	Leaf: Coats on the back	1.00	9.00	1.90	0.11	128.37	0.21
	Petioles: color	1.00	3.00	0.36	0.15	32.78	0.05
	Sterile flowers: sepals are in a raw posture	1.00	3.00	0.48	0.39	23.24	0.19
	Sterile flowers: sepal folds	1.00	3.00	0.52	0.47	38.04	0.25
	Sterile flowers: notched margin of the sepal	1.00	3.00	0.73	0.49	49.79	0.35

cultivars (Fig. 3c and d). Higher principal component scores indicated greater contributions of these traits to morphological variation for

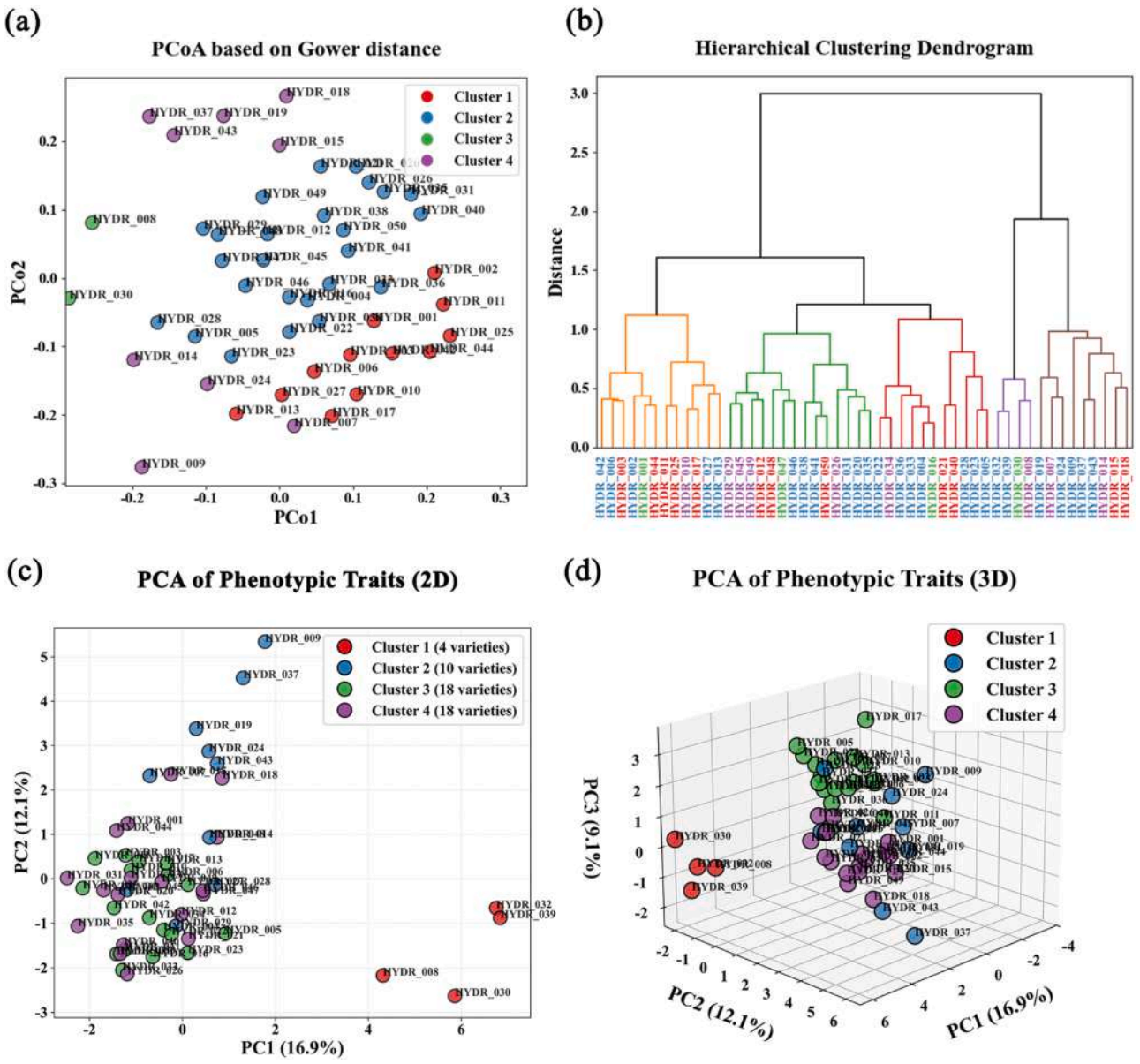

Fig. 3. Principal coordinate analysis (PCoA), principal component analysis (PCA), and hierarchical cluster analysis using Ward's method based on Gower distances among the 50 *Hydrangea* cultivars. In each analysis, a matrix of Gower distances between each variety based on 24 phenotypic traits was analyzed. (a) PCoA scatterplot visualizing the multivariate morphological differences among the cultivars. Each point represents a unique cultivar and colors correspond to the cluster membership in the hierarchical clustering dendrogram in (b). The spatial separation of the varieties on axes PCo1 and PCo2 reflects the degree of phenotypic dissimilarity. (b) Hierarchical clustering dendrogram representing the phenotypic similarity of the varieties based on the 24 phenotypic traits. The vertical axis represents the Euclidean distance at which clusters are merged. Colored branches and labels correspond to the distinct clusters identified, and are consistent with the PCoA groups shown in (a). (c) and (d) Two- and three-dimensional PCA ordinations of the 50 *Hydrangea* varieties based on variation in the 24 phenotypic traits.

cultivar identification. Ultimately, 11 phenotypic traits for identification were selected as the core set of traits for *Hydrangea* DUS testing (Table 6 and Supplementary Table S4). Collectively, these traits accounted for 81.97% of the total morphological variation. Leaf morphology proved to be the most critical indicator for identification, with five leaf-related traits comprising 50% of the core traits. Notably, leaf bullation degree showed the highest contribution. The results indicate that leaf morphological variation serves as the primary basis for distinguishing cultivars and is a preferred criterion for variety identification. Furthermore, floral organs—particularly the number of sterile flowers sepals and notched margin of the sepal—were crucial for variety classification. These floral organ characteristics are core traits in variety identification, providing distinctive diagnostic traits observable during the flowering period.

Table 6
Top-ranked core phenotypic traits and their respective contribution scores calculated using the weighted loading method for the 24 phenotypic traits.

Trait	PCA score
Leaf: Bubble degree	18.26
Petioles: color	18.17
Leaf: The degree of serration of the leaf margin	17.81
Plant type	17.61
Leaf: Tip length	17.43
Leaf: width	17.20
Sterile flowers: number of sepals	17.10
Stem: cutaneous pore size	17.02
Inflorescence: height	17.00
Sterile flowers: notched margin of the sepal	16.87
Leaf: Glossiness of the upper surface	16.62

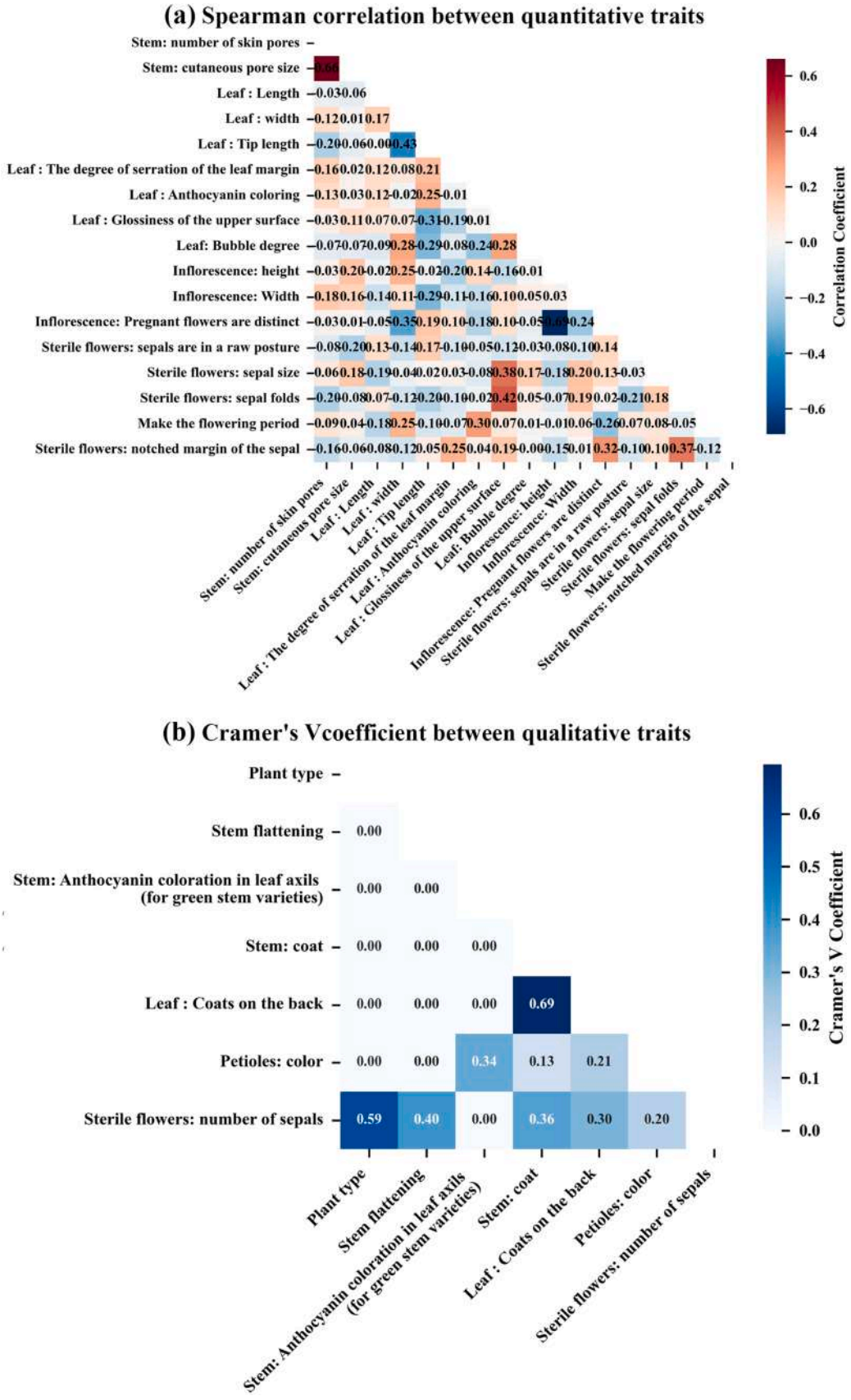

Fig. 4. Correlation analysis of 24 phenotypic traits in 50 *Hydrangea* varieties. (a) Heatmap of Spearman rank correlation coefficients for quantitative traits. (b) Heatmap of Cramer's *V* coefficients for qualitative traits. (c) Kruskal–Wallis test results showing the significance of differences in the distribution of multiple quantitative traits against qualitative traits. The color shade and numerical values in the boxes reflect the degree of association among the traits. Darker colors and higher values indicate a stronger association, whereas smaller values suggest weaker associations.

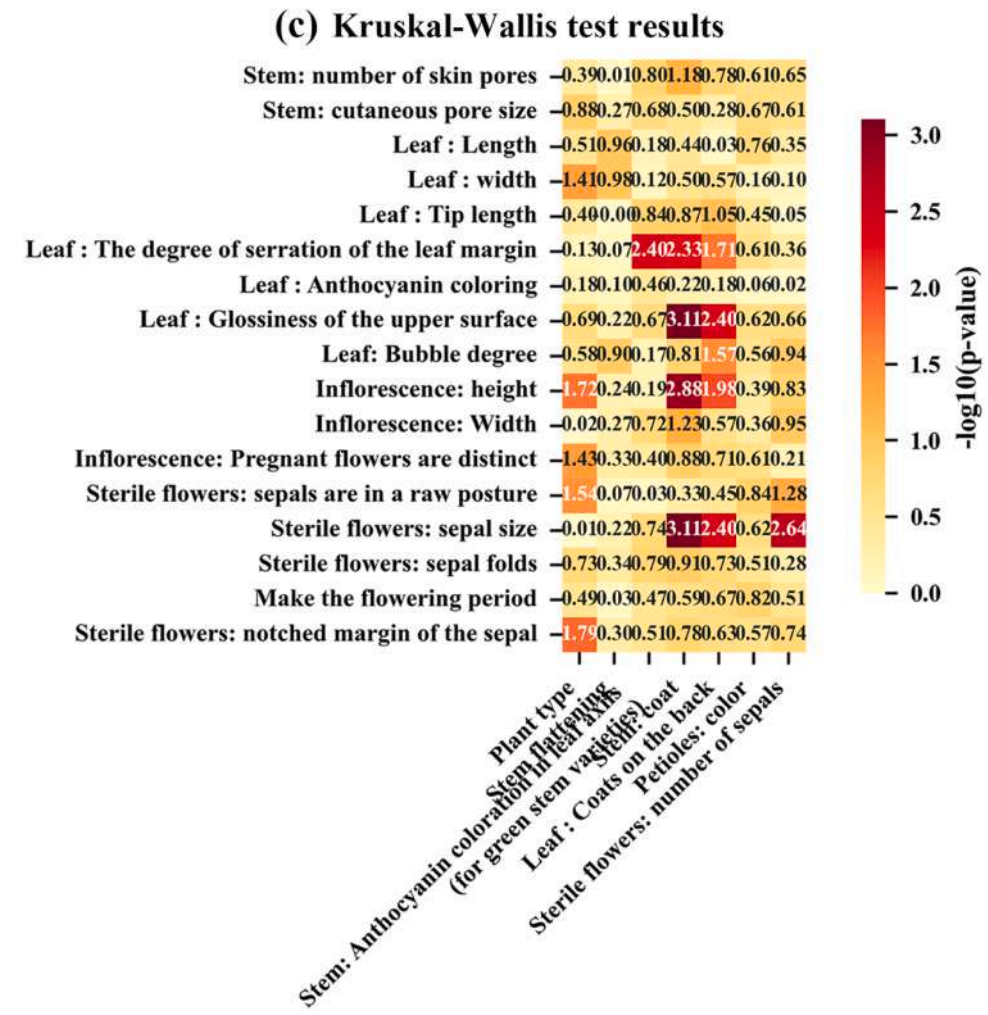

Fig. 4. (*continued*).

3.3. Correlation analysis of phenotypic traits

To further investigate the interrelationships among the phenotypic traits, a systematic analysis of the correlations among the 24 phenotypic traits was performed. A Spearman rank correlation analysis revealed significant correlations among seven qualitative traits (Fig. 4a). Leaf micromorphology and inflorescence size formed core axes of phenotypic variation. Correlated quantitative traits showed a continuous broad distribution and predominantly moderate-to-high positive correlations, consistent with polygenic quantitative inheritance patterns. Notably, leaf length–leaf width and inflorescence height–inflorescence width exhibited significant positive correlations . Leaf margin serration showed a moderate positive correlation with calyx size. The absolute correlation coefficients for leaf anthocyanin-related quantitative traits were all < 0.20, indicating potential for independent color improvement. Cramer's *V* coefficient analysis (Fig. 4b) was used to assess correlations among seven categorical qualitative traits, revealing that plant type and stem flattening were non-correlated with most of the traits. The Kruskal–Wallis test revealed significant correlations of qualitative traits with quantitative traits (Fig. 4c), suggesting that genetic differentiation in qualitative traits may regulate quantitative traits. Crucially, the quantitative trait correlation heatmap (Fig. 4a) suggested that the significant correlations of leaf length–leaf width and inflorescence height–inflorescence width support their potential regulation by a shared core network, whereas weak correlations between leaf anthocyanin-related traits and most of the other quantitative traits suggest the independence of color improvement without compromising plant-type construction or inflorescence structure.

In conclusion, leaf morphology and inflorescence-related traits, together with stem-hair-related traits, jointly determine phenotypic variation among *Hydrangea* cultivars, collectively explaining most of the phenotypic variation. These findings provide theoretical support for subsequent molecular marker development and core germplasm screening in *Hydrangea*, while suggesting practical priority trait combinations useful for multitrait collaborative improvement strategies.

3.4. Comparison of DNA barcode efficiency

Interspecific genetic distance serves to differentiate distinct sections, while intraspecific genetic distance reflects the degree of genetic variation within the same section. We amplified three chloroplast DNA fragments (Table 7). Statistical analyses using Mann–Whitney *U* tests (for pairwise comparisons) and Kruskal–Wallis tests (for multigroup

Table 7
PCR statistics for the three chloroplast DNA fragments amplified.

	*Rbc*L	*Mat*K	*psb*A-*trn*H
Success rates of PCR (%)	100	100	100
Success rates of sequencing (%)	100	100	100
Sequence lengths (bps	713-843	745-795	402-469
%Pairwise Identity	98.1%	99.6%	98.8%
%Identical Sites	31.8%	91.4%	88.6%
%GC	42.9%	35.5%	31.0%

comparisons) revealed that these fragments could serve as effective species demarcation markers to some extent (Fig. 5 and Table 8). The *rbc*L fragment, with its broadest range of genetic distance distribution (0–4.03), serves as a key molecular marker for distinguishing interspecific cultivars of hydrangea. The *mat*K fragment showed the smallest range in distances with extremely narrow interspecific differences and low resolution; its discriminatory capability is comparatively stronger at the intraspecific level than its efficiency in interspecific

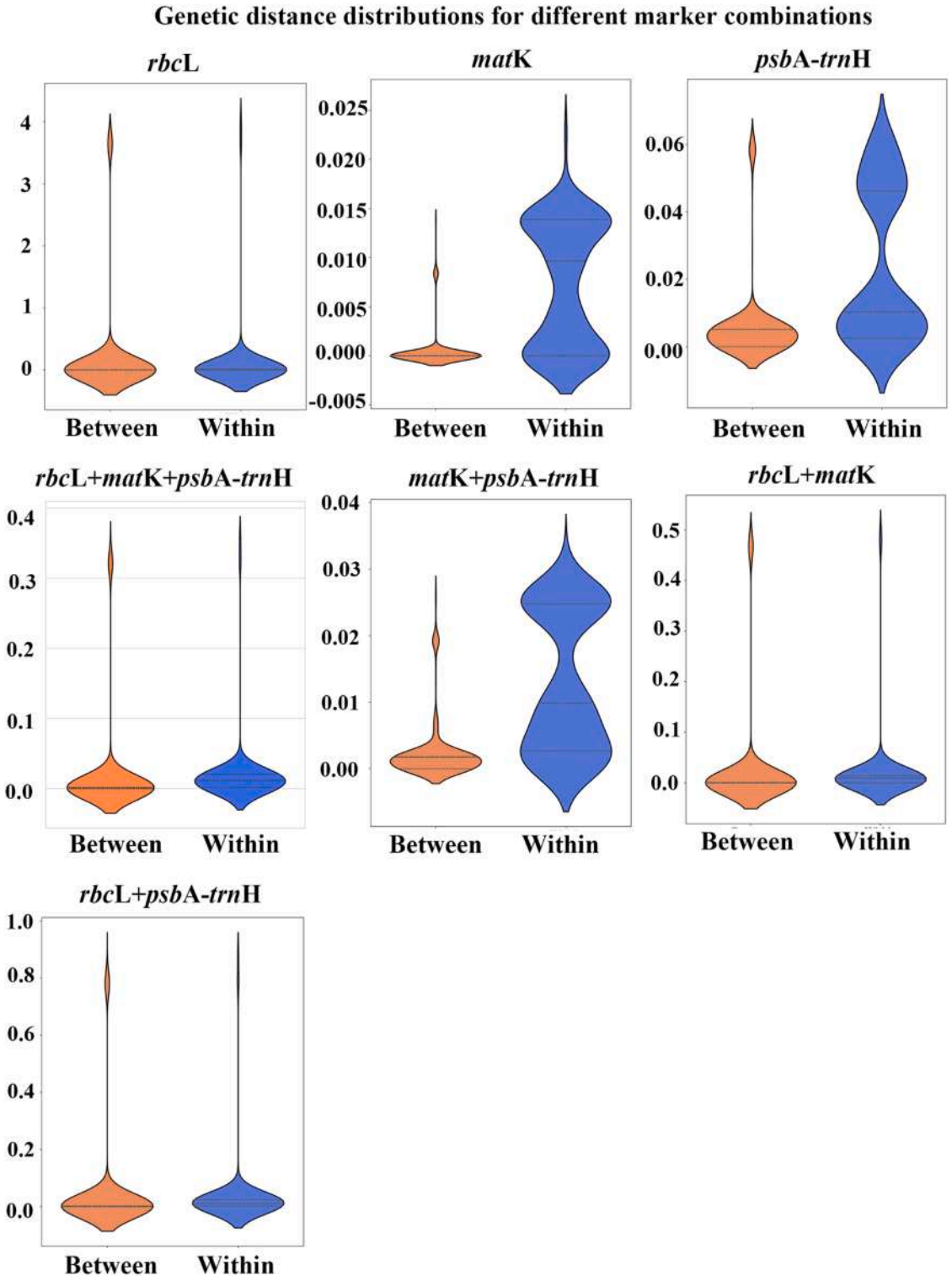

Fig. 5. Violin plots for seven combinations of three chloroplast DNA fragments depicting the distribution of interspecific and intraspecific genetic distances in 50 *Hydrangea* varieties. Violin plots display the genetic distance distributions between section (Between, orange) and within cultivars (Within, blue) for seven chloroplast DNA fragment combinations. Genetic distances were calculated based on sequence alignments and are expressed as standardized substitutions per 100 sites. The width of each violin represents the kernel density estimate of the distribution; the inner white line indicates the median, and the box span shows the interquartile range (IQR, 25th to 75th percentile). Interspecific distances were calculated from pairwise combinations of cultivars across different sections, while intraspecific distances were based on pairwise comparisons among individuals within cultivars of the same section. The vertical axis represents gene distance.

Table 8
Mann–Whitney *U* test (two-way comparisons) or Kruskal–Wallis test (multi-group comparisons) results to evaluate statistical differences in the intraspecific and interspecific variation in sequences for three chloroplast DNA fragments and different combinations of the fragments.

Fragment	Type	Max-Genetic Distance	
		Range	Max
rbcl+matk	Within	0.495564885	0.495564885
rbcl+matk	Between	0.483043974	0.483043974
psbA-trnH	Within	0.061173312	0.061173312
psbA-trnH	Between	0.061173312	0.061173312
rbcl	Within	4.033912435	4.033912435
rbcl	Between	3.726726727	3.726726727
rbcl+psbA-trnH	Within	0.885585905	0.885585905
rbcl+psbA-trnH	Between	0.87402428	0.87402428
matk+psbA-trnH	Within	0.031815498	0.031815498
matk+psbA-trnH	Between	0.026746889	0.026746889
matk	Within	0.022738386	0.022738386
matk	Between	0.013884193	0.013884193
rbcl+matk+psbA-trnH	Within	0.357959484	0.357959484
rbcl+matk+psbA-trnH	Between	0.345742981	0.345742981

Note: Statistical comparisons were performed using the Mann–Whitney *U* test (between vs. within groups) and the Kruskal–Wallis test (multiple fragment comparisons). The significance level was $P < 0.01$.

differentiation.. The *psb*A–*trn*H and *mat*K+*psb*A–*trn*H fragments provided intermediate resolution, where signals of intraspecific differentiation were more pronounced than those for interspecific differentiation, resulting in limited overall discriminatory efficiency. Overall, combined fragments maintained significant differentiation while reducing the extreme variation. Kruskal–Wallis comparisons of interfragment differences revealed that different fragments exhibited distinct interspecific distance distributions.

3.5. Systematic development and clustering analysis

Comprehensive analysis of multiple DNA fragment combinations revealed patterns in the genetic structure of the *Hydrangea* varieties. The variation in PC1 is primarily driven by species-specific 3-mer signatures, reflecting the core genetic differentiation patterns among different *Hydrangea* species. The variation in PC2 is associated with 3-mer combinations linked to intra-specific polymorphic sites, which can be used to analyze the genetic diversity and population differentiation patterns within species. PCA ordinations of the seven DNA fragment combinations (Fig. 6 and Table 9) showed that the markers collectively explained 70.53%–96.25% of the total genetic variation within the dataset. The individual *mat*K fragment had the highest cumulative explanatory power (96.25%), demonstrating its pivotal role in species differentiation. In contrast, the composite marker *rbc*L+*mat*K+*psb*A-*trn*H exhibited a significantly lower cumulative variance of only 70.53%. This phenomenon suggests that combining conservative loci may dilute the signal strength, thereby compromising the accuracy of species identification. The study also found that *rbc*L (cumulative explanatory power: 89.93%) and *psb*A-*trn*H (cumulative explanatory power: 90.72%) exhibited high variance explained when used individually. In *Hydrangea* germplasm studies, using *mat*K alone will provide optimal results. When balancing resolution between PC1 and additional dimensions, the *mat*K+*psb*A–*trn*H combination proved to be advantageous, enabling a comprehensive evaluation of both species differentiation and intraspecific variation. This approach provides a valuable framework for classifying and conserving *Hydrangea* germplasm resources and offers practical guidance for future genetic diversity assessments. The findings further indicated that transitioning from "multi-fragment" approaches to "fragment selection based on variance and sample selection based on coordinates" better offers an improved balance betweens cost-effectiveness and genetic coverage in germplasm resource research.

To further investigate the developmental relationships among the 50 *Hydrangea* varieties, we analyzed statistical differences in intraspecific and interspecific variation in the three DNA fragments using ML analysis with the three-marker and *mat*k+*psb*A–*trn*H dual-marker combinations. The 89 samples were classified into four distinct groups. Notably, the 50 *Hydrangea* varieties were resolved into four clades (Fig. 7), effectively distinguishing *H. arborescens* and *H. paniculata* from *H. macrophylla*, and grouping the *H. macrophylla* accessions into three major subclades. The hybrid varieties 'Runaway Bride' and 'French Bolero', derived from crosses between *H. macrophylla* and *H. serrata*, were isolated, further indicating that the three sequence groups enable identification of genetic relationships among *Hydrangea* varieties.

3.6. Phenotype–molecular marker association analysis

Based on the Mantel test and Wilcoxon rank correlation results, the association between genetic and phenotypic distances across the 50 Hydrangea varieties was evaluated for seven DNA fragment combinations and 24 phenotypic traits (Fig. 8 and Table 10). The fragments matK and psbA–trnH, both individually and in combination (matK+psbA–trnH), showed statistically significant positive correlations in the Mantel test ($q < 0.05$) (Fig. 8a). This indicates that genetic distances derived from these loci are meaningfully associated with overall phenotypic differentiation among varieties. The strong and consistent significance of the Wilcoxon rank correlations for these fragments further supports their utility in capturing trait-associated genetic variation (Fig. 8b). Conversely, all fragment combinations containing rbcL (including rbcL alone, rbcL+matK, rbcL+psbA-trnH, and the triple combination) yielded non-significant ($P > 0.5$) or slightly negative Mantel correlations (Fig. 8c and Fig. 8d). While Wilcoxon tests for these combinations remained statistically significant, the exceedingly weak or negative Mantel r values suggest that the genetic variation captured by rbcL has a negligible or inconsistent relationship with the multivariate phenotypic distance structure among these varieties. This finding underscores the limited utility of rbcL for predicting phenotypic divergence in this context, despite its value as a conservative phylogenetic marker.

3.7. Genetic Structure Revealed by Haplotype Diversity and Its Phenotypic Association

In this study, we conducted a comprehensive analysis of genetic diversity and phenotypic variation in 50 *Hydrangea* cultivars. By integrating DNA barcoding data (*rbc*L, *mat*K, and *psb*A–*trn*H sequences) with detailed phenotypic data (24 DUS traits), we successfully identified haplotypes and SNPs (Tables S5 and S6), enabling exploration of potential associations between genetic composition and phenotypic characteristics. The haplotype analysis revealed 13 haplotypes across the three sequences, with H1 being the most prevalent (frequency 17), comprising three dominant haplotypes (frequency > 5%) and 10 unique haplotypes (Fig. 9a). These results indicate that the three DNA fragments show significant genetic diversity among the tested varieties, comprising numerous distinctive genetic markers for variety identification. The haplotype diagrams provide crucial data and a theoretical foundation for subsequent genetic feature analysis, functional site validation, and molecular marker-assisted breeding for target phenotypes.

In addition, 65 SNP loci were identified in the three DNA fragment sequences, with a density of 3.39% and a transition/neutral mutation ratio of 1.00 (Fig. 9b). These findings suggest a balanced distribution of genetic variation types, with transition and transversion events occurring at comparable frequencies. In the SNP–phenotype association analysis, a total of 246 significantly associated loci (FDR < 0.05) were identified, with core associations concentrated in the stem and leaf phenotypes. Specifically, SNP_1541 and SNP_1543 showed significant negative correlations with stem lenticel number (beta = −0.402, p = 0.000e+00). The SNPs 273, 544, 804, 1147, and 1535 showed strong associations with stem characteristics (beta values of 6.400, 6.400, 6.400, 6.400, and 5.826, all p = 0.000e+0). The SNPs 273, 544, and 804

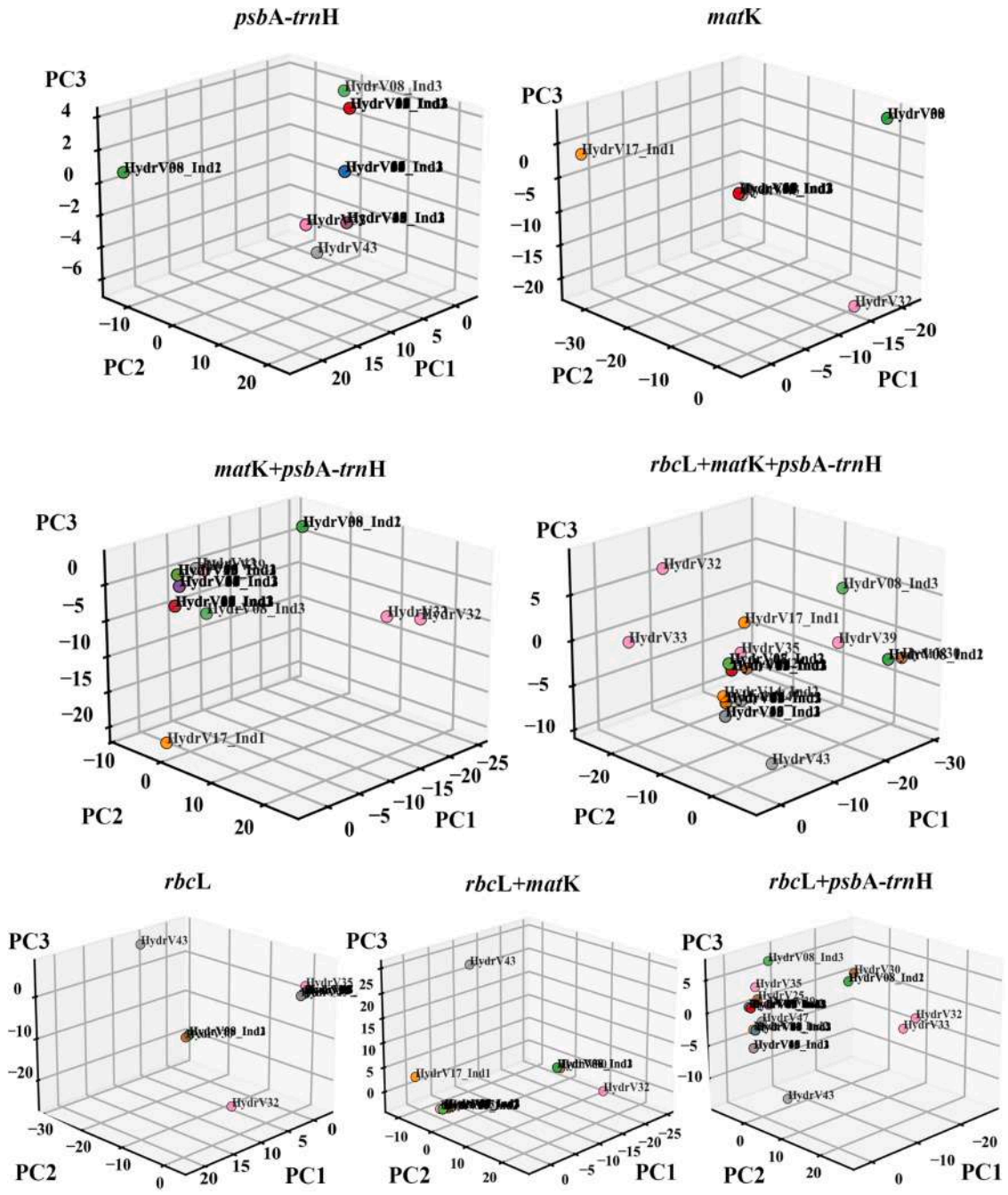

Fig. 6. Three-dimensional principal component analysis ordinations of the seven DNA fragment combinations.

Table 9
Fragment statistics and percentages of the total variance explained in a principal component analysis of seven DNA fragment combinations.

Fragment	Sequence_Length	PC1_Var	PC2_Var	PC3_Var	Total_Var
rbcl_matk	1523	58.61	13.77	11.49	83.87
rbcl_PsbA-trnH	1183	38.70	23.81	9.85	72.36
matk_PsbA-trnH	1128	41.39	21.94	11.37	74.7
rbcl_matk_PsbA-trnH	1917	41.91	19.90	8.72	70.53
rbcl	789	54.5	19.9	15.53	89.93
matk	734	61.98	22.23	12.04	96.25
PsbA-trnH	394	49.63	28.74	12.53	90.72

were significantly correlated with leaf tip length (beta = 0.444, p = 0.000e+00) (Fig. 9c and d).

4. Discussion

This study innovatively and systematically integrates the traditional

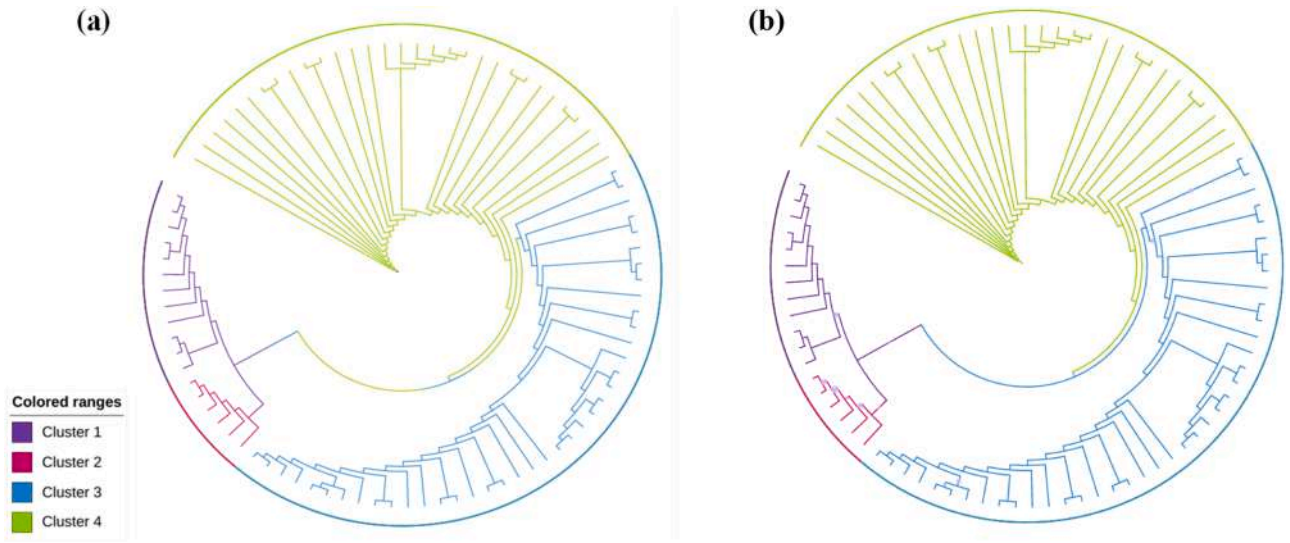

Fig. 7. Bayesian inference and maximum likelihood phylogenetic trees constructed using the triple-marker (*rbc*L+*mat*K+*psb*A–*trn*H) and the dual-marker *mat*K+*psb*A–*trn*H DNA fragment combinations.

Distinctness, Uniformity, and Stability (DUS) phenotypic test with DNA barcoding (*mat*K, *rbc*L, *psb*A-*trn*H) for germplasm identification and genetic diversity assessment. In large-scale DNA barcoding studies of biodiversity hotspots, *mat*K has been demonstrated to provide high resolution at the species level and has proven effective in resolving phylogenetic relationships among closely related taxa (Yan et al., 2015; Yong et al., 2024). This indicates that although rbcL serves as a stable backbone for deep-level phylogenetic inference due to its conserved and alignable sequence (Kress and Erickson, 2007). Meanwhile, the non-coding psbA-trnH intergenic spacer, recognized for its high mutation rate and utility in resolving low-level phylogenetic relationships (Shaw et al., 2014), has been widely adopted in plant identification and resource studies. In *Hydrangea*, chloroplast DNA barcodes (e.g., *rbc*L, *mat*K, *psb*A-*trn*H), known for their sequence stability and significant variation, are progressively being used for cultivar identification and diversity evaluation (Kress, 2017; Li et al., 2021).

In this study, the phenotypic and genetic diversity of 50 *Hydrangea* cultivars were systematically evaluated and 11 core DUS traits were identified. Principal component analysis (PCA), a proven tool for dimensionality reduction in plant phenomics (Greenacre et al., 2022), effectively distilled the complex morphology (Ahlinder et al., 2024). Traits related to leaf morphology, floral organ characteristics (especially sepal number and margin indentation of sterile flowers), and stem pubescence emerged as major contributors to phenotypic variance. This finding aligns with and quantifies the general observation that floral and foliar traits are pivotal for ornamental plant classification (Khaleghi et al., 2025). Further diversity analysis indicated that floral organ traits (e.g., flower color and inflorescence morphology) exhibited the most pronounced variation, consistent with the general pattern that these traits are key diagnostic features in ornamental plants (Zhang et al., 2022). Its molecular basis may be related to the differential expression of key genes in the anthocyanin biosynthesis pathway, and its weak correlation with flower color supports the notion that this trait can be improved independently (Gould, 2004; Schreiber et al., 2011). Regarding stem pubescence structures in *Hydrangeas*, they are predominantly found in *Hydrangea paniculata* and *Hydrangea arborescens*. Previous studies have demonstrated that *Hydrangea* paniculata exhibits superior tolerance and adaptability to abiotic stress (Liu et al., 2025), which may be associated with its unique phenotypic traits or specific gene expression patterns. This study revealed complex correlation patterns among traits. Size-related traits of leaves and inflorescences showed significant positive correlations, supporting the hypothesis of shared developmental regulatory modules (Diggle, 2014; Pigliucci, 2003).

However, phenotypic traits are susceptible to environmental influences, making it difficult for morphological data alone to fully reflect the true genetic relationships among germplasm accessions (Kovuri et al., 2023). Therefore, phylogenetic relationships reconstructed through molecular markers and DNA sequence analysis hold particular importance for germplasm identification, which is also applicable to *Hydrangea* species (Amiteye, 2021). Due to its lack of recombination, moderate evolutionary rate, and high genetic stability, haplotype analysis has become a core tool for deciphering plant population genetic structure, particularly for identifying the wild progenitors of cultivated varieties and elucidating domestication events (Peng et al., 2024). The discriminatory power of DNA barcodes varies considerably across plant lineages. For example, in ferns, the plastid combination "*mat*K + *rbc*L" has been shown to provide strong resolution for species identification, whereas the psbA-trnH intergenic region often performs poorly due to technical limitations such as low primer universality (Ebihara et al., 2010). In contrast, for the orchid genus Cymbidium, the nuclear ribosomal ITS region demonstrated a significantly higher single-locus identification success rate than the plastid markers *mat*K, *rbc*L, and *psb*A-*trn*H (Chen et al., 2024). However, the technical success rate and sequence quality of ITS can vary considerably across different taxa, potentially increasing experimental cost and uncertainty. This study favors selecting the most suitable barcode or combination that balances high success rate and strong discriminatory power. In this study, "rbcL" showed limited ability to distinguish closely related cultivars of *Hydrangea*, while the high explanatory power of individual markers such as "*mat*K" (cumulative variance of 96.25% in principal component analysis) further highlights its exceptional utility for species delimitation within *Hydrangea*. The performance of psbA-trnH in capturing intraspecific variation (explaining 90.72% of the variance) corroborates its established role in resolving phylogeny at the population level. Therefore, the high discriminatory power of *mat*K at the species level and the high sensitivity of *psb*A-*trn*H at the population level may collectively reflect the unique evolutionary trajectory and complex lineage history experienced by the genus *Hydrangea* during its domestication and dispersal.

DNA barcoding provides stable, quantifiable genotypic data, effectively overcoming the drawbacks of traditional methods. This directly enhances the efficiency and reliability of DUS (Distinctness, Uniformity, and Stability) testing, thereby shortening the testing timeline and

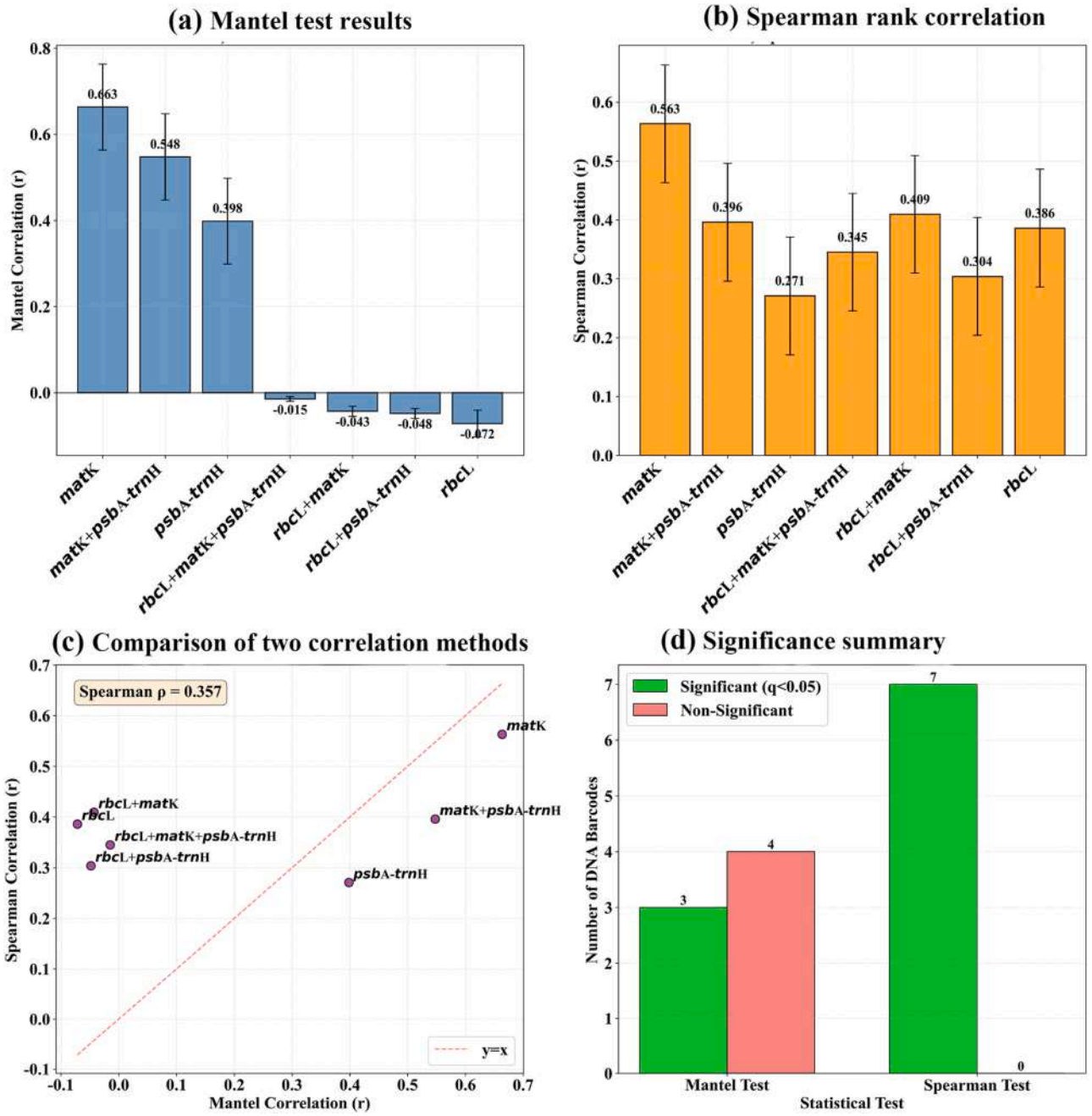

Fig. 8. Comprehensive analysis of genetic-phenotypic correlations across seven DNA barcode fragments. (a) Mantel test correlation coefficients with error bars representing 1 - q-value (confidence level). (b) Spearman rank correlation coefficients with similar error bars. (c) Scatter plot comparing Mantel and Spearman correlation coefficients for each fragment, with the diagonal line (y = x) indicating perfect agreement between methods. The overall Spearman correlation (ρ) between the two methods is displayed in the upper-left corner. (d) Bar chart summarizing the number of significant (q < 0.05) and non-significant fragments for each statistical method. Significance was determined after FDR correction using the Benjamini-Hochberg procedure. All analyses were based on pairwise comparisons between genetic distance matrices (F_{ST}-based) and phenotypic distance matrices (Euclidean distance) across 7 DNA barcode fragments and their concatenations.

reducing the uncertainty associated with phenotypic assessments alone. Furthermore, it enables high-throughput and automated analysis, allowing for more efficient handling of the increasing task of cultivar comparison. This scalability is crucial for modern breeding programs and for processing large volumes of applications in plant variety protection systems. The consistency between clusters formed by core phenotypic traits and genetic distances based on *matK/psb*A-*trn*H validates the method's reliability. From a practical breeding standpoint, our findings translate into actionable strategies. In this study, 246 SNP loci significantly associated with target traits (FDR < 0.05) were identified, providing potential marker resources for subsequent marker-assisted selection (MAS). For instance, SNP_1541 and SNP_1543 showed a significant negative correlation with stem pubescence density (β = -0.402), indicating that their genomic regions may be involved in regulating this trait and pointing to potential candidate intervals for the genetic improvement of stem hair traits in *Hydrangea*. This approach aligns with the strategy of using SNP-phenotype association analysis to guide minor allele frequency (MAS) (Majumdar et al., 2023). It is important to note that the associated SNPs identified based on statistical significance may not themselves be functional genes but could be in linkage disequilibrium with the true causal genes controlling the traits (Zhu and Zhou, 2020). When constructing a core collection oriented towards phenotypic

Table 10
Statistical differences between the fragments and the phenotype were evaluated using the Mann-Whitney U test (two-way comparison) or Kruskal-Wallis test (multi-group comparison).

Matrix number	Fragment	Mantel_r	Mantel_p	Wilcoxon_r	Wilcoxon_p
G1	*rbc*L	-0.0717	0.5325	0.3859	8.6073E-45**
G2	*mat*K	0.6634	0.0001**	0.5632	1.9371E-103**
G3	*mat*K+*psb*A-*trn*H	0.5476	0.0001**	0.3960	2.8097E-47**
G4	*rbc*L+*mat*K+*psb*A-*trn*H	-0.0148	0.9447	0.3451	1.36318E-35**
G5	*rbc*L+*mat*K	-0.0431	0.8300	0.4095	1.02507E-50**
G6	*rbc*L+*psb*A-*trn*H	-0.0483	0.7882	0.3039	1.38075E-27**
G7	*psb*A-*trn*H	0.3984	0.0022**	0.2708	4.89582E-22**

r Correlation coefficient;
** : Regression correlation is significant at p≤0.01 levels, respectively

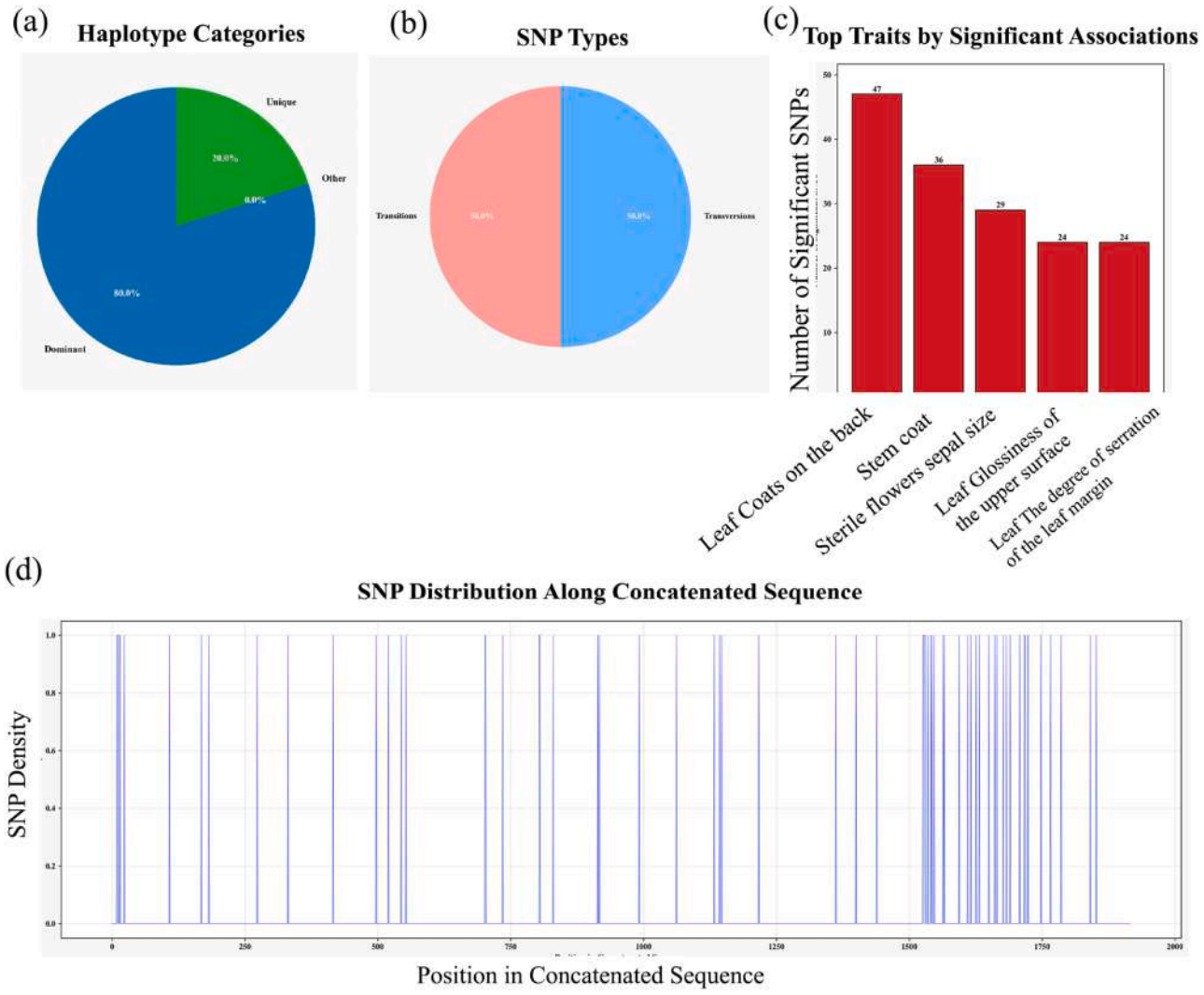

Fig. 9. Integration of DNA barcoding data (*rbc*L, *mat*K, and *psb*A–*trn*H) with detailed phenotypic data (24 DUS traits) to identify haplotypes and single nucleotide polymorphisms (SNPs). (a) Haplotype categories, (b) SNP types, (c) significant SNP–phenotype associations, and (d) SNP distribution along the concatenated sequence.

diversity and breeding applications, priority should be given to molecular markers closely associated with target traits (Brown, 1989; Singh et al., 2024). Given the high discriminative power and phenotypic correlation of the *mat*K and *psb*A-*trn*H barcodes in this study, future research could integrate SNP and DNA barcode data to establish a more efficient and accurate evaluation and breeding system for *Hydrangea* germplasm, thereby promoting the conservation and utilization of its genetic resources.

5. Conclusions

This study presents linkage models between three chloroplast DNA fragments (*mat*K, *rbc*L, *psb*A–*trn*H) and the UPOV DUS testing guidelines for *Hydrangea*. Comparative analyses using the Mann–Whitney *U* test (for two-way comparisons) and the Kruskal–Wallis test (for multigroup comparisons) revealed that *mat*K and *psb*A–*trn*H demonstrated higher discriminative efficiency. Notably, *mat*K exhibited superior discriminative potential compared with *rbc*L, *psb*A–*trn*H, and their combinations,

making it a superior specific DNA barcode marker. Co-analysis of the three DNA fragments identified 13 haplotypes of which H1 was the most prevalent. The dataset comprised three dominant haplotypes and 10 unique haplotypes, revealing genetic diversity among the tested varieties and the presence of distinct genetic markers for variety differentiation. The present evaluation of genetic diversity and phenotypic variation patterns among 50 *Hydrangea* varieties provides crucial insights for *Hydrangea* systematics, germplasm identification, breeding programs, and the formulation of conservation strategies. In conclusion, this study has established a DUS identification system for *Hydrangea* centered on leaf morphology and key floral organ characteristics, elucidated the variation patterns and genetic basis of major traits, and provided a solid phenotypic and theoretical foundation for subsequent core germplasm construction, variety identification, and targeted breeding strategy development.

Data availability

The DNA sequence data of this project has been archived in the GenBank: BankIt2999009: PX280288 - PX280376; BankIt3000343: PX280377 - PX280464; BankIt3001087: PX280465 - PX280553; BankIt3002041 Seq86 PX310743. The raw data of this article will be made available by corresponding author according to the personal requests.

CRediT authorship contribution statement

Yufei Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation. **Bo Jiang:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Jiayue Zhang:** Validation, Software, Investigation, Formal analysis, Data curation. **Beier Zhou:** Software, Investigation, Data curation. **Biao Li:** Resources. **Yue Wu:** Investigation. **Yichen Li:** Resources. **Ye Jiang:** Validation, Formal analysis. **Huanghui Lu:** Resources. **Erxu Pi:** Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Foreign Academician Workstation of Suzhou city (SWY2020001), the Changshu city Science and Technology Development Plan (CN202410, CN202505) and the Gusu Local Talent Development Project (2022), the Zhejiang Provincial Science and Technology Department (grant no. 2025SDXT004-2), the Hangzhou Joint Fund of the Zhejiang Provincial Natural Science Foundation of China (grant no. LHZSZ24C010001)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scienta.2026.114702.

References

Ahlinder, J., Hall, D., Suontama, M., Sillanpää, M.J., 2024. Principal component analysis revisited: fast multitrait genetic evaluations with smooth convergence. G3 Genes, Genomes, Genet. 14 (12), jkae228. https://doi.org/10.1093/g3journal/jkae228.

Amiteye, S., 2021. Basic concepts and methodologies of DNA marker systems in plant molecular breeding. Heliyon 7 (10), e08093. https://doi.org/10.1016/j.heliyon.2021.e08093.

Andrews, S., 2014. FastQC A Quality Control tool for High Throughput Sequence Data (Version 0.11.2)[Software]. Babraham Institute. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Antil, S., Abraham, J.S., Sripoorna, S., Maurya, S., Dagar, J., Makhija, S., Bhagat, P., Gupta, R., Sood, U., Lal, R., Toteja, R., 2023. DNA barcoding, an effective tool for species identification: a review. Mol. Biol. Rep. 50 (1), 761–775. https://doi.org/10.1007/s11033-022-08015-7.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brown, A., 1989. Core collections: A practical approach to genetic resource management. Genome 31 (2), 818–824. https://doi.org/10.1139/g89-144.

Chac, L.D., Thinh, B.B., 2023. Species identification through DNA barcoding and its applications: A review. Biol. Bull. 50, 1143–1156. https://doi.org/10.1134/S106235902360229X.

Chen, Y., Chen, J., Zhong, S., Zhong, H., Liu, Z., 2023. DUS traits and classification of *antirrhinum majus* L. Fujian J. Agric. Sci 38 (8), 944–952. https://doi.org/10.19303/j.issn.1008-0384.2023.08.008.

Chen, Z., Gao, L., Wang, H., Feng, S., 2024. Molecular identification and phylogenetic analysis of Cymbidium species (Orchidaceae) based on the potential DNA barcodes *matK, rbcL, psbA-trnH*, and internal transcribed spacer. Agronomy 14 (5), 933. https://doi.org/10.3390/agronomy14050933.

Clegg, M.T., 1993. Chloroplast gene sequences and the study of plant evolution. Proc. Natl. Acad. Sci. U.S.A. 90 (2), 363–367. https://doi.org/10.1073/pnas.90.2.363.

Dabbu, M., Karuppusamy, L., Pulugu, D., Vootla, S.R., Reddyvari, V.R., 2022. Water atom search algorithm-based deep recurrent neural network for the big data classification based on spark architecture. Int. J. Mach. Learn. Cyber. 13, 2297–2312. https://doi.org/10.1007/s13042-022-01524-8.

Diggle, P.K., 2014. Modularity and intra-floral integration in metameric organisms: plants are more than the sum of their parts. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369, 20130253. https://doi.org/10.1098/rstb.2013.0253.

Ebihara, A., Nitta, J.H., Ito, M., 2010. Molecular species identification with rich floristic sampling: DNA barcoding the pteridophyte flora of Japan. PLoS ONE 5, e15136. https://doi.org/10.1371/journal.pone.0015136.

Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. Nat. Rev. Genet. 17, 422–433. https://doi.org/10.1038/nrg.2016.58.

Gould, K.S., 2004. Nature's Swiss Army knife: the diverse protective roles of anthocyanins in leaves. Biomed Res. Int. 2004, 415423. https://doi.org/10.1155/S1110724304406147.

Greenacre, M., Groenen, P.J.F., Hastie, T., D'Enza, A.I., Markos, A., Tuzhilina, E., 2022. Principal component analysis. Nat. Rev. Methods Primers 2, 100. https://doi.org/10.1038/s43586-022-00184-w.

Hartvig, I., Czako, M., Kjær, E.D., Nielsen, L.R., Theilade, I., 2015. The use of DNA barcoding in identification and conservation of Rosewood (*Dalbergia* spp.). PLoS ONE 10 (9), e0138231. https://doi.org/10.1371/journal.pone.0138231.

He, Q., Yao, J., Fang, P., Qi, J., Zhang, L., 2022. DUS test and DNA fingerprinting construction of jute varieties. In: Zhang, L., Khan, H., Kole, C. (Eds.), The Jute Genome. Compendium of Plant Genomes. Springer, Cham. https://doi.org/10.1007/978-3-030-91163-8_5.

Hirota, S.K., Yahara, T., Fuse, K., Sato, H., Tagane, S., Fujii, S., Minamitani, T., Suyama, Y., 2022. Molecular phylogeny and taxonomy of the *hydrangea serrata* complex (Hydrangeaceae) in western Japan, including a new subspecies of *H. acuminata* from Yakushima. PhytoKeys 188, 49–71. https://doi.org/10.3897/phytokeys.188.64259.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. PLoS ONE 6 (5), e19254. https://doi.org/10.1371/journal.pone.0019254.

Huang, Z., Li, M., Liang, C., Fan, Q., Yu, W., Wang, S.a., Lin, M., Li, P., Zhang, R., Gong, W., 2025. Evolutionary genetics analysis for complete plastomes and diversification rate estimation of Deutzia and Philadelphus in tribe Philadelpheae (Hydrangeaceae). BMC Plant Biol 25 (1), 1104. https://doi.org/10.1186/s12870-025-07073-w.

Hufford, L., Moody, M.L., Soltis, D.E., 2001. A phylogenetic analysis of hydrangeaceae based on sequences of the plastid gene *matK* and their combination with *rbcL* and morphological data. Int. J. Plant Sci. 162, 835–846. https://doi.org/10.1086/320789.

Hunter, J.D., 2007. Matplotlib (Version 3.10.5) [software]. IEEE Computer Society. https://doi.org/10.1109/MCSE.2007.55.

IBM Corporation, 2020. IBM SPSS Statistics (Version 27) [software]. IBM Corporation. https://www.ibm.com/products/spss-statistics.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589. https://doi.org/10.1038/nmeth.4285.

Kang, Y., Deng, Z., Zang, R., Long, W., 2017. DNA barcoding analysis and phylogenetic relationships of tree species in tropical cloud forests. Sci. Rep. 7, 12564. https://doi.org/10.1038/s41598-017-13057-0.

Kapoor, S., Sood, S., Jayaswall, K., Sood, V.K., Kumar, N., Sood, T., Jayaswal, D., Sood, V., 2025. Deciphering genetic diversity and population structure of onion (*Allium cepa* L.) using agro-morphological and molecular markers. Hortic. Environ. Biotechnol. 66, 149–162. https://doi.org/10.1007/s13580-024-00644-0.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Khaleghi, A., Khadivi, A., Tunç, Y., 2025. Multivariate analysis of *Iris meda* Stapf based on phenological and morphological characteristics. PLoS ONE 20, e0336783. https://doi.org/10.1371/journal.pone.0336783.
Kovuri, P., Yadav, A., Sinha, H., 2023. Role of genetic architecture in phenotypic plasticity. Trends Genet 39 (9), 703–714. https://doi.org/10.1016/j.tig.2023.04.002.
Kress, W.J., 2017. Plant DNA barcodes: applications today and in the future. J. Syst. Evol. 55, 291–307. https://doi.org/10.1111/jse.12254.
Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE 2 (6), e508. https://doi.org/10.1371/journal.pone.0000508.
Kroon, E.E., Swart, Y., van Rensburg, R., Cuttler, K., Müller-Nedebock, A.C., Kotze, M.J., 2025. Chapter 7 - genotype versus phenotype versus environment. In: Patrinos, G.P., Möller, M., Uren, C. (Eds.), Population Genomics in the Developing World. Academic Press, pp. 107–141.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
LaPierre, N., Ju, C.J.T., Zhou, G., Wang, W., 2019. MetaPheno: A critical evaluation of deep learning and machine learning in metagenome-based disease prediction. Methods 166, 74–82. https://doi.org/10.1016/j.ymeth.2019.03.003.
Letunic, I., Bork, P., 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res 52, W78–W82. https://doi.org/10.1093/nar/gkae268.
Li, H., Xiao, W., Tong, T., Li, Y., Zhang, M., Lin, X., Zou, X., Wu, Q., Guo, X., 2021. The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants. Sci. Rep. 11, 1424. https://doi.org/10.1038/s41598-021-81087-w.
Liao, Y., Zhao, W., Wang, Y., Zhao, S., Gu, S., Zhang, C., Peng, J., Sheng, S., Wang, S., 2025. Phylogenetic and functional analysis of MYB genes unraveling its role involved in anthocyanin biosynthesis in *H. macrophylla*. Sci. Rep. 15, 29125. https://doi.org/10.1038/s41598-025-14216-4.
Liberty, J.T., Lin, H., Kucha, C., Sun, S., Alsalman, F.B., 2025. Innovative approaches to food traceability with DNA barcoding: beyond traditional labels and certifications. Ecol. Genet. Genomics 34, 100317. https://doi.org/10.1016/j.egg.2024.100317.
Liu, Y., Chen, S., Chen, W., Feng, J., Chen, H., Han, Y., Du, X., Deng, Y., 2025. Comparisons on the lead-tolerance and physiological response of different *Hydrangea paniculata* cultivars to lead stress. Environ. Technol. Innov. 40, 104376. https://doi.org/10.1016/j.eti.2025.104376.
Luo, F., Zhang, Q., Xin, H., Liu, H., Yang, H., Doblin, M.S., Bacic, A., Li, L., 2022. A phytochrome B-PIF4-MYC2/MYC4 module inhibits secondary cell wall thickening in response to shaded light. Plant Commun 3, 100416. https://doi.org/10.1016/j.xplc.2022.100416.
Majumdar, S., Basu, S., McGue, M., Chatterjee, S., 2023. Simultaneous selection of multiple important single nucleotide polymorphisms in familial genome wide association studies data. Sci. Rep. 13, 8476. https://doi.org/10.1038/s41598-023-35379-y.
McHugh, M.L., 2012. Interrater reliability: the kappa statistic. Biochem. Med. 22 (3), 276–282.
Mohanty, S., Mishra, B.K., Dasgupta, M., Acharya, G.C., Singh, S., Naresh, P., Bhue, S., Dixit, A., Sarkar, A., Sahoo, M.R., 2023. Deciphering phenotyping, DNA barcoding, and RNA secondary structure predictions in eggplant wild relatives provide insights for their future breeding strategies. Sci. Rep. 13, 13829. https://doi.org/10.1038/s41598-023-40797-z.
Muino, J.M., Ruwe, H., Qu, Y., Maschmann, S., Chen, W., Zoschke, R., Ohler, U., Kaufmann, K., Schmitz-Linneweber, C., 2024. MatK impacts differential chloroplast translation by limiting spliced tRNA-K(UUU) abundance. Plant J 119, 2737–2752. https://doi.org/10.1111/tpj.16945.
National Forestry and Grassland Administration, 2024. Guidelines for Specificity, Consistency, and Stability Testing of New Plant Varieties—Hydrangea. Chinese Standard LY/T 3397-2024. Issued 2024-03-01.
Pang, X., Liu, C., Shi, L., Liu, R., Liang, D., Li, H., Cherny, S.S., Chen, S., 2012. Utility of the *trnH-psbA* intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. PLoS ONE 7, e48833. https://doi.org/10.1371/journal.pone.0048833.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. scikit-learn (Version 1.7.1) [software]. Zenodo. https://doi.org/10.18637/jmlr.v12.pedregosa11a.
Peng, J., Xie, J., Gu, Y., Guo, H., Zhang, S., Huang, X., Luo, X., Qian, J., Liu, M., Wan, X., Chen, L., Huang, X., Zhang, F., He, F., Zhu, P., Zhong, Y., Yang, H., 2024. Assessing population genetic structure and diversity and their driving factors in *Phoebe zhennan* populations. BMC Plant Biol 24, 1091. https://doi.org/10.1186/s12870-024-05810-1.
Pigliucci, M., 2003. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Ecol. Lett. 6, 265–272. https://doi.org/10.1046/j.1461-0248.2003.00428.x.
Python Software Foundation, 2024. Python (Version 3.13) [software]. Python Software Foundation. https://www.python.org/.
Qin, M., Zhu, C.-J., Yang, J.-B., Vatanparast, M., Schley, R., Lai, Q., Zhang, D.-Y., Tu, T.-Y., Klitgård, B.B., Li, S.-J., Zhang, D.-X., 2022. Comparative analysis of complete plastid genome reveals powerful barcode regions for identifying wood of *Dalbergia odorifera* and *D. tonkinensis* (Leguminosae). J. Syst. Evol. 60, 73–84. https://doi.org/10.1111/jse.12598.
Rather, S.A., Wang, T., Liu, H., Schneider, H., 2023. Characterization of the complete chloroplast genome of *Dalbergia congesta* (Fabaceae), an endangered legume endemic to the Nilgiri Hills of Tamil Nadu. India. Funct. Integr. Genomics 23, 126. https://doi.org/10.1007/s10142-023-01047-7.
Rinehart, T., Scheffler, B., Reed, S., 2006. Genetic diversity estimates for the genus *hydrangea* and development of a molecular key based on SSR. J. Am. Soc. Hortic. Sci. 131, 787–792. https://doi.org/10.21273/JASHS.131.6.787.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302. https://doi.org/10.1093/molbev/msx248.
Schreiber, H.D., Jones, A.H., Lariviere, C.M., Mayhew, K.M., Cain, J.B., 2011. Role of aluminum in red-to-blue color changes in *Hydrangea macrophylla* sepals. Biometals 24, 1005–1015. https://doi.org/10.1007/s10534-011-9458-x.
Seabold, S., Perktold, J., 2010. StatsModels (Version 0.14.5) [software]. Zenodo. https://doi.org/10.25080/Majora-92bf1922-011.
Shaw, J., Lickey, E.B., Schilling, E.E., Small, R.L., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288. https://doi.org/10.3732/ajb.94.3.275.
Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., Morris, A.B., 2014. Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am. J. Bot. 101, 1987–2004. https://doi.org/10.3732/ajb.1400398.
Shen, F., Bai, X., Li, L., Fan, X., Song, Y., Yu, M., Cui, M., Jiang, S., Li, Z., Zhao, J., Shi, S., 2023. DBALM: A novel method for identifying ornamental flowering plants based on DNA barcodes–leaf morphology. Ecol. Evol. 13, e10250. https://doi.org/10.1002/ece3.10250.
Singh, A., Prasad, S.S., Ingle, K.P., Das, U., Ramteke, P.W., Kurubar, A.R., Shukla, P.K., Geethika, P., Madala, R.S., 2024. Molecular marker-assisted selection in plant breeding. In: Al-Khayri, J.M., Ingle, K.P., Jain, S.M., Penna, S. (Eds.), Plant Molecular Breeding in Genomics Era: Concepts and Tools. Springer Nature Switzerland, Cham, pp. 95–111.
Sommer, R.J., 2020. Phenotypic plasticity: from theory and genetics to current and future challenges. Genetics 215, 1–13. https://doi.org/10.1534/genetics.120.303163.
Sun, C., Lei, Y., Li, B., Gao, Q., Li, Y., Cao, W., Yang, C., Li, H., Wang, Z., Li, Y., Wang, Y., Liu, J., Zhao, K.T., Gao, C., 2024a. Precise integration of large DNA sequences in plant genomes using PrimeRoot editors. Nat. Biotechnol. 42, 316–327. https://doi.org/10.1038/s41587-023-01769-w.
Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38, 3022–3027. https://doi.org/10.1093/molbev/msab120.
Tian, M., Li, W., Luo, P., He, J., Zhang, H., Yan, Q., Ye, Y., 2024. Genetic diversity analysis and core germplasm bank construction in cold resistant germplasm of rubber trees (*Hevea brasiliensis*). Sci. Rep. 14, 14533. https://doi.org/10.1038/s41598-024-65464-9.
Tiburtini, M., Scrucca, L., Peruzzi, L., 2025. Using gaussian Mixture Models in plant morphometrics. Perspect. Plant Ecol. Evol. Syst. 69, 125902. https://doi.org/10.1016/j.ppees.2025.125902.
van Oosterhout, C., Supple, M.A., Morales, H.E., Birley, T., Tatayah, V., Jones, C.G., Whitford, H.L., Tollington, S., Ruhomaun, K., Groombridge, J.J., Brickson, L., Keyte, A.L., Shapiro, B., James, M., Turner, S.D., 2025. Genome engineering in biodiversity conservation and restoration. Nat. Rev. Biodivers. 1, 543–555. https://doi.org/10.1038/s44358-025-00065-6.
Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy (Version 1.16.1) [software]. Zenodo. https://doi.org/10.1038/s41592-019-0686-2.
Wang, L., Zheng, Y., Duan, L., Wang, M., Wang, H., Li, H., Li, R., Zhang, H., 2022. Artificial selection trend of wheat varieties released in Huang-Huai-Hai region in China evaluated using DUS testing characteristics. Front. Plant Sci. 13, 898102. https://doi.org/10.3389/fpls.2022.898102.
Waskom, M.L., 2021. Seaborn (Version 0.13.2) [software]. J. Open Source Softw. 6, 3021. https://doi.org/10.21105/joss.03021.
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x.
Wilson, C.A., 2004. Phylogeny of Iris based on chloroplast *matK* gene and *trnK* intron sequence data. Mol. Phylogenet. Evol. 33, 402–412. https://doi.org/10.1016/j.ympev.2004.06.013.
Wu, L., Nie, L., Wang, Q., Xu, Z., Wang, Y., He, C., Song, J., Yao, H., 2021. Comparative and phylogenetic analyses of the chloroplast genomes of species of Paeoniaceae. Sci. Rep. 11, 14643. https://doi.org/10.1038/s41598-021-94137-0.
Wu, X., Alexander, L.W., 2020. Genome-wide association studies for inflorescence type and remontancy in *Hydrangea macrophylla*. Hortic. Res. 7, 27. https://doi.org/10.1038/s41438-020-0255-y.
Wu, X., Simpson, S.A., Youngblood, R.C., Liu, X.F., Scheffler, B.E., Rinehart, T.A., Alexander, L.W., Hulse-Kemp, A.M., 2023. Two haplotype-resolved genomes reveal important flower traits in bigleaf *Hydrangea* (*Hydrangea macrophylla*) and insights

into Asterid evolution. Hortic. Res. 10, uhad217. https://doi.org/10.1093/hr/uhad217.

Yan, L.J., Liu, J., Möller, M., Zhang, L., Zhang, X.M., Li, D.Z., Gao, L.M., 2015. DNA barcoding of Rhododendron (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. Mol. Ecol. Resour. 15, 932–944. https://doi.org/10.1111/1755-0998.12353.

Yang, C.J., Russell, J., Ramsay, L., Thomas, W., Powell, W., Mackay, I., 2021. Overcoming barriers to the registration of new plant varieties under the DUS system. Commun. Biol. 4, 302. https://doi.org/10.1038/s42003-021-01840-9.

Yao, R., Guo, R., Liu, Y., Kou, Z., Shi, B., 2022. Identification and phylogenetic analysis of the genus Syringa based on chloroplast genomic DNA barcoding. PLoS ONE 17, e0271633. https://doi.org/10.1371/journal.pone.0271633.

Yong, W.T.L., Mustafa, A.A., Derise, M.R., Rodrigues, K.F., 2024. DNA barcoding using chloroplast *matK* and *rbcL* regions for the identification of bamboo species in Sabah. Adv. Bamboo Sci. 7, 100073. https://doi.org/10.1016/j.bamboo.2024.100073.

Zhang, L., Yu, X., Zhang, X., Zhang, D., Li, W., Xiang, L., Yang, Y., Li, Y., Xu, L., 2022. Phenotypic diversity analysis of the progeny variation of a 'Mosaic leaf' *Loropetalum chinense* var. *Rubrum* based on flower organ characteristics. Diversity. 14 (11), 913. https://doi.org/10.3390/d14110913.

Zhang, M., Zhai, X., He, L., Wang, Z., Cao, H., Wang, P., Ren, W., Ma, W., 2025. Morphological description and DNA barcoding research of nine Syringa species. Front. Genet. 16, 1544062. https://doi.org/10.3389/fgene.2025.1544062.

Zheng, J., Wang, Y., Pei, S., Hu, Q., 2024. Exploring and exploiting hierarchical structures for large-scale classification. Int. J. Mach. Learn. Cyber. 15, 2427–2437. https://doi.org/10.1007/s13042-023-02039-6.

Zheng, S., Kong, Q., Yan, H., Liu, J., Tang, R., Zhou, L., Yang, H., Jiang, X., Feng, S., Ding, C., Chen, T., 2025. Phenotypic diversity analysis and integrative evaluation of camellia oleifera germplasm resources in Ya'an, Sichuan Province. Plants 14, 2249. https://doi.org/10.3390/plants14142249.

Zhu, H., Zhou, X., 2020. Statistical methods for SNP heritability estimation and partition: A review. Comput. Struct. Biotechnol. J. 18, 1557–1568. https://doi.org/10.1016/j.csbj.2020.06.011.

Zhu, T., Feng, Y., Dong, X., Yang, X., Liu, B., Yuan, P., Song, X., Chen, S., Sui, S., 2023. Optimizing DUS testing for Chimonanthus praecox using feature selection based on a genetic algorithm. Front. Plant Sci. 14, 1328603. https://doi.org/10.3389/fpls.2023.1328603.