agronomy

MDPI

Article

Comprehensive Evaluation of *Nanhaia speciosa* Germplasm Resources Using Agronomic Traits, Molecular Markers, and Metabolomics

Jing Yang [1,†], Nanchen Lai [1,†], Yiqin Zheng [1], Ruifeng Ji [1], Ping Wang [2], Wei Dai [2,*], Gantao Cheng [3] and Xin He [1,4,5,*]

[1] School of Chinese Materia Medica, Guangdong Pharmaceutical University, Guangzhou 510006, China; yangjing@gdpu.edu.cn (J.Y.); lainanchen0203@163.com (N.L.); zhengyiqin2022@163.com (Y.Z.); jiruifeng@gdpu.edu.cn (R.J.)
[2] Teaching and Experimental Center, Guangdong Pharmaceutical University, Guangzhou 510006, China; yyplwp@163.com
[3] Guangdong Xiaoyang Ecological Agriculture Co., Ltd., Yunfu 527500, China; zyzyxyky@163.com
[4] Key Specialty of Clinical Pharmacy, The First Affiliated Hospital of GuangDong Pharmaceutical University, Guangzhou 510080, China
[5] Guangdong Provincial Key Laboratory for Research and Evaluation of Pharmaceutical Preparations, Guangdong Pharmaceutical University, Guangzhou 510006, China
* Correspondence: dai_gdpu_2018@gdpu.edu.cn (W.D.); hexintn@163.com (X.H.); Tel.: +86-020-39352174 (X.H.)
† These authors contributed equally to this work.

Abstract: *Nanhaia speciosa* is a valuable industrial crop known for its significant edible and medicinal properties, attributed to its abundance of secondary metabolites. This study aims to elucidate the genetic relationships among *N. speciosa*, enhance species identification accuracy, and select genetically stable, high-quality germplasms of *N. speciosa*. To achieve this, we conducted chloroplast gene amplification and sequencing, alongside an analysis of the agronomic traits of 19 *N. speciosa* germplasms. Additionally, non-targeted metabolomic profiling using UPLC-Q-Orbitrap/MS and chemometric methods was employed to assess their metabolic diversity and genetic relationships. The results revealed notable differences in DNA barcodes and secondary structures across the germplasms, with the *atp*F-*atp*H sequence successfully distinguishing all 19 germplasms into distinct varieties. Multivariate statistical analyses, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and hierarchical clustering, identified 12 key marker metabolites that effectively differentiated the germplasms, consistent with the observed phenotypic diversity. A clustering analysis integrating genetic, phenotypic, and metabolomic data further validated the utility of DNA barcoding for species identification. The results of the comprehensive analysis showed that NDL-2 and NDL-18 exhibited relatively good edible and medicinal properties, while NDL-1 and NDL-19 exhibited relatively poor edible and medicinal properties. This study provides crucial insights for the breeding and genetic improvement of *N. speciosa* and related species, offering a valuable reference for the identification, conservation, and utilization of germplasm resources within the *N. speciosa*.

Keywords: *Nanhaia speciosa*; genetic diversity; DNA barcoding; UPLC-Q-Orbitrap/MS; nontargeted metabonomic; germplasm resources

check for updates

Academic Editor: Yong-Bao Pan

Received: 21 January 2025
Revised: 18 February 2025
Accepted: 19 February 2025
Published: 20 February 2025

Citation: Yang, J.; Lai, N.; Zheng, Y.; Ji, R.; Wang, P.; Dai, W.; Cheng, G.; He, X. Comprehensive Evaluation of *Nanhaia speciosa* Germplasm Resources Using Agronomic Traits, Molecular Markers, and Metabolomics. *Agronomy* **2025**, *15*, 508. https://doi.org/10.3390/agronomy15030508

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

N. speciosa (Champ. ex Benth.) J. Compton & Schrire, a member of the Fabaceae family [1,2], is a perennial liana native to subtropical regions of southern China. In its natural habitat, it thrives in thickets, sparse forests, and open areas at altitudes below 1500 m, with its geographical distribution, between 18° and 30° N latitude, notably in Hainan, Guangxi, and Guangdong provinces [3]. This species is valued for its dual edible and medicinal properties. It is commonly used as a raw material for food products such as *N. speciosa* rice flour, a traditional Chinese snack [4]. More significantly, its medicinal applications are well-documented in traditional Chinese medicine (TCM). As early as the Ming dynasty, the text Essential Properties of Raw Herbs described *N. speciosa* for its efficacy in "strengthening tendons, invigorating blood circulation, nourishing the lungs, and treating lumbar pain, rheumatism, chronic hepatitis, and pulmonary tuberculosis".

Recent pharmacological studies have confirmed its multifunctional bioactivities, including anti-hepatic fibrosis, hypoglycemic, antioxidant, anti-inflammatory, antidepressant, and analgesic effects [5–7]. These properties are attributed to its rich repertoire of secondary metabolites [8], comprising sugars (α,α-trehalose, rhamnose, lactose, glucose, mannose, and fructose) [9], flavonoids (esculetin, diosmetin, glycitein, formononetin, (−)-maackiain, and medicarpin) [10], alkaloids (anthranilic acid, tyrosine, phenylalanine, and hypaphorine), and organic acids (D-(−)-quinic acid, cis-aconitic acid, pantothenic acid, salicylic acid, 2-hydroxycinnamic acid, and 3-hydroxybenzoic acid) [11]. These bioactive compounds underpin its applications in health products such as medicinal wines and herbal teas [12,13].

Despite its economic and therapeutic value importance, the sustainable utilization of *N. speciosa* faces critical challenges. Overharvesting has severely depleted the wild populations, necessitating the reliance on cultivated sources [14]. However, challenges such as germplasm misidentification, growth heterogeneity, and inconsistent phytochemical profiles have compromised the product quality and market value [15,16]. Systematic germplasm resource management (encompassing collection, preservation, characterization, and evaluation) is imperative for breeding programs aimed at developing elite cultivars [17]. The current germplasm assessment can be analyzed using various methods, including morphological markers, molecular markers, and non-targeted metabolomics [18,19].

Morphological markers are simple and effective for visually distinguishing and classifying plant varieties without the need for complex analytical techniques [20]. Molecular markers, such as DNA barcoding, are highly efficient for species identification and genetic evaluation [21]. DNA barcoding utilizes species-specific DNA sequences as molecular markers, facilitating the accurate identification of species and the study of their evolutionary relationships [22]. Non-targeted metabolomics is an emerging field that provides comprehensive insights into the chemical diversity of plants, enabling applications in plant variety identification, the origin tracing of plant-based foods, and chemical variation analysis among different species [23]. These methods have been successfully applied to various species. For example, phenotypic traits and SNP markers have been used to assess the genetic diversity of cowpea germplasm [24]; Kress W.J., et al. evaluated the genetic variation of invasive plants in Ontario, Canada through DNA barcoding technology, and studied intraspecific differences, providing a basis for the management of popular species [25]; and enzyme markers have been employed for genetic diversity analysis in celery [26]. The integration of DNA barcoding with morphological and chemical analyses has proven effective for the quality evaluation of *Pruni Semen* [27]. Similarly, chemometrics-assisted metabolomics has been utilized for the identification of spices, herbs, and food varieties [28–30].

In this study, we integrate DNA barcoding with UPLC-Q-Orbitrap/MS-based metabolomics to systematically evaluate the genetic diversity, agronomic traits, and key

metabolites of *N. speciosa* germplasms. Our objective was to establish a reliable framework for germplasm identification, and identify key metabolites linked to medicinal efficacy, and then provide a scientific basis for breeding superior cultivars and sustainable resource utilization. This multidisciplinary approach advances the genetic improvement and commercial development of *N. speciosa* resources.

2. Materials and Methods

2.1. Experimental Reagents and Plant Materials

2.1.1. Chemical Reagents

Genomic DNA was extracted using the Plant Genomic DNA Extraction Kit (Accurate Biotechnology Co., Ltd., Changsha, China). Agarose powder and ultrapure water were purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China). Tris and EDTA were supplied by Aladdin (Shanghai, China). HPLC-grade formic acid and acetonitrile were procured from Thermo Fisher Scientific (Guangzhou, China), methanol from Guangzhou Chemical Reagent Factory (Guangzhou, China), and ultrapure water from A.S. Watson Tm Limited (Guangzhou, China).

2.1.2. Plant Materials

A total of 19 *N. speciosa* germplasms were collected from three major production regions in southern China: Guangdong (n = 7), Guangxi (n = 6), and Hainan (n = 6). Detailed sample codes, geographic origins, and GPS positioning are listed in Table 1. All germplasms were subsequently transplanted to the Xiaoyang Agricultural *Nanhaia speciosa* Standardized Cultivation Base (23°12′ N, 113°25′ E; GAP-certified) under uniform agroecological conditions. Cultivation followed standardized protocols: planting density: 1.2 m × 1.5 m spacing; soil management: lateritic red soil (pH 5.8–6.3), and organic matter ≥ 2.5%; agronomic practices: drip irrigation, monthly growth monitoring, and pest control according to national medicinal plant cultivation guidelines (GB/T 29344-2023).

Table 1. Naming and source of origin of plant materials.

Naming	Source (Town, City, Province)	GPS Positioning (North Latitude, East Longitude)
NDL-1	Fulin, Yunfu, Guangdong	22.6830161, 111.9125324
NDL-2		22.6939529, 111.9153171
NDL-3	Xingping, Yangshuo, Guangxi	24.9220809, 110.5309201
NDL-4		24.9396041, 110.5366816
NDL-5	Yao Gu, Yunfu, Guangdong	22.8870317, 112.2879645
NDL-6		22.8872130, 112.2874863
NDL-7		22.8504033, 112.3070249
NDL-8		22.8504489, 112.3072114
NDL-9		22.8485882, 112.3068318
NDL-10		22.8469868, 112.3067984
NDL-11	Gaoliang, Deqing, Guangdong	23.1771246, 111.9555338
NDL-12		23.2899006, 111.9535036
NDL-13		23.2899567, 111.9535581
NDL-14	Bolao, Lingshan, Guangxi	22.1199661, 109.1271791
NDL-15		22.1071753, 109.1192201
NDL-16		22.1215900, 109.1172618
NDL-17	Heshe, Danzhou, Hainan	19.5938770, 109.7374584
NDL-18	Heqing, Danzhou, Hainan	19.5279646, 109.6673995
NDL-19		19.5355961, 109.6783948

Nineteen *N. speciosa* germplasms with three biological replicates per accession were analyzed in this study. For DNA extraction, healthy young leaves from each sample were collected, washed, dried, flash-frozen with liquid nitrogen, and stored in an ultra-low temperature freezer. For biochemical and metabolomic analyses, the entire plant of each sample was washed, dried, flash-frozen with liquid nitrogen, and stored in an ultra-low-temperature freezer. Botanical identification of all specimens was performed by Dr. Yang Jing (School of Traditional Chinese Medicine Resources, Guangdong Pharmaceutical University) through morphological characterization. All corresponding genomic/metabolomic samples are stored in Prof. He Xin's laboratory under controlled conditions (−80 °C for tissues, 4 °C for extracts). A representative field photograph of *N. speciosa* is provided in Figure S1 (Supporting Materials).

2.2. Genome DNA Extraction, PCR Amplification, and Sequence Analysis

2.2.1. Genomic DNA Extraction and Quality Control

Total genomic DNA (gDNA) was isolated from 1 g fresh leaf tissues using the using the Steady Pure Plant Genomic DNA Extraction Kit (Accurate Biotechnology Co., Ltd., Changsha, China) following the manufacturer's protocol. DNA integrity was verified via 1% (*w*/*v*) agarose gel electrophoresis (Bio-Rad Gel Doc™ XR+ Imaging System, Hercules, CA, USA), and purity was assessed by spectrophotometry (A260/A280 ratio: 1.8–2.0).

2.2.2. PCR Amplification and Optimization

Zhang C.Y., et al. showed that *atp*F-*atp*H and *psbk-psbl* are non-coding regions in the chloroplast genome with a high rate of variation, which is suitable for species identification [31]. These regions are easy to amplify, and the primers are versatile enough for a wide range of plant groups. Specific primer sequences and amplification conditions are provided in the literature as a reference for the design of this experiment. Two DNA barcode, *psbk-psbl* and *atp*F-*atp*H, were amplified using gene specific primers (Table 2). The PCR reactions (50 µL final volume) containing 25 µL of 2 × Phanta Max Buffer (Vazyme Biotech Co., Ltd., Nanjing, China), 2 µL of gDNA (20–50 ng/µL), 2 µL of each forward and reverse primer (10 µM), 1 µL of Phanta Max Super-Fidelity DNA Polymerase (2.5 U/µL), 1 µL of dNTP mix (10 mM), and 17 µL PCR-grade ddH_2O. Optimal annealing temperatures for each primer pair were determined through gradient PCR (Table 2). The PCR products were visualized by 2% (*w*/*v*) agarose gel electrophoresis in 1 × TAE buffer and visualized under UV light.

Table 2. Information of DNA barcodes and PCR reaction program.

DNA Barcode	Forward and Reverse Primers	Primer Sequences (5′–3′)	PCR Reaction Program
psbk-psbl	*psbk-psbl*-F	TTAGCATTTGTTTGGCAAG	95 °C, 3 min; 95 °C, 15 s, 49.6 °C, 15 s, and 72 °C, 90 s, 35×; 72 °C, 5 min
	psbk-psbl-R	AAAGTTTGAGAGTAAGCTA	
*atp*F-*atp*H	*atp*F-*atp*H-F	ACTCGCACACACTCCCTTTCC	95 °C, 3 min; 95 °C, 15 s, 56.1 °C, 15 s, and 72 °C, 90 s, 35×; 72 °C, 5 min
	*atp*F-*atp*H-R	GCTTTTATGGAAGCTTTAACAAT	

2.2.3. Sequencing and Bioinformatics

An analysis using bidirectional sequencing was performed by Tsingke Biotechnology Co. (Guangzhou, China). Raw chromatograms were processed via BioEdit 7.2.5, with low-quality regions removed and sequences corrected manually. For taxonomic identification, consensus sequences were aligned against the NCBI nucleotide database using BLASTn (E-value $\leq 1 \times 10^{-10}$). Similar sequences from closely related species were downloaded from GenBank (detailed sequence information is provided in Table S1 in Supplementary

Material). We employed Kimura two-parameter model "https://pubmed.ncbi.nlm.nih.gov/7463489/ (accessed on 15 February 2025)" to analyze genetic distance model, and the Bootstrap test which set the number of repetitions (1000) to evaluate the node support rate. The Neighbor-Joining (NJ) algorithm is used for the construction of phylogenetic trees.

2.3. Genetic Diversity Analysis of Agronomic Traits

Agronomic phenotypic traits were analyzed using the Shannon–Wiener index (H′) "https://pubmed.ncbi.nlm.nih.gov/14078577/ (accessed on 15 February 2025)". Phenotypic traits and biochemical characteristics were classified and graded (Table 3). The average observed values (X) and standard deviations (σ) were calculated for each trait, and relative frequencies were determined based on $X \pm K\sigma$ (K = 0, 1, 2). The diversity index (H′) was calculated using the following formula:

$$H' = -\sum(Pi \times lnPi) \quad (1)$$

where Pi is the relative frequency of a trait level, N is the total number of samples, and Ni is the number of individuals at a specific level.

Table 3. Grading of observed values for agronomic traits.

Level	Observed Value
1	$X1 \leq X - 2\sigma$
2	$X - 2\sigma < X2 \leq X - \sigma$
3	$X - \sigma < X3 \leq X$
4	$X < X4 \leq X + \sigma$
5	$X + \sigma < X5 \leq X + 2\sigma$
6	$X6 > X + 2\sigma$

Standard deviation, mean, maximum value, minimum value, coefficient of variation, and H′ values were calculated using Excel 16.0, while SPSS 26.0 was employed for correlation analysis, cluster analysis, and principal component analysis (PCA). Leaf length, leaflet width, leaflet length, and root diameter were measured using a ruler. For biochemical traits, plant roots were ground into fine powder and extracted at 60 °C with a 1:10 (*w*/*v*) solid-to-solvent ratio using 50% methanol, with extraction performed using an ultrasonic device for 30 min. The extracted solution was used as Sample 1, and Sample 2 was prepared by mixing with a color reagent for absorbance measurement. High-performance liquid chromatography (HPLC) was utilized to measure the peak areas of Vitamin B_3 and Hypaphorine, which were then used to calculate the final content values.

2.4. Preparation of MS Samples

For each batch of the 19 *N. speciosa* germplasms, biological triplicates (*n* = 3) were processed through an optimized extraction protocol. Precisely 0.2 g of fine plant powder was mixed with 50% methanol at a solid-to-liquid ratio of 1:20 (*w*/*v*). Ultrasonication-assisted extraction was conducted at 50 °C for 60 min using a thermostatic ultrasonic bath (40 kHz, 500 W), and subsequently centrifuged at 3500 rpm for 10 min using a high-speed centrifuge. The supernatants were filtered through 0.22 μm PTFE membranes prior to ultra-high-performance liquid chromatography–quadrupole–orbitrap mass spectrometry (UPLC-Q-Orbitrap/MS) analysis. For quality assurance, a quality control (QC) sample was prepared by pooling 20 μL of each of the 57 individual samples. The pooled QC sample was thoroughly vortexed to ensure phase homogeneity. All samples were maintained at 4 °C in amber glass vials during autosampler storage to prevent photodegradation.

2.5. UPLC-Q-Orbitrap/MS Analysis

Chromatographic analysis was performed on a Vanquish Flex UPLC system (Thermo Fisher Scientific), equipped with a Hypersil GOLD C_{18} column (100 × 2.1 mm, 1.9 µm). The mobile phase comprised (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile. A multistep gradient was implemented: 0–12 min from 95% to 5% A (linear), 12–13.001 min isocratic at 5% A, 13.001–13.1 min return to initial conditions (95% A), followed by 1.9 min column re-equilibration. The system operated at 0.3 mL/min flow rate with column temperature maintained at 35 °C. Mass spectrometric detection was performed using an Orbitrap Exploris 120 mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a heated electrospray ionization (HESI) source. Key parameters were optimized as follows: sheath gas 50 psi, auxiliary gas 8 psi; capillary temperature 325 °C; vaporizer temperature 350 °C; and S-lens RF 60%. Full scan MS1 spectra (m/z 100–1500) were acquired at 60,000 FWHM resolution, with data-dependent MS^2 scans triggered above 1×10^4 intensity threshold at 15,000 resolutions. Dynamic exclusion was set to 15 s with ±10 ppm mass tolerance. Collision energy ramping (20%, 40%, and 60%) enabled comprehensive fragmentation. Instrument control and data acquisition were managed through Xcalibur 4.3 software under positive/negative ion switching mode (spray voltage +3.5/−2.8 kV).

Mass spectrometric data were processed using Compound Discoverer 3.3 (Thermo Fisher Scientific, USA). This processing involved peak extraction, alignment, and compound identification by matching MS/MS spectra against the mzCloud and mzVault databases, with identification criteria requiring a mass deviation of less than 5 ppm and a matching score of at least 80.

Use MetaboAnalyst V5.0 "https://www.metaboanalyst.ca/ (accessed on 15 February 2025)" for data processing, with specific parameters: upload data first, select "Remove missing values", select Normalization by sum, select logarithmic transformation, and remove low variance metabolites (variance < 10%). Statistical analysis: multivariate analysis: principal component analysis (PCA): default parameters for data dimensionality reduction and visualization. Partial least squares discriminant analysis (PLS-DA): categorical variables are set for analysis of differences between groups. Metabolite annotation: metabolite annotation is performed using the built-in database (KEGG). Set the mass error range (±10 ppm) and retention time error range (±0.2 min). Pathway analysis: metabolite Set Enrichment Analysis (MSEA) was selected and the significance threshold was set ($p < 0.05$).

2.6. Multicriteria Evaluation System for Germplasm Performance

To assess the edible and medicinal potential of the 19 germplasms of *N. speciosa*, a comprehensive scoring system was developed based on key phenotypic, edible, and medicinal indicators. These quality indicators are determined by reference to Guo Peng's master's thesis [32]. The scoring process involved calculating the average value (X) and standard deviation (σ) for each indicator and categorizing the results into six grades according to the formula X ± Aσ, where A = 0, 1, or 2. According to Kumar S., et al., X (mean) ± Aσ (standard deviation) were used as grading scoring criteria for the analysis of variation in plant agronomic phenotype and metabolite content [33]. In this study, we also referred to the grading scoring method of Kumar S, and further proposed the score: the narrower and shorter the leaves, the shorter the root diameter, the lower the vitamin B_3 content, the lower the protein content, the lower the total flavonoid content, and the smaller the VIP value, indicating that the agronomic phenotypic traits and biochemical characteristics and metabolite levels were worse. Therefore, as shown in Tables 4 and 5, the worst agronomic phenotypic traits, biochemical characteristics, and metabolite levels were divided into the first level, which was numerically $X1 \leq X - 2\sigma$, and assigned a score of 1.

Those with poor agronomic phenotypic traits, biochemical characteristics, and metabolite levels were divided into the second level, which were $X - 2\sigma < X2 \leq X - \sigma$ numerically, and were assigned two points, and so on. As Kim M. J., et al. have shown, leaflet length and leaflet width are relatively unimportant in terms of edible indices [34], so leaflet length and leaflet width are assigned a score of 0.5 points in the first level. The specific weights and assigned points are as follows:

Table 4. Scoring table of edible quality indicators.

Level	Observed Value	Leaflet Length	Leaflet Width	Root Diameter	Vitamin B_3	Protein	Total Sugars
1	$X1 \leq X - 2\sigma$	0.5	0.5	1.0	1.0	1.0	1.0
2	$X - 2\sigma < X2 \leq X - \sigma$	1.0	1.0	2.0	2.0	2.0	2.0
3	$X - \sigma < X3 \leq X$	1.5	1.5	3.0	3.0	3.0	3.0
4	$X < X4 \leq X + \sigma$	2.0	2.0	4.0	4.0	4.0	4.0
5	$X + \sigma < X5 \leq X + 2\sigma$	2.5	2.5	5.0	5.0	5.0	5.0
6	$X6 > X + 2\sigma$	3.0	3.0	6.0	6.0	6.0	6.0

Table 5. Scoring table of medicinal quality indicators.

Level	Observed Value	Vitamin B_3	Total Sugars	Total Flavonoids	VIP Score	
					α,α-Trehalose	Hypaphorine
1	$X1 \leq X - 2\sigma$	1.0	1.0	1.0	1.0	1.0
2	$X - 2\sigma < X2 \leq X - \sigma$	2.0	2.0	2.0	2.0	2.0
3	$X - \sigma < X3 \leq X$	3.0	3.0	3.0	3.0	3.0
4	$X < X4 \leq X + \sigma$	4.0	4.0	4.0	4.0	4.0
5	$X + \sigma < X5 \leq X + 2\sigma$	5.0	5.0	5.0	5.0	5.0
6	$X6 > X + 2\sigma$	6.0	6.0	6.0	6.0	6.0

For edible indicators, the weights were assigned as follows: leaflet length and leaflet width were each assigned a weight of 0.1, while root diameter, vitamin B_3 content, protein content, and total sugars were each assigned a weight of 0.2 (Table 4). For medicinal indicators, the weights were distributed as follows: vitamin B_3 content, total sugars, and total flavonoids were each assigned a weight of 0.2, while the variable importance in projection (VIP) value, representing the importance of differential metabolites (with two metabolites having VIP > 5 selected as observed indicators), was assigned a weight of 0.4 (Table 5).

The comprehensive performance score for each germplasm was calculated using a weighted average of the edible and medicinal scores. Statistical analyses, including mean value calculation, standard deviation, and coefficient of variation, were performed using Microsoft Excel. Additionally, correlation analysis, cluster analysis, and principal component analysis (PCA) were conducted using SPSS 26.0 to evaluate the genetic diversity among the germplasms and identify promising candidates for future breeding programs.

3. Results

3.1. Molecular Characterization of Chloroplast Barcodes

The chloroplast intergenic regions *psbk-psbl* and *atp*F-*atp*H were universally amplified from all 19 *N. speciosa* germplasms, with distinct electrophoretic bands observed in 2% agarose gels (Figure 1). Bidirectional sequencing confirmed high-quality sequences (QV > 30) for all amplicons. A comparative analysis against 29 NCBI reference sequences

(Table S1) was performed using BLASTn with default parameters (E-value < 1×10^{-5}). After multiple sequences with MegAlign alignment and manual correction, the key molecular features were evaluated and summarized in Table 6.

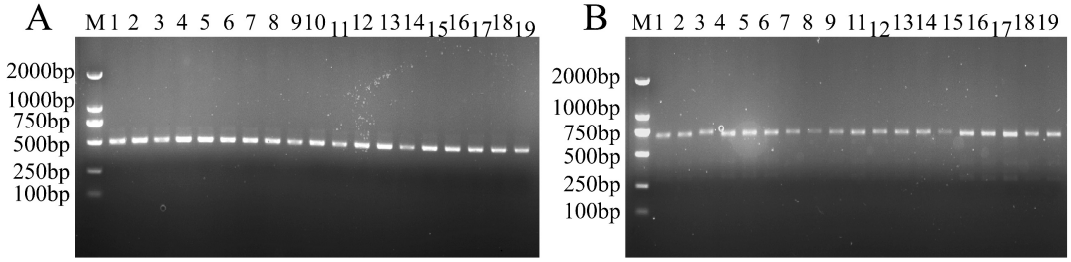


Figure 1. (**A**) Agarose gel electrophoresis of *psbk-psbl* sequence from samples. (**B**) Agarose gel electrophoresis of *atp*F-*atp*H sequence from samples. 1–19 correspond to NDL-1 to NDL-19.

Table 6. DNA barcoding sequence analysis.

Sequence	Alignment Length (bp)	GC Content (%)	Conserved Site	Conservation Index (%)	Variability Index (%)	Variable Sites	Parsimony Informative Sites
psbk-psbl	444	28.05	426	95.9	4.1	18	10
*atp*F-*atp*H	599	25.04	547	91.3	8.7	52	33

The *atp*F-*atp*H region demonstrated superior length polymorphism (599 bp vs. 444 bp in *psbk-psbl*), while *psbk-psbl* exhibited a higher thermostability potential as evidenced by the elevated GC content (28.05% vs. 25.04%). The level of sequence conservation calculated by MEGA 11.0. showed *psbk-psbl* maintained 95.9% conserved sites (426/444), significantly higher than *atp*F-*atp*H's 91.3% (547/599). Conversely, *atp*F-*atp*H displayed greater genetic variability with the following: (1) 8.7% variable sites (52/599) vs. 4.1% (18/444) in *psbk-psbl*; and (2) 5.5% parsimony-informative sites (33/599) vs. 2.3% (10/444).

Among the variable sites in the *atp*F-*atp*H region, specific contributions were observed from NDL-8, NDL-11, NDL-2, NDL-3, and NDL-4, with germplasms NDL-10 and others showing minimal contributions. These findings suggest that, among the 19 germplasms of *N. speciosa* analyzed, the *atp*F-*atp*H barcode exhibits the highest sequence variability, making it a promising candidate for a DNA barcode to differentiate *N. speciosa*.

3.2. Molecular Divergence Assessment and Barcode Gap Evaluation

Genetic distances were calculated using the Kimura 2-Parameter (K2P) model to assess the interspecific and intraspecific variability of the *psbk-psbl* and *atp*F-*atp*H barcode sequences (Table 7). As shown in Table 7, the *atp*F-*atp*H intergenic spacer demonstrated significantly higher molecular variation than *psbk-psbl* at both taxonomic levels ($p < 0.01$, Mann–Whitney U test). The average intraspecific genetic distance was larger for *atp*F-*atp*H compared to *psbk-psbl*. Similarly, the average interspecific genetic distance was also greater for *atp*F-*atp*H than for *psbk-psbl*. A comparative analysis of the genetic distance distributions revealed that *psbk-psbl* exhibited a minimal genetic distance both within and between species, with a certain degree of overlap between intraspecific and interspecific distances. In contrast, *atp*F-*atp*H displayed larger genetic distances within and between species, with a clear separation between the two categories. The average interspecific genetic distance for *atp*F-*atp*H was significantly greater than its average intraspecific genetic distance, further supporting its higher discriminatory power. These results highlight the superior variability and barcoding gap of the *atp*F-*atp*H barcode, making it a more effective molecular marker for distinguishing species within the *N. speciosa*.

Table 7. Comparison of intraspecific and interspecific genetic distance differences of two candidate barcodes in *N. speciosa*.

Sequence	Intraspecific Genetic Distance			Interspecific Genetic Distance		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
psbk-psbl	0	0.0027	0.0021	0.0017	0.0568	0.0360
*atp*F-*atp*H	0	0.0786	0.0644	0.1743	1.2937	0.7431

3.3. Phylogenetic Analysis

The phylogenetic reconstruction conclusively validated the discriminatory capacity of the *atp*F-*atp*H intergenic spacer as a molecular barcode. Utilizing sequence polymorphisms and phylogenetically informative sites identified through preliminary analyses, we generated a neighbor-joining (NJ) phylogenetic tree with 1000 bootstrap replicates in MEGA 11.0 (Figure 2). The resulting topology demonstrated significant phylogenetic resolution, segregating the germplasms into distinct clades with notable evolutionary divergence. Notably, five germplasms (NDL-2, NDL-3, NDL-4, NDL-8, and NDL-11) formed well-defined monophyletic clusters, while the remaining germplasms clustered together.

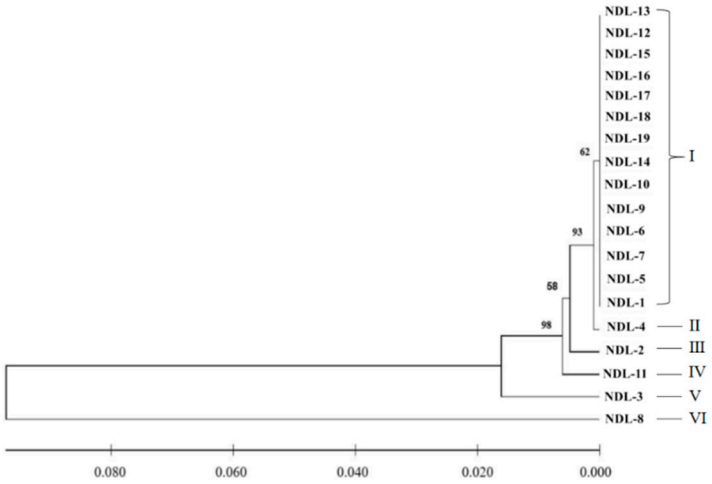


Figure 2. Phylogenetic tree constructed based on the *atp*F-*atp*H sequence. In the figure, I–VI represent different clusters.

The phylogenetic architecture revealed NDL-8 as the most divergent lineage, followed sequentially by NDL-3, both exhibiting substantial genetic distance from the primary cluster (branch length > 0.012 substitutions/site). All major nodes received strong statistical support (bootstrap values 62–98%), confirming the reliability of the inferred relationships. These findings underscore the utility of the *atp*F-*atp*H sequence as a core DNA barcode for accurately identifying and classifying *N. speciosa*, particularly within the context of plant biodiversity and conservation efforts in China.

3.4. Analysis of Variance in Agronomic Phenotypic Traits and Biochemical Characteristics

An analysis of variance revealed significant inter-germplasm differences ($p < 0.05$) in both agronomic phenotypic traits and biochemical characteristics among *N. speciosa* accessions (Table 8). The comprehensive phenotypic characterization demonstrated substantial morphological and biochemical diversity across the germplasm collection.

Table 8. Phenotypic trait variance analysis and biochemical characteristic variance analysis for agronomic traits.

Sample	Leaf Length (cm)	Leaflet Length (cm)	Leaflet Width (cm)	Root Diameter (cm)	Vitamin B3 Content (μg/g)	Protein Content (mg/g)	Total Sugars Content (mg/g)	Total Flavonoids Content (mg/g)	Hypaphorine Content (mg/g)
NDL-1	19.7 ± 2.06 A	8.97 ± 4.25 AB	2.83 ± 1.30 BD	0.70 ± 0.15 C	29.25 ± 1.01 A	0.89 ± 0.11 I	43.00 ± 0.33	0.89 ± 0.07	1.94 ± 0.18 B
NDL-2	15.77 ± 1.72 ABC	8.56 ± 2.20 ABC	4.77 ± 1.45 A	1.10 ± 0.36 ABC	39.60 ± 0.50 C	1.09 ± 0.14 J	128.87 ± 0.81 A	1.00 ± 0.02 E	1.63 ± 0.09 A
NDL-3	12.26 ± 2.10 CD	4.47 ± 0.38 E	2.53 ± 0.45 C	0.92 ± 0.22 ABC	42.69 ± 0.75 D	1.06 ± 0.07 J	80.96 ± 1.94 D	1.00 ± 0.03 F	1.62 ± 0.11 A
NDL-4	7.77 ± 2.55 EF	4.93 ± 0.45 DE	1.70 ± 0.46 C	1.17 ± 0.20 ABC	43.51 ± 0.55 D	0.36 ± 0.06 BC	121.01 ± 0.23 E	0.72 ± 0.08 A	1.63 ± 0.14 A
NDL-5	12.70 ± 1.45 CD	10.20 ± 1.97 A	5.07 ± 1.45 A	1.18 ± 0.36 ABC	23.01 ± 0.48	0.62 ± 0.04 EF	123.24 ± 0.27	0.86 ± 0.07 D	1.93 ± 0.21 B
NDL-6	15.23 ± 2.00 BC	6.06 ± 1.75 BCDE	2.27 ± 0.45 C	0.83 ± 0.35 BC	36.50 ± 1.00 B	0.59 ± 0.08 EF	113.21 ± 0.48	0.80 ± 0.10 BC	1.66 ± 0.54
NDL-7	12.63 ± 1.65 CD	4.81 ± 0.48 E	4.88 ± 0.16 A	1.13 ± 0.24 ABC	60.69 ± 0.77	0.51 ± 0.04 DE	109.21 ± 0.19	0.79 ± 0.08 B	1.78 ± 0.38
NDL-8	9.93 ± 1.08 DE	8.93 ± 1.84 ABC	1.60 ± 0.20 C	1.52 ± 0.11 A	11.57 ± 0.53	0.52 ± 0.03 DE	122.58 ± 0.32	0.85 ± 0.01	1.91 ± 0.25 B
NDL-9	15.37 ± 1.35 BC	6.08 ± 0.75 E	1.81 ± 0.18 C	0.73 ± 0.37 C	43.83 ± 0.46 D	0.84 ± 0.07 HI	112.69 ± 0.57	0.83 ± 0.05	1.58 ± 0.13
NDL-10	7.68 ± 2.38 EF	4.98 ± 0.58 DE	1.75 ± 0.52 C	1.21 ± 0.17 ABC	45.10 ± 0.64 E	0.54 ± 0.03 DE	122.21 ± 0.13 E	0.73 ± 0.06 A	1.53 ± 0.28
NDL-11	4.83 ± 1.45 FG	6.50 ± 2.10 BCDE	4.97 ± 1.56 A	1.31 ± 0.32 BC	51.74 ± 0.62	0.37 ± 0.02 BC	117.26 ± 0.36 B	0.92 ± 0.10 E	2.07 ± 0.33 C
NDL-12	9.73 ± 1.27 DE	5.37 ± 1.90 CDE	2.67 ± 0.65 C	0.99 ± 0.31 ABC	29.72 ± 0.50 A	0.44 ± 0.02 CD	109.42 ± 0.24 C	1.04 ± 0.10 F	1.96 ± 0.16 B
NDL-13	12.53 ± 2.09 CD	4.37 ± 1.20 DE	1.60 ± 0.20 C	0.97 ± 0.20 ABC	29.92 ± 0.61 A	0.67 ± 0.03 FG	95.93 ± 0.25	0.91 ± 0.06	0.94 ± 0.11
NDL-14	3.78 ± 0.61 G	7.73 ± 1.46 ABCD	4.37 ± 1.42 AB	1.42 ± 0.21 A	47.78 ± 0.72	0.19 ± 0.05 A	116.86 ± 0.66 B	1.16 ± 0.09 F	2.08 ± 0.34 C
NDL-15	12.53 ± 1.70 CD	4.20 ± 1.05 E	2.53 ± 0.45 C	0.91 ± 0.24 ABC	28.84 ± 0.65 A	0.19 ± 0.06 A	81.34 ± 0.23 D	0.81 ± 0.06 C	1.02 ± 0.23
NDL-16	17.07 ± 2.42 AB	3.83 ± 0.35 E	1.67 ± 0.21 C	0.81 ± 0.20 BC	49.66 ± 0.61	0.27 ± 0.03 AB	137.76 ± 0.26	0.92 ± 0.11 E	1.61 ± 0.14 A
NDL-17	5.67 ± 2.50 FG	4.30 ± 0.95 DE	1.50 ± 0.20 C	0.96 ± 0.46 ABC	45.00 ± 0.78 E	0.29 ± 0.07 AB	108.61 ± 0.36 C	0.87 ± 0.09 D	1.35 ± 0.25
NDL-18	15.83 ± 1.67 ABC	8.59 ± 2.21 ABC	4.81 ± 1.47 A	1.12 ± 0.37 ABC	38.96 ± 0.77 C	0.38 ± 0.05 BC	129.42 ± 1.58 A	1.23 ± 0.10	1.71 ± 0.26
NDL-19	7.77 ± 2.45 EF	5.93 ± 0.65 E	2.00 ± 0.10 C	0.69 ± 0.27 C	37.44 ± 0.44 B	0.75 ± 0.06 GH	88.91 ± 0.26	0.91 ± 0.07 E	1.16 ± 0.13

Note: Values represent mean ± SD (n = 3). Capital letters indicate significant differences ($p < 0.05$, Tukey's HSD) within columns.

In terms of leaf morphological variation, germplasms NDL-1, NDL-2, NDL-6, NDL-9, NDL-16, and NDL-18 exhibited significantly longer leaf lengths (>15 cm), with NDL-1 showing the maximum length of 19.7 ± 2.06 cm. In contrast, NDL-14 displayed the shortest leaf length (3.78 ± 0.61 cm). Leaflet dimensions showed distinct patterns, where NDL-5 displayed the longest leaflet (length: 10.20 ± 1.97 cm; width: 5.07 ± 1.45 cm), while NDL-8 and NDL-14 had the narrowest leaflet width (1.60 ± 0.20 cm and 4.37 ± 1.42 cm, respectively).

Regarding root traits, root diameter measurements revealed NDL-15 had the longest root diameter (1.42 ± 0.21 cm), contrasting with NDL-19's minimal root diameter (0.69 ± 0.27 cm). Most germplasms maintained intermediate root diameters ranging from 0.70 cm to 1.30 cm, suggesting moderate genetic variability in this trait.

In biochemical composition profiling, the biochemical analysis uncovered notable metabolic diversity. The vitamin B_3 content ranged from 11.57 ± 0.53 μg/g (NDL-8) to 60.69 ± 0.77 μg/g (NDL-7). Protein content extremes were observed between NDL-2/NDL-3 (1.09 ± 0.14 mg/g) and NDL-14/NDL-16 (0.19 ± 0.05 mg/g). Total sugars showed a 3.2-fold variation, from 43.00 ± 0.33 mg/g (NDL-1) to 137.76 ± 0.26 mg/g (NDL-16). Flavonoid accumulation peaked in NDL-18 (1.23 ± 0.10 mg/g), while NDL-4 and NDL-10 showed minimal synthesis (0.72–0.73 mg/g). Hypaphorine content demonstrated a 2.2-fold variation, with maxima in NDL-11/NDL-14 (2.07–2.08 mg/g) and minima in NDL-2/NDL-3/NDL-4/NDL-16 (1.62–1.63 mg/g).

3.5. Diversity Analysis of Agronomic and Biochemical Traits in N. speciosa Germplasms

A comprehensive diversity analysis of 19 *N. speciosa* germplasms revealed substantial phenotypic and biochemical variation, with coefficients of variation (CV) ranging from 14.29% to 48.43% across measured parameters (Table 9). This quantitative characterization provides critical insights into the germplasm's genetic potential for breeding applications.

Table 9. Variation parameters of agronomic and biochemical traits in *N. speciosa* germplasms.

Traits	Mean ± SD	Range	CV (%)
Morphological			
Leaf length/cm	11.51 ± 4.22	3.78–19.70	36.66
Leaflet length/cm	6.25 ± 1.87	3.83–10.20	29.92
Leaflet width/cm	2.92 ± 1.31	1.50–5.07	44.86
Root diameter/cm	1.03 ± 0.22	0.69–1.52	21.36
Biochemical			
Vitamin B_3 content μg/g	38.67 ± 11.09	11.57–60.69	28.68
Protein content mg/g	0.56 ± 0.26	0.19–1.09	48.43
Total sugar content mg/g	108.55 ± 1.75	43.00–137.76	20.04
Total flavonoids content mg/g	0.91 ± 0.13	0.72–1.23	14.29
Hypaphorine content mg/g	1.64 ± 0.32	1.02–2.08	19.51

The morphological analysis demonstrated distinct patterns of variation: leaf dimensions exhibited moderate to high variability (CV: 29.92–44.86%); leaf length showed the greatest absolute variation (extreme difference: 15.92 cm); leaflet width displayed the highest morphological (CV: 44.86%); and root architecture showed relatively conserved variation (CV: 21.36%), with the maximum diameter (1.52 cm) 2.2-fold greater than the minimum (0.69 cm). These findings suggest a strong environmental adaptability in foliar development compared to the more genetically constrained root system development.

Biochemical traits displayed exceptional diversity: protein content showed maximum variability (CV: 48.43%), ranging 5.7-fold from 0.19 to 1.09 mg/g; vitamin B_3 exhibited the largest absolute biochemical difference (49.12 μg/g range); total flavonoids demonstrated remarkable stability (CV: 14.29%), with values confined to 0.72–1.23 mg/g; and hypaphorine content maintained a moderate variation (CV: 19.51%) despite a 2.04-fold concentration range. The inverse relationship between protein content variability and flavonoid stability suggests distinct regulatory mechanisms governing nitrogenous compound versus secondary metabolite biosynthesis.

Notable differential patterns emerged: the leaflet width (CV: 44.86%) vs. length (CV: 29.92%) variation indicates the independent genetic control of these foliar dimensions; protein content's exceptional variability (CV: 48.43%) highlights the potential for nutritional enhancement through selective breeding; the total sugars' moderate variation (CV: 20.04%) contrasts with its wide concentration range (43.00–137.76 mg/g), suggesting threshold-regulated accumulation.

These patterns provide strategic targets for germplasm utilization: high-variation traits (protein and leaflet width) offer immediate selection potential, while stable traits (flavonoids) may require molecular interventions for modification.

The Shannon–Wiener diversity index (H′) analysis revealed substantial genetic variability among 10 key traits in *N. speciosa* germplasms, with indices ranging from 1.118 to 1.470 (mean = 1.282), indicating moderate to high genetic diversity (Table 10). Notably, the vitamin B_3 content demonstrated the highest genetic diversity (H′ = 1.417), followed by leaflet length (H′ = 1.412) and protein content (H′ = 1.470). From the multidimensional trait comparison analysis, it can be observed that the contents of vitamin B_3 (H′ = 1.417), protein (H′ = 1.470), and the length of small leaves (H′ = 1.412) exhibit the highest genetic diversity,

with their distribution frequencies showing typical multimodal distribution characteristics. Specifically, the content of vitamin B_3 displays an even distribution across levels 1, 2, and 6 (with frequencies of approximately 15.5% for each level), while the cumulative frequency for levels 3 and 4 reaches 71.5%. This distribution pattern suggests that this trait may be regulated by the synergistic action of multiple genetic loci or significantly influenced by environmental factors. In contrast, the content of total flavonoids ($H' = 1.118$) shows a unimodal distribution across levels 2–5 (with level 4 reaching 32.8%), indicating a relatively conservative genetic basis.

Table 10. Genetic diversity index and distribution frequency.

Traits and Biochemical Characteristics	Distribution Frequency						H'
	1	2	3	4	5	6	
Leaf length	0.000	0.291	0.351	0.368	0.328	0.000	1.338
Leaflet length	0.000	0.328	0.364	0.237	0.328	0.155	1.412
Leaflet width	0.000	0.291	0.354	0.155	0.364	0.000	1.167
Root diameter	0.000	0.328	0.364	0.364	0.237	0.155	1.211
Vitamin B_3 content	0.155	0.155	0.351	0.364	0.237	0.155	1.417
Protein content	0.000	0.328	0.368	0.328	0.291	0.155	1.470
Total sugar content	0.155	0.237	0.291	0.316	0.237	0.000	1.236
Total flavonoids content	0.000	0.237	0.316	0.328	0.237	0.000	1.118
Hypaphorine content	0.155	0.237	0.368	0.368	0.237	0.000	1.365

Notably, traits with high genetic diversity ($H' > 1.3$) exhibit two typical distribution patterns: (1) the "bimodal distribution" of protein content and small leaf length (with frequencies of 32.8% at levels 2 and 5, respectively), which may reflect the segregation of dominant/recessive alleles; and (2) the "plateau distribution" of the content of hypaphorine (with frequencies of 36.8% at levels 3 and 4), which implies the continuous variation characteristic of quantitative traits. The equal frequency distribution of the root diameter thickness at levels 3 and 4 (each at 36.4%) further validates the stable genetic characteristics of underground organ traits.

These findings suggest that *N. speciosa* germplasms maintain considerable genetic variability, particularly in nutrition-related biochemical traits (vitamin B_3 and protein content) and foliar morphological characteristics. The bimodal distribution patterns observed in several traits may indicate disruptive selection pressures or the presence of major effect genes governing these characteristics. The high diversity indices for pharmacologically relevant compounds (hypaphorine and flavonoids) highlight the species' potential for medicinal applications through targeted breeding programs.

3.6. Multivariate Analysis of Agronomic and Biochemical Traits in N. speciosa Germplasms

A Pearson correlation analysis of nine agronomic and biochemical traits across 19 *N. speciosa* germplasms revealed distinct relationship patterns (Figure 3A). The results revealed that intra-group correlations showed that Total sugars exhibited the strongest positive correlation ($r = 0.910$), while leaflet width showed a minimal association ($r = -0.015$). The inter-group correlation analysis showed the following: (1) Leaflet width correlated positively with leaflet length ($r = 0.523$, $p < 0.05$); (2) leaf length was negatively associated with root diameter ($r = -0.491$, $p < 0.01$); (3) in terms of biochemical–morphological interactions, the root diameter positively correlated with total sugars ($r = 0.512$, $p < 0.05$); and (4) hypaphorine had significant positive correlations with leaflet width ($r = 0.652$), leaflet length ($r = 0.630$), and root diameter ($r = 0.504$) ($p < 0.01$). These results suggest a coordinated regulation between photosynthetic structures (leaf dimensions) and root

development, with hypaphorine potentially serving as a biochemical mediator in resource allocation strategies.

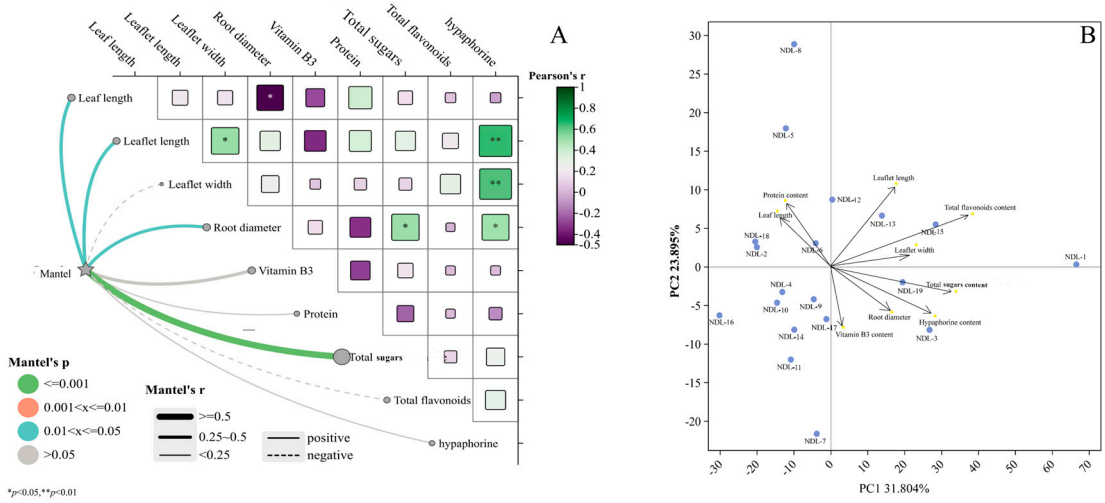


Figure 3. (**A**) Correlation analysis of agronomic phenotypic traits and biochemical characteristics; and (**B**) principal component analysis of agronomic phenotypic traits and biochemical characteristics.

Principal component analysis (PCA) was performed on the nine traits and characteristics, resulting in the extraction of three principal components with a cumulative contribution rate of 69.313% (Figure 3B and Table 11). The first principal component (PC1) had an eigenvalue of 2.862, contributing 31.804% of the total variance, with leaflet width, root diameter, and hypaphorine content being the primary indicators (loadings of 0.743, 0.794, and 0.767, respectively). The second principal component (PC2), with an eigenvalue of 2.151, explained 23.895% of the total variance, with leaf length and protein content as the key indicators (loadings of 0.708 and 0.712). The third component (PC3), with an eigenvalue of 1.225, contributed 13.614% of the variance, with vitamin B_3 content being the main indicator (loading of 0.822). A PCA biplot revealed that the first principal component, explaining 31.804% of the variance, was significantly influenced by leaflet width, root diameter, and hypaphorine content, while the second principal component, explaining 23.895% of the variance, was influenced by leaf length and protein content. The analysis highlighted that leaflet width, leaf length, root diameter, hypaphorine content, and protein content had significant contributions to the classification of the germplasms, serving as important evaluation criteria for *N. speciosa* germplasms.

The cluster analysis, based on Euclidean distance, classified the 19 *N. speciosa* germplasms into two phylogenetically significant clusters (Figure 4). Cluster I (NDL-2, NDL-4, NDL-5, NDL-8, NDL-10, NDL-11, and NDL-18) exhibited characteristics such as a longer leaf length, a larger root diameter, higher protein and sugar contents, and a higher hypaphorine content. In contrast, Cluster II (remaining 12 germplasms), showed a compact architecture (shorter leaf length, narrower leaflet width, and smaller root diameter), vitamin B_3-enriched but with reduced protein and hypaphorine. The convergent results from multivariate analyses demonstrate the following: (1) trait co-evolution: this results in a strong integration between root morphology and alkaloid biosynthesis (PC1); (2) resource allocation trade-offs: negative leaf–root correlations suggest carbon partitioning constraints; and (3) breeding implications: PC1-associated traits (hypaphorine and root diameter) represent key targets for medicinal cultivar development.

Table 11. Principal component analysis of agronomic phenotypic traits and biochemical characteristics.

Traits and Biochemical Characteristics	PCA		
	PC1	PC2	PC3
Vitamin B_3 content	0.057	−0.483	0.822
Protein content	−0.266	0.712	0.139
Total sugars content	0.532	−0.427	0.011
Total flavonoids content	0.497	0.269	0.327
Hypaphorine content	0.767	−0.310	0.362
Eigenvalue (λ)	2.862	2.151	1.225
Contribution rate (%)	31.804	23.895	13.614
Cumulative contribution rate (%)	31.804	55.699	69.313
Leaf length	−0.308	0.708	0.226
Leaflet length	0.626	0.667	−0.269
Leaflet width	0.743	0.270	0.406
Root diameter	0.794	−0.310	−0.362

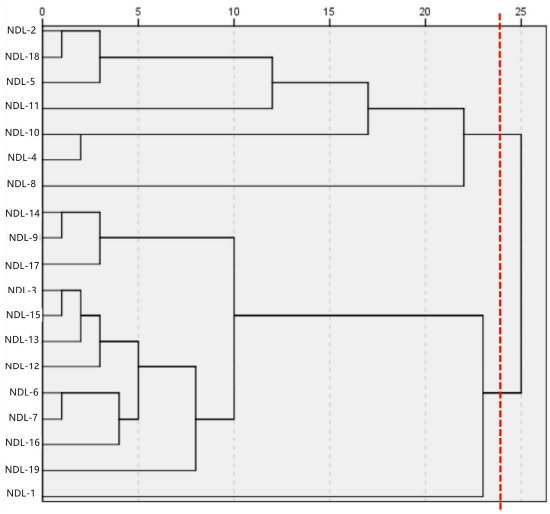


Figure 4. Euclidean distance clustering figure of agronomic phenotypic traits and biochemical characteristics. The red line in the diagram divides all germplasm into two clusters.

3.7. Metabolomic Differences Analysis of Different Germplasms of N. speciosa

3.7.1. UPLC-Q-Orbitrap/MS Was Employed for the Qualitative Analysis of Samples

The results showed that 123 metabolites were identified in this study, as depicted in Figure 5. These metabolites were classified into eight major phytochemical classes: organic acids (32 compounds, 26.02%), amino acids (23 compounds, 18.70%), alkaloids (21 compounds, 17.07%), flavonoids (16 compounds, 13.01%), sugars (7 compounds, 5.69%), purines (5 compounds, 4.06%), quinones (2 compounds, 1.63%), and miscellaneous metabolites (17 other compounds, 13.82%). Notably, the metabolic profile was dominated by nitrogen-containing compounds (amino acids and alkaloids, collectively 35.77%) and antioxidative phytochemicals (flavonoids and organic acids, 39.03%), suggesting an evolutionary adaptation to environmental stressors. The complete annotated metabolite list is provided in Supplementary Table S2.

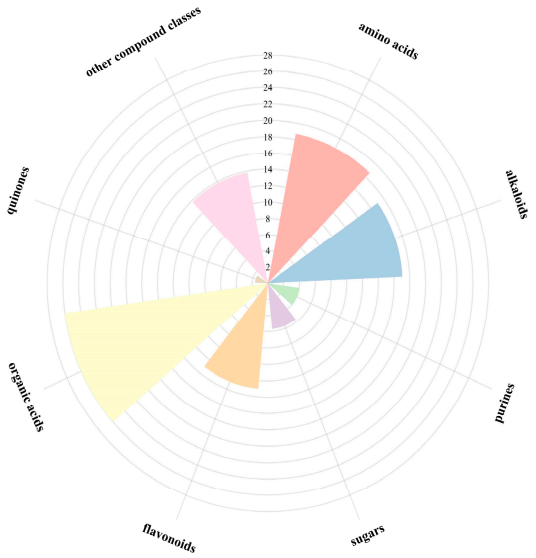


Figure 5. Composition chart of metabolite categories. Each color represents a metabolite category (the area of the sector indicates the proportion of that category).

3.7.2. Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA)

Metabolomic analysis was conducted on 57 samples representing 19 *N. speciosa* germplasms (NDL1–NDL19). Base peak ion chromatograms (BPI) in positive ion mode were generated (Figure 6). Principal component analysis (PCA) was employed to reduce dimensionality and identify key data features. The first three principal components accounted for 59.2%, 28.4%, and 7.3% of the variance, respectively. Both 2D and 3D PCA plots (Figure 7A,B) revealed a significant overlap among germplasms, suggesting the limited group separation and challenges in identifying major metabolite differences.

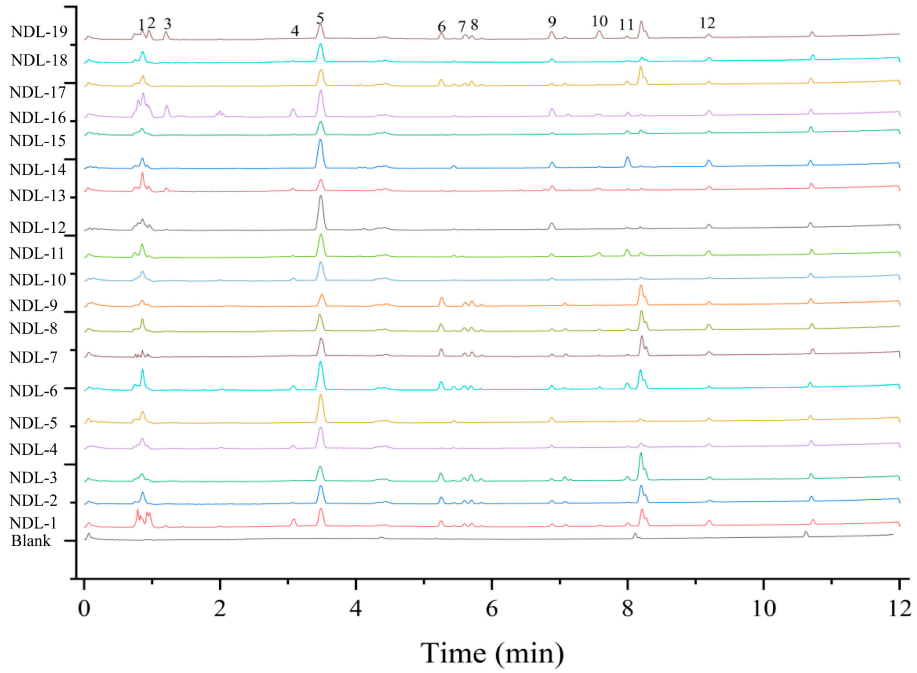


Figure 6. Representative base peak ion chromatograms (BPI) in the positive ion mode were obtained for the QA ions. Peaks 1–12 correspond to the 12 key metabolites listed in Table 12.

Table 12. Key metabolites of *N. speciosa*.

Number	Key Metabolite Markers	Molecular Formula	Molecular Weight	Retention Time (min)	VIP Score
1	α,α-Trehalose	$C_{12}H_{22}O_{11}$	342.1159	0.881	6.5543
2	Citric acid	$C_6H_8O_7$	192.0269	1.204	2.6019
3	2-Hydroxycinnamic acid	$C_9H_8O_3$	164.0474	1.221	0.7748
4	trans-3-Indoleacrylic acid	$C_{11}H_9NO_2$	187.0633	3.075	3.6106
5	Hypaphorine	$C_{14}H_{18}N_2O_2$	246.1367	3.496	7.6284
6	4′,6-Dimethoxyisoflavone-7-O-β-D-glucopyranoside	$C_{23}H_{24}O_{10}$	460.1373	5.766	0.1194
7	3-Hydroxybenzoic acid	$C_7H_6O_3$	138.0316	5.767	0.1050
8	1-Octen-3-yl-6-O-[(2R,3R,4R)-3,4-dihydroxy-4-(hydroxymethyl)tetrahydro-2-furanyl]-β-D-glucopyranoside	$C_{19}H_{34}O_{10}$	422.2151	5.82	0.3250
9	6-Methoxyflavanone	$C_{16}H_{14}O_3$	254.0943	6.953	0.6759
10	Formononetin	$C_{16}H_{12}O_4$	268.0736	7.484	1.7185
11	(−)-Maackiain	$C_{16}H_{12}O_5$	284.0687	7.886	0.1110
12	1,3:2,4-Bis(3,4-dimethylobenzylideno) sorbitol	$C_{24}H_{30}O_6$	414.2045	9.195	4.7971

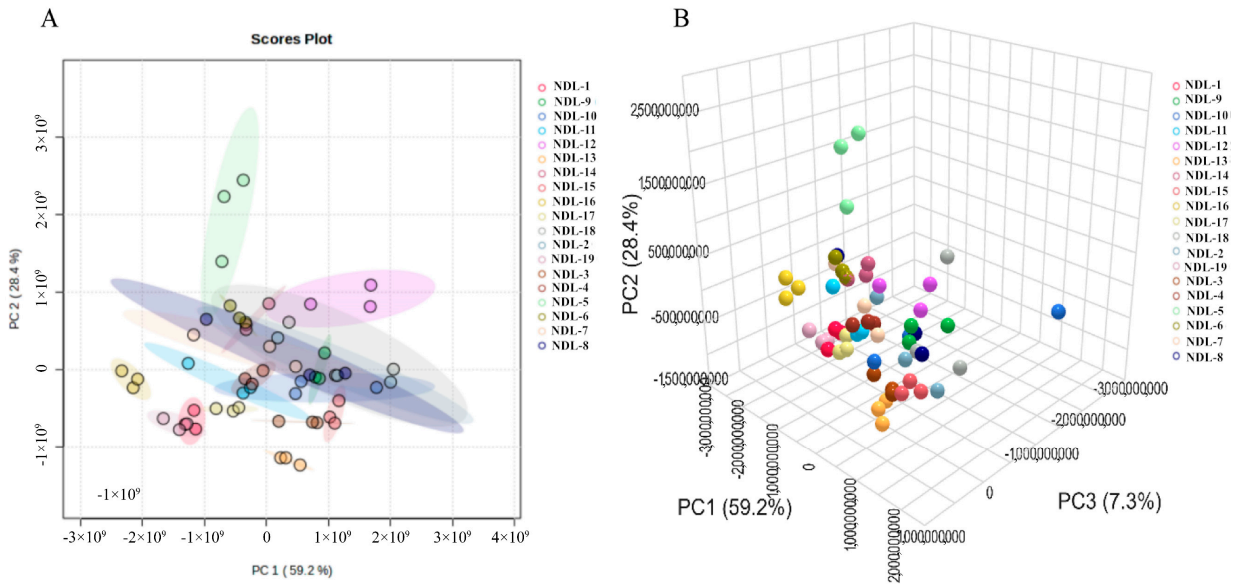


Figure 7. (**A**) PCA 2D score plot. (**B**) PCA 3D score plot.

To further investigate metabolite variations, partial least squares discriminant analysis (PLS-DA) was applied, yielding an R^2 of 0.557 and a Q^2 of 0.297 (Figure 8D). A permutation test (100 iterations) confirmed the model reliability ($p = 0.02$, $p < 0.05$) (Figure 8C). While 2D PLS-DA plots showed overlapping clusters, 3D PLS-DA plots (Figure 8A,B) effectively distinguished group differences, establishing PLS-DA as the superior method for metabolomic data analysis.

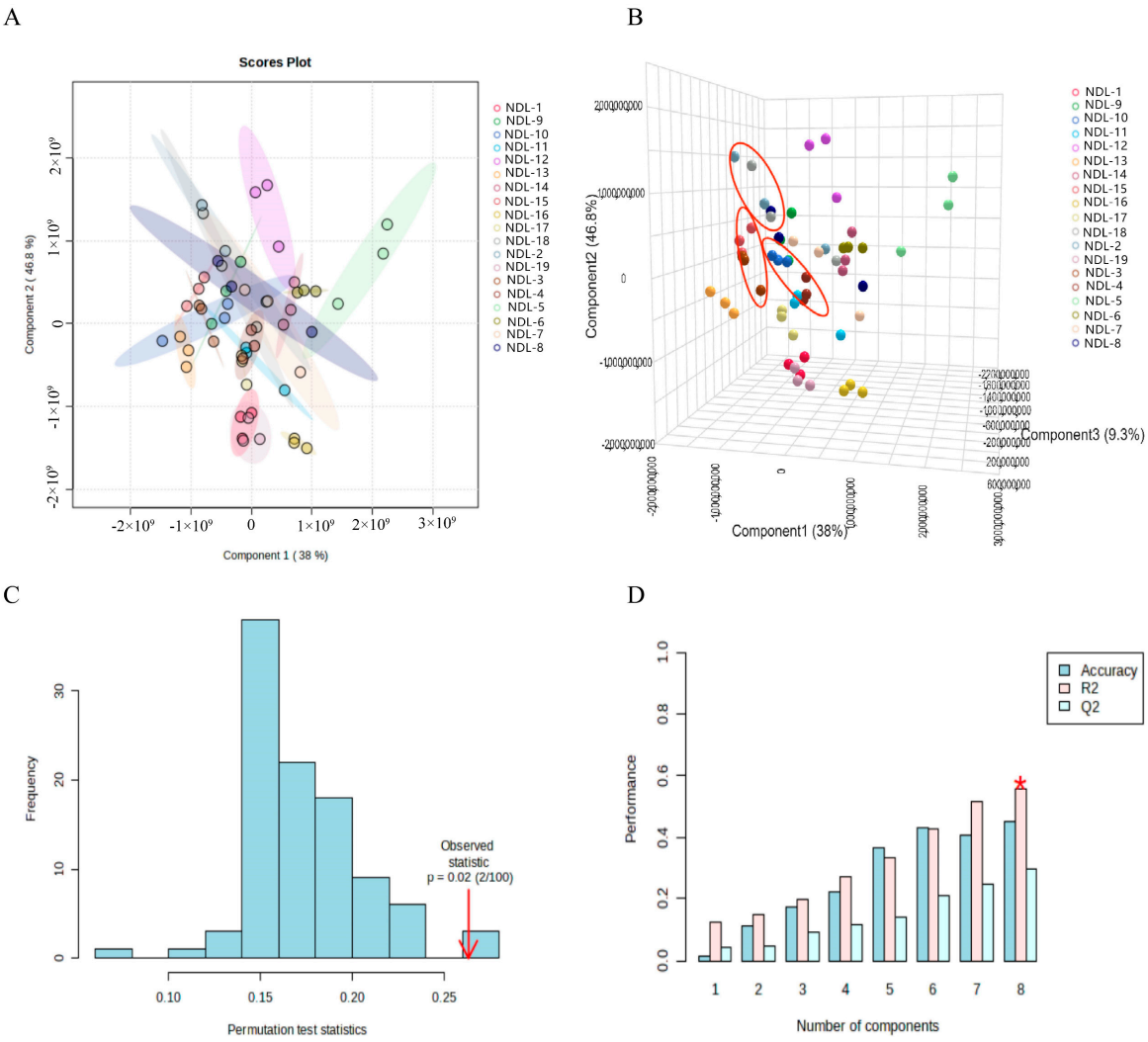


Figure 8. (**A**) PLS-DA 2D score plot. (**B**) PLS-DA 3D score plot. The red circle represents the clustering of the corresponding metabolites. (**C**) Cross-validation plot of 100 permutation tests of PLS-DA model. The red arrow indicates the confidence value. (**D**) Cross-validated variance histogram of 100 permutation tests for PLS-DA model. *, it means that there is a strong significance.

3.7.3. Determination of Key Metabolites and Assessment of Their Differential Abundance Through Heat Map Analysis in Professional Plant Metabolomics

Through comprehensive analysis of PLS-DA VIP scores and positive ion chromatograms, we identified 12 key metabolites that significantly contribute to the metabolic differentiation among 19 *N. speciosa* germplasms (Table 12, Figure 6). To elucidate the metabolic variation patterns, we performed hierarchical clustering analysis and visualized the results through heat map construction (Figure 9B). This heat map analysis revealed distinct clustering patterns among the 19 *N. speciosa* germplasms based on their metabolic profiles. Three major clusters were identified: Cluster I comprised NDL-15 and NDL-8, which exhibited remarkably similar metabolite profiles, suggesting a close genetic relationship; Cluster II included NDL-3, NDL-4, NDL-6, NDL-11, NDL-12, NDL-13, and NDL-16; while Cluster III contained the remaining germplasms.

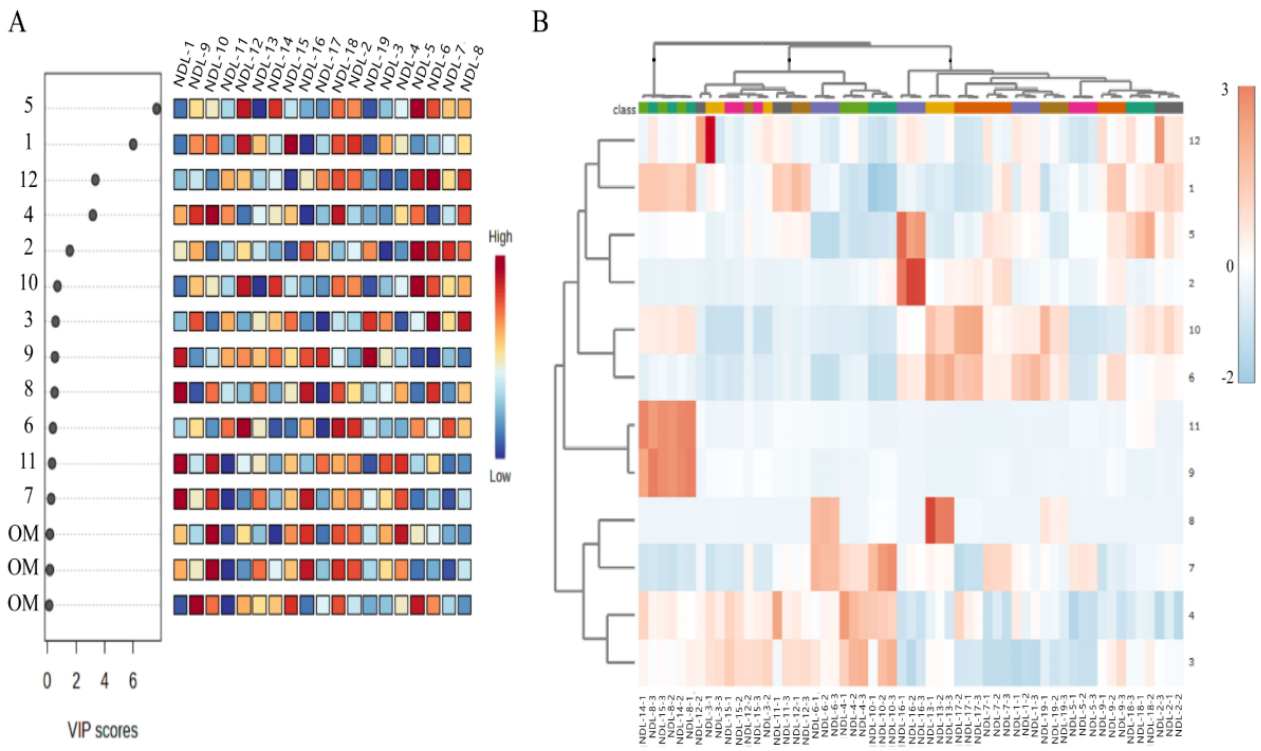


Figure 9. (**A**) Variable importance in projection (VIP) scores from PLS-DA. OM represents other metabolites. (**B**) Hierarchical clustering heat map of key metabolites. Number 1–12 correspond to the 12 key metabolites listed in Table 12.

Notably, we observed significant variations in metabolite abundance across different germplasms. Metabolites 9 (6-Methoxyflavanone) and 11 (−)-Maackiain) demonstrated the highest relative abundance in NDL-15 and NDL-8. In contrast, metabolites 2 (Citric acid) and 5 (Hypaphorine) showed significantly elevated levels in NDL-17 compared to other germplasms. Conversely, NDL-4 and NDL-11 exhibited a notably lower relative abundance of metabolites 1 (α,α-Trehalose) and 5 (Hypaphorine). These findings provide substantial evidence for the metabolic diversity among *N. speciosa* germplasms, offering valuable insights into their genetic relationships and biochemical variations. The differential abundance patterns of key metabolites may serve as potential biomarkers for germplasm characterization and selection in breeding programs.

3.8. Comprehensive Performance Evaluation and Analysis of Different Germplasms of N. speciosa

The evaluation of edible quality in *N. speciosa* germplasms was conducted based on six key indicators: leaf length, leaf width, root diameter, vitamin B_3, protein, and total sugars (Table 13). The mean values (X) and standard deviations (σ) for these indicators were as follows: leaf length (6.25 ± 1.92 cm), leaf width (2.91 ± 1.35 cm), root diameter (1.04 ± 0.23 cm), vitamin B_3 content (38.67 ± 11.09 mg/100 g), protein content (0.56 ± 0.26 g/100 g), and total sugars (108.55 ± 21.75 mg/g). The overall scores for edibility are presented in Table 13, with germplasms NDL-2, NDL-7, NDL-11, and NDL-18 exhibiting good edibility (Total score ≥ 20), and NDL-1, NDL-15, NDL-17, and NDL-19 showing inferior edibility (Total score ≤ 15).

Table 13. Total score table of edible quality indicators.

Germplasm	Leaflet Length	Leaflet Width	Root Diameter	Vitamin B_3	Protein	Total Sugars	Total Score
NDL-1	2.5	1.5	2.0	1.0	5.0	1.0	13.0
NDL-2	2.5	2.5	4.0	4.0	6.0	4.0	23.0
NDL-3	1.5	1.5	3.0	4.0	5.0	2.0	17.0
NDL-4	1.5	1.5	4.0	4.0	3.0	4.0	18.0
NDL-5	2.5	2.5	4.0	2.0	4.0	4.0	19.0
NDL-6	1.5	1.5	3.0	3.0	4.0	4.0	17.0
NDL-7	1.5	2.5	4.0	5.0	3.0	4.0	20.0
NDL-8	2.5	1.5	6.0	1.0	3.0	4.0	18.0
NDL-9	1.5	1.5	2.0	4.0	5.0	4.0	18.0
NDL-10	1.5	1.5	4.0	4.0	3.0	4.0	18.0
NDL-11	2.0	2.5	5.0	5.0	3.0	4.0	21.5
NDL-12	1.5	1.5	3.0	3.0	3.0	4.0	16.0
NDL-13	1.5	1.5	3.0	3.0	4.0	3.0	16.0
NDL-14	2.0	2.5	5.0	4.0	2.0	4.0	19.5
NDL-15	1.0	1.5	3.0	3.0	2.0	2.0	12.5
NDL-16	1.0	1.5	2.0	5.0	2.0	5.0	16.5
NDL-17	1.0	1.0	3.0	4.0	2.0	4.0	15.0
NDL-18	2.5	2.5	4.0	4.0	3.0	4.0	20.0
NDL-19	1.5	1.5	2.0	3.0	4.0	3.0	15.0

For the evaluation of medicinal quality, the analysis focused on the vitamin B_3 content, total sugars, total flavonoids, and the variable importance in projection (VIP) values of the selected metabolites (VIP > 5, totaling 2 metabolites) (Table 14). The mean values and standard deviations for these indicators were as follows: vitamin B_3 content (38.67 $\pm$ 11.09 mg/100 g), total sugars (108.55 $\pm$ 21.75 mg/g), total flavonoids (0.91 $\pm$ 0.13 mg/g), VIP values (6.55 $\pm$ 0.83), and selected metabolites (total VIP score: 7.63 $\pm$ 0.77). The overall scores for medicinal properties are provided in Table 14, with germplasms NDL-2, NDL-12, and NDL-18 exhibiting excellent medicinal properties (Total score > 20), and NDL-1, NDL-13, and NDL-19 showing poor medicinal properties (Total score < 15).

The comprehensive evaluation revealed that germplasms NDL-2 and NDL-18 exhibited the highest overall performance in both edibility and medicinal properties. These germplasms were characterized by longer and wider leaves, larger root diameters, a higher vitamin B_3 and protein content, elevated levels of total sugars and flavonoids, and significant VIP values for metabolites with medicinal potential. In contrast, germplasms NDL-1 and NDL-19 demonstrated inferior performance in both edibility and medicinal quality, as evidenced by shorter and narrower leaves, smaller root diameters, and lower levels of vitamin B_3, protein, total sugars, total flavonoids, and VIP values.

This comprehensive evaluation provides critical insights for the selection of *N. speciosa* germplasms with optimal characteristics for both edibility and medicinal applications. The findings underscore the importance of integrating multiple quality indicators in the assessment of germplasm performance, thereby facilitating the identification of superior genotypes for targeted breeding programs and commercial utilization.

Table 14. Total score table of medicinal quality indicators.

Germplasm	Vitamin B_3	Total Sugars	Total Flavonoids	VIP Score		Total Score
				α,α-Trehalose	Hypaphorine	
NDL-1	1.0	1.0	3.0	2.0	4.0	11.0
NDL-2	4.0	4.0	4.0	6.0	3.0	21.0
NDL-3	4.0	2.0	4.0	4.0	3.0	17.0
NDL-4	4.0	4.0	2.0	3.0	3.0	16.0
NDL-5	2.0	4.0	3.0	2.0	4.0	15.0
NDL-6	3.0	4.0	3.0	2.0	4.0	16.0
NDL-7	5.0	4.0	3.0	2.0	4.0	18.0
NDL-8	1.0	4.0	3.0	3.0	4.0	15.0
NDL-9	4.0	4.0	3.0	4.0	3.0	18.0
NDL-10	4.0	4.0	2.0	4.0	3.0	17.0
NDL-11	5.0	4.0	4.0	2.0	5.0	20.0
NDL-12	3.0	4.0	5.0	6.0	5.0	23.0
NDL-13	3.0	3.0	4.0	3.0	1.0	14.0
NDL-14	4.0	4.0	5.0	2.0	5.0	20.0
NDL-15	3.0	2.0	3.0	6.0	2.0	16.0
NDL-16	5.0	5.0	4.0	1.0	3.0	18.0
NDL-17	4.0	4.0	3.0	2.0	3.0	16.0
NDL-18	4.0	4.0	6.0	5.0	5.0	24.0
NDL-19	3.0	3.0	3.0	1.0	2.0	12.0

4. Discussion

The integration of multi-omics technologies has emerged as a transformative approach for enhancing the precision of plant germplasm identification and utilization. This study combines DNA barcoding, agronomic phenotyping, and untargeted metabolomics to systematically evaluate the intraspecific diversity in *N. speciosa* germplasms, addressing both the strengths and limitations of single-method approaches.

4.1. DNA Barcoding: Potential and Limitations

DNA barcoding is widely advocated as a tool for resolving taxonomic ambiguities in plant breeding, particularly in biodiversity-rich regions with limited taxonomic expertise [35,36]. It enables rapid species identification by targeting standardized gene regions, and it can also be utilized for intraspecific delineation. Sucher N.J., et al.'s identification of medicinal plants using DNA barcoding technology revealed intraspecific genetic differences to ensure the authenticity and quality control of medicinal materials [37]. Challenges arise from the inconsistent evolutionary rates between organellar and nuclear genomes, potential horizontal gene transfer, and insufficient data on the intraspecific sequence variation [38–40]. Critics argue that single-gene phylogenies may fail to reflect true evolutionary histories, necessitating complementary approaches for robust taxonomic revision [41,42].

In this study, *atp*F-*atp*H outperformed *psbk-psbl* in differentiating *N. speciosa* germplasms, with distinct genetic distances observed among germplasms (e.g., NDL-8 showed the highest variability). However, these findings align with concerns about relying solely on plastid markers, as nuclear–cytoplasmic discordance can obscure true genetic relationships [43]. For instance, nuclear loci such as ITS or low-copy nuclear genes often resolve discrepancies caused by incomplete lineage sorting in plastid barcodes [44], highlighting the need for multilocus analyses.

4.2. Multi-Omics Integration Enhances Germplasm Evaluation

To mitigate the limitations of DNA barcoding, we integrated agronomic phenotyping and metabolomics—a strategy supported by recent advances in multi-omics resource characterization [45]. Agronomic trait assessments play a vital role in cultivar identification and the evaluation of genetic resources [46,47]. In our study, agronomic traits, including leaf morphology and hypaphorine content, effectively distinguished germplasms such as NDL-2 and NDL-12, corroborating studies where the phenotypic diversity exceeded the molecular variation in crops like radish and soybean [48,49]. These traits reflect adaptive evolution under selection pressures, which may not always correlate with neutral genetic markers [50].

Untargeted metabolomics further resolved classification ambiguities. While DNA barcoding grouped NDL-8 and NDL-14 as genetically distant, metabolomic clustering placed them in the same subgroup, implying shared biochemical pathways despite genetic divergence. Such discrepancies may arise from the convergent evolution of metabolic traits under environmental stress or post-transcriptional regulation [51]. For example, the flavonoid biosynthesis in *Dalbergia odorifera* and terpenoid variation in *Artemisia argyi* were driven by ecological adaptation rather than phylogenetic relatedness [52,53]. In our study, the metabolite profiles of NDL-3, NDL-4, and NDL-11 clustered together, suggesting environmental or epigenetic influences on secondary metabolism that override sequence-based relationships.

4.3. Reconciling Molecular and Metabolic Classification Differences

From the phenotypic and biochemical clusters, the samples were divided into two clusters, in which the sample NDL-8 was in cluster I., and NDL-14 was in cluster II, and there was a significant difference between NDL-8 and other samples in cluster I, which was consistent with the phylogenetic tree where the sample NDL-8 formed cluster VI. The discordance between the DNA barcoding and metabolomic results underscores the complex interplay between genetic divergence and phenotypic plasticity. Three factors may explain this: (1) Evolutionary Rate Disparity: Plastid genes (e.g., *atp*F-*atp*H) evolve more slowly than metabolic pathways, leading to conserved sequences but divergent metabolomes under selection [54]; (2) Horizontal Gene Transfer (HGT): Microbial or endophytic gene transfer can introduce metabolic capabilities without altering host genomes, as observed in stress-adapted plants [55]; and (3) Epigenetic Regulation: DNA methylation and miRNA-mediated silencing can decouple metabolite production from genetic relatedness, as reported in medicinal plants with environment-dependent alkaloid synthesis [56].

These mechanisms highlight the necessity of multi-omics frameworks for a comprehensive germplasm characterization. For instance, integrating transcriptomics could bridge the DNA barcode data with metabolic phenotypes by identifying regulatory genes influencing both traits [57].

4.4. Implications for Breeding and Conservation

Our multi-omics approach identified NDL-2 and NDL-18 as superior germplasms for edible and medicinal use, whereas NDL-1 and NDL-19 exhibited undesirable traits for edible and medicinal use. This aligns with the growing emphasis on "phytochemical barcoding", where metabolite profiles complement DNA data to predict bioactivity and adaptability [58]. Future studies should incorporate genomic selection models to prioritize germplasms with synergistic genetic and metabolic advantages [59]. By integrating DNA barcoding, phenotyping, and metabolomics, this study advances the systematic evaluation of *N. speciosa* germplasms. While *atp*F-*atp*H serves as a reliable molecular marker, metabolomics captures the functional diversity driven by ecological and epigenetic factors.

The combined approach not only resolves taxonomic uncertainties but also provides actionable insights for breeding programs. To our knowledge, this is the first multi-omics investigation of *N. speciosa*, establishing a model for studying the genetic–chemical diversity in underutilized species.

5. Conclusions

By integrating DNA barcoding, phenotyping, and metabolomics, this study advances the systematic evaluation of *N. speciosa* germplasms, addressing both the potential and limitations of single-method approaches in plant resource characterization. *atp*F-*atp*H serves as a reliable molecular marker in resolving intraspecific genetic relationships among the 19 germplasms. Untargeted metabolomics based on UPLC-Q-Orbitrap/MS allows for the differentiation of germplasms using key metabolites. Clustering analysis, supported by metabolic data, successfully distinguishes *N. speciosa* germplasms, assisting in the assessment of quality. Our integrated analysis also showed that NDL-2 and NDL-18 are priority candidates (among these 19 *N. speciosa* germplasms) for breeding programs due to their edibility and medicinal properties, whereas NDL-1 and NDL-19 exhibited undesirable traits for edible and medicinal use.

This study contributes valuable insights into the genetic and chemical diversity of *N. speciosa*, and provides actionable insights for breeding programs. To our knowledge, this is the first multi-omics investigation of *N. speciosa*, establishing a model for studying the genetic–chemical diversity in underutilized species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy15030508/s1, Figure S1: The physical image of *Nanhaia speciosa*. Table S1: Sequence information of *Nanhaia speciosa* related species. Table S2: Compounds identified in the methanol extract of *N. speciosa* by UPLC-Q-Orbitrap/MS.

Author Contributions: Conceptualization, J.Y. and W.D.; methodology, J.Y.; software, Y.Z., P.W. and G.C.; validation, N.L., R.J. and P.W.; formal analysis, N.L., Y.Z. and P.W.; investigation, N.L., R.J. and G.C.; resources, G.C.; data curation, N.L. and R.J.; writing—original draft preparation, J.Y. and N.L.; writing—review and editing, X.H. and W.D.; visualization, Y.Z.; supervision, X.H.; project administration, W.D.; funding acquisition, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Project of the Traditional Chinese Medicine (southern medicinal herbs) Industry Innovation Team in Yunfu City in 2022 (YKSTB Letter [2022] No); and the Project of Universities (College) and Enterprise joint funding project of Guangzhou, China (SL2024A03J01134).

Data Availability Statement: The data are contained within the article.

Conflicts of Interest: Author Gantao Cheng was employed by the company Guangdong Xiaoyang Ecological Agriculture Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

N. speciosa	*Nanhaia speciosa*

References

1. Su, Y.X.; He, C.M.; Wang, C. Research progress on cultivation and application of *Nanhaia speciosa*. *Trop. For.* **2019**, *47*, 26–29. (In Chinese)
2. Wan, X.; Zhang, L. Global new species of vascular plants published in 2020. *Biodivers. Sci.* **2021**, *29*, 1003. [CrossRef]
3. Mo, Y.; Zheng, H.; Huang, Y.; Li, P.; Cai, G.; Zeng, Q. Herbal Textual Research and Character Identification of Southern Medicine *Nanhaia speciosa*. *Lishizhen Med. Mater. Med. Res.* **2018**, *29*, 1361–1362. (In Chinese)
4. Zu, B.; Jiang, J.; Long, L.; Liang, Y.; Tan, Z.; Lai, K. Research Overview of Edible *Nanhaia speciosa* Root Soup Products. *Chin. Seas.* **2020**, *45*, 133–136. (In Chinese)
5. Niu, P.; Meng, Q.; He, X.; Wang, S.; Yi, Y.J.; Zhang, Y.J. Study on ultrasonic-assisted extraction process and antioxidant activity of total alkaloids from *Millettia speciosa* Champ. *Food Sci. Technol.* **2023**, *48*, 216–222.
6. Wu, Y.R.; Peng, L.S.; Xiao, J. Research progress on main chemical constituents and pharmacological effects of *Millettia speciosa* Champ. *J. Hunan Univ. Chin. Med.* **2020**, *40*, 503–506.
7. Zhang, C.; Mo, Y.Y.; Feng, S.S.; Meng, M.W.; Chen, S.Y.; Huang, H.M.; Ling, X.; Song, H.; Liang, Y.H.; Ou, S.F.; et al. Urinary metabonomics study of anti-depressive mechanisms of *Millettia speciosa* Champ on rats with chronic unpredictable mild stress-induced depression. *J. Pharm. Biomed. Anal.* **2021**, *205*, 114338. [CrossRef] [PubMed]
8. Zhang, J.; Wang, J.; Wang, Y.; Chen, M.; Shi, X.; Zhou, X.; Zhang, Z. Phytochemistry and antioxidant activities of the rhizome and radix of *Millettia speciosa* based on UHPLC-Q-Exactive Orbitrap-MS. *Molecules* **2022**, *27*, 7398. [CrossRef]
9. Chen, X.; Sun, W.; Xu, B.; Wu, E.; Cui, Y.; Hao, K.; Zhang, G.; Zhou, C.; Xu, Y.; Li, J.; et al. Polysaccharides from the roots of *millettia speciosa* champ modulate gut health and ameliorate cyclophosphamide-induced intestinal injury and immunosuppression. *Front. Immunol.* **2021**, *12*, 766296. [CrossRef] [PubMed]
10. Wang, M.; Zhang, M.; Yang, Q.; Wang, Q.; Ma, B.; Li, Z.; Cheng, W.; Tang, H.; Feng, S.; Wang, Z. Metabolomic profiling of *M. speciosa* champ at different growth stages. *Food Chem.* **2022**, *376*, 131941. [CrossRef]
11. Wang, M.Y.; Ma, W.Y.; Wang, Q.L.; Yang, Q.; Yan, X.X.; Tang, H.; Li, Z.Y.; Li, Y.Y.; Feng, S.X.; Wang, Z.N. Flavonoid-enriched extract from *Millettia speciosa* Champ prevents obesity by regulating thermogenesis and lipid metabolism in high-fat diet–induced obese C57BL/6 mice. *Food Sci. Nutr.* **2022**, *10*, 445–459. [CrossRef]
12. Chen, C.; Liu, P.; Lou, N.; He, Y.; Zhang, Y.; Li, A. Overview of research on the consumption of *Nanhaia speciosa*. *Food Res. Dev.* **2016**, *47*, 168–172. (In Chinese)
13. Yang, D.; Liu, X.; Fan, P.; Wu, Y. Traditional usage of wild edible plants reflecting the dietary habits and the awareness of health care of LI minority in Baoting and Lingshui, Hainan Island, China: An ethnobotanical approach. *Bangladesh J. Bot.* **2019**, *48*, 279–287. [CrossRef]
14. Hang, Y.; Gong, L.; Dong, X.; Xue, D. Research Progress on Development and Utilization of Nanyao *Nanhaia speciosa*. *Spec. Wild Econ. Anim. Plant Res.* **2023**, *1*, 1–13. (In Chinese)
15. Feng, Y.M.; Huang, C.Y.; Lin, Z.; Li, J.; Zhao, B. Research progress on cultivation of *Millettia speciosa* in Lingnan area. *For. Sci. Technol.* **2022**, *6*, 51–54.
16. Gao, W.; Jiang, Z.T.; Mo, Y.L.; Jiang, X.J.; Wei, X.; Shi, Y.C. Effects of different cultivation measures on yield of *Nanhaia speciosa*. *J. Guangxi Acad. Sci.* **2023**, *39*, 71–77. (In Chinese)
17. Boczkowska, M.; Łapiński, B.; Kordulasińska, I.; Dostatny, D.F.; Czembor, J.H. Promoting the use of common oat genetic resources through diversity analysis and core collection construction. *PLoS ONE* **2016**, *11*, 0167855. [CrossRef] [PubMed]
18. Fallahi, F.; Abdossi, V.; Bagheri, M.; Ghanbari Jahromi, M.; Mozafari, H. Genetic diversity analysis of Eggplant Germplasm from Iran: Assessments by morphological and SSR markers. *Mol. Biol. Rep.* **2022**, *49*, 11705–11714. [CrossRef] [PubMed]
19. Wu, W.; Lu, M.; Peng, J.; Lv, H.; Shi, J.; Zhang, S.; Liu, Z.; Duan, J.; Chen, D.; Dai, W.; et al. Nontargeted and targeted metabolomics analysis provides novel insight into nonvolatile metabolites in Jianghua Kucha tea germplasm (*Camellia sinensis* var. *assamica* cv. Jianghua). *Food Chem. X* **2022**, *13*, 100270. [CrossRef] [PubMed]
20. Chapepa, B.; Mapuranga, R.; Manjeru, P. Characterisation of upland cotton grown in Zimbabwe using agronomical and morphological markers for Verticillium wilt tolerance. *Cogent. Food Agric.* **2020**, *6*, 1766734. [CrossRef]
21. Hebert, P.D.; Cywinska, A.; Ball, S.L.; DeWaard, J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B Biol. Sci.* **2003**, *270*, 313–321. [CrossRef] [PubMed]
22. Hajibabaei, M.; Singer, G.A.; Hebert, P.D.; Hickey, D.A. DNA barcoding: How it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* **2007**, *23*, 167–172. [CrossRef]
23. Liang, L.; Li, Y.; Mao, X.; Wang, Y. Metabolomics applications for plant-based foods origin tracing, cultivars identification and processing: Feasibility and future aspects. *Food Chem.* **2024**, *449*, 139227. [CrossRef] [PubMed]
24. Nkhoma, N.; Shimelis, H.; Laing, M.D.; Shayanowako, A.; Mathew, I. Assessing the genetic diversity of cowpea [*Vigna unguiculata* (L.) Walp.] germplasm collections using phenotypic traits and SNP markers. *BMC Genet.* **2020**, *21*, 110. [CrossRef]
25. Kress, W.J.; Erickson, D.L.; Swenson, N.G.; Thompson, J.; Uriarte, M.; Zimmerman, J.K. DNA barcoding of invasive species in the flora of Ontario, Canada. *Mol. Ecol. Resour.* **2011**, *11*, 1012–1021.

26. Arus, P.; Ortan, T.J. Inheritance patterns and linkage relationships of eight genes of celery. *J. Hered.* **1984**, *75*, 11–14. [CrossRef]
27. Li, P.; Zhao, Z.; Li, Z.; Zeng, R.; Li, W. Distinguishing features of *Prunus humilis*, *P. japonica*, *P. pedunculata* seeds and their adulterant based on DNA barcoding, morphological characterization, and chemical profiles. *Fitoterapia* **2024**, *175*, 105942. [CrossRef] [PubMed]
28. Oliveira, M.M.; Cruz-Tirado, J.P.; Barbin, D.F. Nontargeted analytical methods as a powerful tool for the authentication of spices and herbs: A review. *Compr. Rev. Food Sci. Food Saf.* **2019**, *18*, 670–689. [CrossRef]
29. Shen, S.; Yang, Y.; Wang, J.; Chen, X.; Liu, T.; Zhuo, Q. Analysis of differences between unifloral honeys from different botanical origins based on non-targeted metabolomics by ultra-high performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *J. Chromatogr.* **2021**, *39*, 291–300. [CrossRef] [PubMed]
30. Esteki, M.; Shahsavari, Z.; Simal-Gandara, J. Gas chromatographic fingerprinting coupled to chemometrics for food authentication. *Food Rev. Int.* **2020**, *36*, 384–427. [CrossRef]
31. Zhang, C.Y.; Zhou, Q.; Yang, J.; Hong, D.Y.; Zhang, D.Q. DNA barcoding of Paeonia (Paeoniaceae) using *psbk-psbl* and *atp*F-*atp*H spacer regions. *PLoS ONE.* **2012**, *7*, e41870.
32. Guo, P. Research on Quality Standards of Nanhaia speciosa Medicinal Material. Master's Thesis, Guangzhou University of Chinese Medicine, Guangzhou, China, 2019. (In Chinese)
33. Kumar, S.; Sharma, V.; Chaudhary, S.; Tyagi, A.; Mishra, P.; Priyadarshini, A.; Singh, G.P. Evaluation of phenotypic and metabolic diversity in wheat genotypes under drought stress. *Front. Plant Sci.* **2020**, *11*, 553.
34. Kim, M.J.; Moon, Y.; Tou, J.C.; Mou, B.; Waterland, N.L. Nutritional and phytochemical variability in lettuce cultivars: Implications for breeding and consumer health. *Food Chem.* **2016**, *205*, 375–383.
35. Yu, J.; Wu, X.I.; Liu, C.; Newmaster, S.; Ragupathy, S.; Kress, W.J. Progress in the use of DNA barcodes in the identification and classification of medicinal plants. *Ecotoxicol. Environ. Saf.* **2021**, *208*, 111691. [CrossRef]
36. Blaxter, M.L. The promise of a DNA taxonomy. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2004**, *359*, 669–679. [CrossRef]
37. Sucher, N.J.; Carles, M.C. DNA barcoding of medicinal plant species for authentication and quality control. *Planta Med.* **2014**, *80*, 288–294.
38. Little, D.P.; Stevenson, D.W. A comparison of algorithms for the identification of specimens using DNA barcodes: Examples from gymnosperms. *Cladistics* **2007**, *23*, 1–21. [CrossRef] [PubMed]
39. Meier, R.; Shiyang, K.; Vaidya, G.; Ng, P.K. DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Syst. Biol.* **2006**, *55*, 715–728. [CrossRef] [PubMed]
40. Moritz, C.; Cicero, C. DNA barcoding: Promise and pitfalls. *PLoS Biol.* **2004**, *2*, 354. [CrossRef]
41. Will, K.W.; Rubinoff, D. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* **2004**, *20*, 47–55. [CrossRef] [PubMed]
42. Will, K.W.; Mishler, B.D.; Wheeler, Q.D. The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* **2005**, *54*, 844–851. [CrossRef] [PubMed]
43. Rose, J.P.; Toledo, C.A.; Lemmon, E.M.; Lemmon, A.R.; Sytsma, K.J. Out of sight, out of mind: Widespread nuclear and plastid-nuclear discordance in the flowering plant genus Polemonium (Polemoniaceae) suggests widespread historical gene flow despite limited nuclear signal. *Syst. Biol.* **2021**, *70*, 162–180. [CrossRef]
44. Hollingsworth, P.M.; Forrest, L.L.; Spouge, J.L.; Hajibabaei, M.; Ratnasingham, S.; van der Bank, M.; Chase, M.W.; Cowan, R.S.; Erickson, D.L.; Fazekas, A.J. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* **2009**, *9*, 439–457. [CrossRef] [PubMed]
45. Chen, W.; Gao, Y.; Xie, W.; Gong, L.; Lu, K.; Wang, W.; Li, Y.; Liu, X.; Zhang, H.; Dong, H. Multi-omics revolution in crop improvement: From data integration to predictive breeding. *Trends Plant Sci.* **2022**, *27*, 844–856.
46. Hörandl, E.; Emadzade, K. Evolutionary classification: A case study on the diverse plant genus *Ranunculus* L. (Ranunculaceae). *Perspect. Plant Ecol. Evol. Syst.* **2012**, *14*, 310–324. [CrossRef]
47. Drisya Ravi, R.S.; Nair, B.R.; Siril, E.A. Morphological diversity, phenotypic and genotypic variance and heritability estimates in Moringa oleifera Lam: A less used vegetable with substantial nutritional value. *Genet. Resour. Crop Evol.* **2021**, *68*, 3241–3256. [CrossRef]
48. Huang, T.; Lu, J.; Zhang, W.; Wang, W.; Wan, Y.; Pei, Y.; Mao, F.; Wang, L.; Li, J. Evaluation of genetic variation of morphological and clubroot-resistance traits of radish and metabonomic analysis of clubroot-resistant cultivar. *Sci. Hortic.* **2023**, *321*, 112272. [CrossRef]
49. Yang, C.; Yan, J.; Jiang, S.; Li, X.; Min, H.; Wang, X.; Hao, D. Resequencing 250 soybean accessions: New insights into genes associated with agronomic traits and genetic networks. *Geno. Prot. Bioinform.* **2022**, *20*, 29–41. [CrossRef]
50. Lasky, J.R.; Des Marais, D.L.; McKay, J.K.; Richards, J.H.; Juenger, T.E.; Keitt, T.H. Characterizing genomic variation of Arabidopsis thaliana: The roles of geography and climate. *Mol. Ecol.* **2012**, *21*, 5512–5529. [CrossRef] [PubMed]
51. Jacob, F.; Vernaldi, S.; Maekawa, T. The evolutionary origin of complex features in metabolites and enzymes. *Nat Commun.* **2022**, *13*, 7141.

52. Li, M.; Liu, M.; Wang, B.; Shi, L. Metabolomics analysis of stem extracts from Dalbergia sissoo. *Molecules* **2022**, *27*, 1982. [CrossRef]
53. Alami, M.M.; Ouyang, Z.; Zhang, Y.; Shu, S.; Yang, G.; Mei, Z.; Wang, X. The current developments in medicinal plant genomics enabled the diversification of secondary metabolites' biosynthesis. *Int. J. Mol. Sci.* **2022**, *23*, 15932. [CrossRef] [PubMed]
54. Pichersky, E.; Lewinsohn, E. Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* **2011**, *62*, 549–566. [CrossRef] [PubMed]
55. Hacquard, S.; Spaepen, S.; Garrido-Oter, R.; Schulze-Lefert, P. Interplay between innate immunity and the plant microbiota. *Annu. Rev. Phytopathol.* **2017**, *55*, 565–589. [CrossRef] [PubMed]
56. Strauss, J.; Reyes-Dominguez, Y. Regulation of secondary metabolism by chromatin structure and epigenetic codes. *Fungal Genet. Biol.* **2011**, *48*, 62–69. [CrossRef] [PubMed]
57. Bose, U.; Juhász, A.; Broadbent, J.A.; Komatsu, S.; Colgrave, M.L. Multi-omics strategies for decoding smoke-assisted germination pathways and seed vigour. *Int. J. Mol. Sci.* **2020**, *21*, 7512. [CrossRef]
58. Wolfender, J.L.; Marti, G.; Thomas, A.; Bertrand, S. Plant metabolomics: From holistic data to relevant biomarkers. *Curr. Med. Chem.* **2019**, *26*, 1925–1946.
59. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y. Genomic selection in plant breeding: Methods, models, and perspectives. *Trends Plant Sci.* **2017**, *22*, 961–975. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.