

# Enhancing genomic association studies in slash pine through close-range UAV-based morphological phenotyping

Ruiye Yan<sup>1,2#</sup>, Yihan Dong<sup>1#</sup>, Yanjie Li<sup>1\*</sup>, Cong Xu<sup>3</sup>, Qifu Luan<sup>1</sup>, Shu Diao<sup>1</sup> and Chunyan Wu<sup>4</sup>

<sup>1</sup> State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, Zhejiang, China

<sup>2</sup> College of Landscape and Travel, Agricultural University of Hebei, Baoding 071051, Hebei, China

<sup>3</sup> School of Forestry, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

<sup>4</sup> State Key Laboratory of Tree Genetics and Breeding, Key Laboratory of Tree Breeding and Cultivation of National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China

# Authors contributed equally: Ruiye Yan, Yihan Dong

\* Corresponding author, E-mail: [aj7105@gmail.com](mailto:aj7105@gmail.com)

## Abstract

In forestry genetics and industry, tree morphological traits such as height, crown size, and shape are critical for understanding growth dynamics and productivity. Traditional methods for measuring these traits are limited in efficiency, scalability, and accuracy, posing challenges for large-scale forest assessments. This study focuses on integrating unmanned aerial vehicle (UAV) technology with GWAS to improve genomic association studies in slash pine (*Pinus elliottii*). Seven key morphological traits have been identified (canopy area (CA), crown base height (CBH), crown length (CL), canopy volume (CV), crown width (CW), crown width height (CWH), and tree height (H)) through advanced UAV-based phenotyping. These associations account for a remarkable range of heritability in slash pine, with traits such as CBH, CL, CV, and H showing relatively high heritability across both Single nucleotide polymorphisms (SNP) and pedigree methods, indicating strong genetic influence, while traits such as CWH show lower heritability, suggesting greater environmental influence or non-additive genetic variance. The GWAS identified 28 associations, including 22 different SNPs localized to 16 candidate genes, that were significantly associated with the morphological traits of Slash Pine. Notably, two of these candidate genes, annotated as putative DEAD-like helicase and ethylene-responsive element binding factor (ERF), were present at different mutation sites and were significantly associated with CW and CA traits, respectively. These results demonstrate that the UAV imaging enables a comprehensive analysis of the Morphological growth response of slash pine and can facilitate the discovery of informative alleles to elucidate the genetic structure underlying complex phenotypic variation in conifers.

**Citation:** Yan R, Dong Y, Li Y, Xu C, Luan Q, et al. 2024. Enhancing genomic association studies in slash pine through close-range UAV-based morphological phenotyping. *Forestry Research* 4: e025 <https://doi.org/10.48130/forres-0024-0022>

## Introduction

The morphological phenotypes of trees, including tree height, crown spread, crown length, crown width, height to the lowest branch, crown volume, and height at the widest part of the crown, are fundamental to understanding ecological and environmental dynamics<sup>[1,2]</sup>. These phenotypic attributes are critical not only as indicators of a tree's health and growth patterns but also as integral components in the broader context of global ecological balance. The study of these morphological traits extends beyond simple physical measurements, significantly contributing to our understanding of ecosystem services, biodiversity, and the impact of environmental changes on forest health<sup>[3]</sup>.

Digital Whole-Community Phenotyping (DWCP) exemplifies advancements in capturing these essential phenotypic data by collecting high-resolution multispectral and structural data on plant communities<sup>[4]</sup>. This novel approach underscores the importance of precise phenotypic data collection in advancing ecological research.

In the field of forestry genetics and tree breeding, the detailed analysis of tree morphological phenotypes is paramount. Tree architecture and functional traits are closely interlinked, providing insights into tree life history and demography,

essential for genetic studies and breeding programs<sup>[5-7]</sup>. These characteristics directly influence the timber industry, impacting wood quality, yield, and overall forest productivity. Rapid advancements in sensor technologies and image-based phenotyping are transforming how these traits are measured and analyzed, offering new avenues for accelerating forest breeding<sup>[8]</sup>. Accurate measurement and analysis of these phenotypic traits are crucial for advancing genetic selection and breeding programs aimed at enhancing timber production and forest sustainability. Ongoing research and development in forest tree functional genomics and breeding highlight the critical role of these traits in meeting the growing global demand for wood fiber and other bioproducts from forest trees<sup>[9]</sup>. The development and improvement of these traits through selective breeding and genetic studies hold immense potential for the future of the timber industry and ecological conservation efforts globally<sup>[10]</sup>.

In forestry research, traditional methods for assessing tree morphological phenotypes have primarily relied on manual and direct observational techniques. These approaches involved measuring various physical attributes, including tree height, crown dimensions, and other relevant phenotypic traits. While these methods have historically provided foundational

insights, they have notable limitations in efficiency and scalability<sup>[8,11,12]</sup>. Manual measurements are labor-intensive and subjective, often resulting in significant time investments and potential inconsistencies in data accuracy. These limitations are particularly pronounced when dealing with extensive datasets or conducting wide-area forest surveys, where rapid and accurate data collection is essential<sup>[13]</sup>. Additionally, the low-throughput nature of traditional methods poses a significant challenge in large-scale ecological studies and comprehensive forest monitoring efforts. The difficulty in applying these techniques in varied and often inaccessible terrains further restricts their utility across diverse ecological landscapes<sup>[14]</sup>.

The demands of contemporary forestry research, especially in contexts requiring prompt and precise data gathering—such as monitoring dynamic environmental impacts on forest health or managing extensive commercial forestry operations—necessitate the evolution of phenotyping methodologies<sup>[15,16]</sup>. This need has driven the exploration and adoption of more sophisticated, efficient, and accurate phenotypic data collection techniques, paving the way for innovative approaches in forestry and ecological studies<sup>[17]</sup>.

The advent of Unmanned Aerial Vehicle (UAV) technology and Structure from Motion (SfM) imaging methods has revolutionized forestry phenotyping. SfM, a photogrammetric technique that uses 2D images to reconstruct 3D structures, has emerged as a powerful tool in forestry research. By capturing multiple overlapping images of a forested area from different angles, SfM processes these images to create a 3D model or point cloud of the area. This method is particularly advantageous for providing high-resolution data on forest canopy structure and tree morphology, which are essential for assessing biomass, forest health, and growth patterns<sup>[18]</sup>.

The superiority of SfM methods lies not only in their accuracy and detail but also in their ability to cover large areas rapidly and frequently, allowing for dynamic monitoring of forests<sup>[19]</sup>. This technology offers a non-invasive, cost-effective, and highly scalable approach to forest monitoring, making it invaluable for modern forestry management, ecological research, and conservation efforts. By facilitating detailed and regular assessments of forest structures, UAV technology helps in understanding forest dynamics, carbon sequestration potential, and the impact of environmental changes, playing a crucial role in sustainable forest management and climate change studies<sup>[20]</sup>.

Genome-Wide Association Studies (GWAS) have become increasingly vital in forest genetics, particularly for identifying key genes associated with morphological phenotypic traits<sup>[21]</sup>. This molecular breeding approach holds significant promise for enhancing our understanding of the genetic basis of critical traits like tree height, crown architecture, and overall growth patterns. Identifying these key genes is crucial for advancing forestry genetics and breeding programs, as it enables the development of tree varieties with improved timber quality, disease resistance, and adaptability to environmental changes<sup>[22–24]</sup>.

Despite the popularity of UAV technology in phenotyping and the significance of GWAS in molecular breeding, the integration of high-throughput phenotyping with genotyping remains underexplored, especially in species like slash pine (*Pinus elliottii*)<sup>[25]</sup>. The challenges are compounded by the fact that slash pine, a large coniferous tree, and a gymnosperm, has

not yet had its genome fully sequenced. The complexity of its genome, typical of many large woody perennials, adds another layer of difficulty. While UAV and GWAS individually contribute significantly to the field, their combined application in the genetic improvement and selection of key candidate genes in species like slash pine represent a burgeoning field of study with immense potential yet to be fully explored<sup>[26,27]</sup>.

In response to these challenges, the present research group is pioneering the resequencing of slash pine. Utilizing genomic data from closely related sequenced species and combining it with transcriptome sequencing, a 51k liquid-phased probe array has been developed and specifically designed for slash pine<sup>[28]</sup>. This tool enables targeted resequencing to capture the genetic variations most pertinent to slash pine's morphological traits. This approach significantly enhances our understanding of slash pine's genetic architecture and paves the way for future breeding programs. By linking phenotypic traits with their genetic underpinnings, the aim is to improve the growth rate and wood quality in slash pine.

Therefore, in this study, we used a slash pine plantation to achieve two primary objectives:

- 1) Employ UAV imaging for efficient, high-throughput quantification of slash pine morphological phenotypes. The aim is to systematically collect detailed data on vital traits (canopy area (CA), crown base height (CBH), crown length (CL), canopy volume (CV), crown width (CW), crown width height (CWH) and tree height (H)), which are essential for deciphering growth patterns and structural variations in slash pine.

- 2) Integrate phenotypic data derived from UAV technology with genomic resequencing data to perform a targeted GWAS. This phase focuses on identifying candidate genes that show a significant correlation with the morphological attributes of slash pine, thereby providing valuable insights into the genetic determinants of these traits within the constraints of our sample size.

By achieving these objectives, this study addresses a critical gap in current research by combining advanced UAV phenotyping with genomic analyses, providing comprehensive insights into the genetic and phenotypic variations in slash pine.

## Materials and methods

### Site description

This study was conducted across two meticulously managed slash pine (*Pinus elliottii*) plantations within the Matou National Forest Farm, located in Jingxian County, Xuancheng City, Anhui Province, China. The first site is positioned precisely at the coordinates of 30°45'N and 118°29'E, while the specifics of the second site's location remain closely aligned within the same geographic and climatic zone, ensuring comparable environmental conditions. Both sites are typified by a temperate climate, exhibiting an average annual temperature of 15.7 °C. The thermal profile spans from the chilliest month, January, averaging 2.9 °C, to the peak of warmth in July, averaging 28.1 °C. Precipitation across these regions averages 1,525 mm annually, coupled with an average relative humidity of 84%. The soil in these plantations predominantly consists of yellow loam, characterized by an acidic to neutral pH of 5.5 to 6.0 and depths ranging from 70 to 150 cm. Established in 2013, each plantation spans three hectares, collectively covering six hectares, and is composed of 20 genetically diverse open-pollinated

## UAV-based phenotyping in slash pine genomics

families, each tracing back to the same maternal progenitor. The plantation's design is a lattice incomplete block with single tree plots, where each block, measuring 20 trees with a spacing of 6 m × 8 m, represents an individual family without replication within the block. The entire experimental area across both sites comprises 40 such blocks, effectively splitting the previously singular experimental setup into two distinct but similar environments for comparative analysis. Detailed characteristics and historical data of these plantations have been documented in preceding publications by the same research team<sup>[29–31]</sup>.

### UAV RGB images acquisition

The UAV imagery was acquired using a DJI Matrice 300 RTK (M300RTK) drone, which was equipped with a P1 35 mm camera, provided by Dà-Jiāng Innovations Science and Technology Co., Ltd., China. The flight operation was conducted on December 24, 2023, under optimal weather conditions characterized by clear skies and minimal wind. The M300RTK drone, renowned for its stability and precision, was remotely controlled to maintain a consistent flight altitude, ensuring uniform image quality and coverage.

The camera, designed for advanced aerial photography, employed a swing shooting mode to capture detailed images of the plantation. We meticulously planned the flight path to mirror a double grid pattern, similar to our previous methodology, ensuring 80% image overlap. This strategic approach was crucial for achieving comprehensive coverage of the entire plantation area in a single flight. The M300RTK's advanced flight capabilities allowed us to maintain a forward speed of 8 m/s, optimizing the efficiency of the image capture process.

One of the key features of the M300RTK drone is its integration with a networked real-time kinematic (RTK) positioning system. This system provided highly accurate waypoint positioning, significantly reducing horizontal and vertical positioning errors to 0.03 m and 0.06 m, respectively. The precision of this technology was instrumental in ensuring the high quality of the spatial data acquired during the flight, which lasted approximately 40 min.

To ensure the reproducibility and robustness of the data, multiple flights were conducted under similar conditions, maintaining the same flight altitude of 40 m and clear weather for each session. The use of RTK precise positioning across all flights ensured consistency and accuracy in georeferencing. These repeated flights validated the reproducibility of the image acquisition process, as evidenced by the consistency of the phenotypic data. Previous studies have further confirmed the reliability and robustness of the UAV-based phenotyping approach<sup>[29,30]</sup>.

### Image processing

Utilizing high-precision RGB imagery, the data processing employed the DJI Terra software (version 3.3.0, Shenzhen, China) for generating orthoimages and 3D point cloud data. In this study, the algorithm originally proposed by Song et al. was refined and enhanced, focusing on increased accuracy and speed in data extraction<sup>[29]</sup>. The approach minimized manual intervention by employing automated processes, significantly improving the efficiency of handling large data volumes. Previous studies have validated the accuracy of these methods using field data such as tree height and diameter at breast height (DBH)<sup>[29–31]</sup>. These validations demonstrated that the

UAV-based measurements are highly reliable and consistent with traditional ground-truth measurements.

The lidR package (version 4.0.0) and the terra package (version 1.7-39) in R (R Core Team, 2023) were instrumental in our individual tree analysis and phenotypic data extraction. The workflow commenced with the generation of Digital Terrain Models (DTMs) and Digital Surface Models (DSMs) from the 3D point cloud data. The Cloth Simulation Filtering (CSF) algorithm<sup>[32]</sup> effectively differentiated ground points to produce DTMs, while DSMs were constructed using a point-to-raster algorithm. Subsequently, the Canopy Height Model (CHM) was derived from the differential between DTM and DSM.

For tree detection and segmentation within the plantation, the `dalponte2016` function from the lidR package was used. This function was configured to detect trees with a minimum height of 2.6 m and a maximum crown radius of 2.5 m. Subsequently, individual tree point clouds were extracted, with a focus on removing noise and irrelevant data points to ensure the purity of the dataset.

Following this initial processing of the point cloud, the attention turned to the in-depth analysis of individual tree data to accurately determine morphological traits. This intricate process was facilitated by a series of custom-developed R functions, designed to handle specific aspects of the point cloud data. The following steps delineate the methodology:

The `calculate_CBH_LAS` function estimated Crown Base Height by identifying the branching point above the trunk, using a combination of height thresholds and standard deviation calculations.

`calculate_canopy_height_outside_trunk` assessed the canopy height outside the trunk's radius, providing a comprehensive view of the tree's overall height.

The `calculate_crown_width` function evaluated the crown's maximum width by calculating the greatest horizontal distance across the canopy.

`calculate_extreme_points_height` determined the maximum height reached by the canopy, offering insights into the vertical extent of the tree's growth.

Finally, `calculate_crown_volume_ellipsoid` computed the crown volume, applying an ellipsoidal model to the canopy dimensions derived from the point cloud.

These functions, fine-tuned for slash pine characteristics, were applied to the individual tree point cloud data. By processing the point cloud through these custom R functions, detailed morphological measurements for each tree were extracted. This included data on tree height, crown base height, crown width, the height at the widest part of the canopy, and overall crown volume. By integrating these measurements, a comprehensive phenotypic profile for each tree was ensured, providing a detailed dataset for subsequent GWAS studies.

### SNP probe

The genotyping aspect of the present study was underpinned by the creation of a novel 51k SNP probe array for slash pine, devised using the genotyping by target sequencing (GBTS) method based on solution hybridization. This development was informed by a comprehensive collection of SNP data and EST-probe sequences from related pine species, as well as the present research groups own extensive slash pine specimen and transcriptome data. Specifically, the following were incorporated:

1. A combination of SNP loci from loblolly pine (*Pinus taeda*), fast-growing slash pine specimens, and transcriptomes of open-pollinated slash pine offspring.

2. EST-probe sequences from loblolly pine.

3. SNPs associated with various traits in loblolly pine.

These components were meticulously mapped against the loblolly pine reference genome, resulting in a robust SNP array encompassing over 51,000 probes. In December 2022, a targeted collection of 1-year-old needle samples from 200 trees, spanning 20 families, across four canopy directions was conducted. These samples were immediately preserved and processed for targeted sequencing using the custom SNP array. The initial dataset comprised hundreds of thousands of SNP loci. After stringent filtering for biallelic genes, minor allele frequency (MAF), and call rate criteria, this dataset was refined to approximately 51,630 high-quality SNP loci. These loci formed the basis of our GWAS analysis, providing a critical genetic insight into the studied population. Details on the development and validation of the 51k SNP probe array are provided in a recently published article<sup>[28]</sup>, which thoroughly describes the SNP array design pipeline, the number of high-quality validated SNPs, and the distribution of SNPs across various functional groups and chromosomes.

### Genetic variation and correlations for morphological traits

In this study, three types of data were integrated: individual morphological phenotypes (CA, CBH, CL, CV, CW, CWH, and H), pedigree information, and SNP data. This combination provided the input data for estimating the genetic parameters of growth traits in slash pine. A Generalized Linear Mixed Model (GLMM) was applied, using the restricted maximum likelihood (REML) method *via* the SommeR package in R software<sup>[33]</sup>. In the GLMM:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \end{bmatrix} = Xm + Z_1b + Z_2f + e \quad (1)$$

$y$  represents the vector of observed growth traits (crown area, tree height, etc.).  $Xm$  captures the fixed effects, including environmental and treatment effects.  $Z_1b$  and  $Z_2f$  represent the random effects matrices for block and family effects, respectively.  $b$  is the vector of random block effects.  $f$  is the vector of random family effects.  $e$  denotes the residual effects.

To estimate the narrow-sense heritability ( $h^2$ ) and genetic correlations ( $r_{gij}$ ) between traits, both pedigree-based and SNP-based methods were utilized. The pedigree-based method inferred heritability based on familial relationships, and the SNP-based method provided a more precise estimation by using individual genetic variants.

The equations for calculating  $h^2$  and  $r_{gij}$  using the pedigree-based method were:

$$h_i^2 = \frac{2.5\sigma_{fi}^2}{\sigma_{fi}^2 + \sigma_{bi}^2 + \sigma_{ei}^2} \quad (2)$$

where  $\sigma_{fi}^2$ ,  $\sigma_{bi}^2$ , and  $\sigma_{ei}^2$  represent the variances due to family, block, and residual effects, respectively.

$$r_{gij} = \frac{\sigma_{fij}}{\sqrt{\sigma_{fi}^2 + \sigma_{fj}^2}} \quad (3)$$

where  $\sigma_{fij}$  is the covariance between traits  $i$  and  $j$  due to family effects.

For the SNP-based method:

$$h_{SNP}^2 = \frac{2.5\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \quad (4)$$

where  $\sigma_g^2$  and  $\sigma_e^2$  are the variances due to genotype and environmental effects, respectively.

$$r_{gij} = \frac{\sum(\beta_{i1}\beta_{i2})/m}{\sqrt{\sum\left(\frac{\beta_{i1}^2}{m}\right)\sum\left(\frac{\beta_{i2}^2}{m}\right)}} \quad (5)$$

with  $\beta_{i1}$  and  $\beta_{i2}$  being the effect sizes of the  $i$ -th SNP on traits 1 and 2, and  $m$  representing the total number of SNPs. The Breeding Value (BV) was calculated as  $BV = G + G \times E$ , where,  $G$  represents the Genetic Effect, encompassing additive genetic effects, and  $G \times E$  indicates the gene-by-environment interaction.

Analysis and visualization of these genetic parameters, particularly in relation to the morphological traits of slash pine, were conducted using the ggplot2 package<sup>[34]</sup> in R. This methodology enabled a detailed understanding of the genetic influences on the growth and development of slash pine.

### GWAS for morphological traits

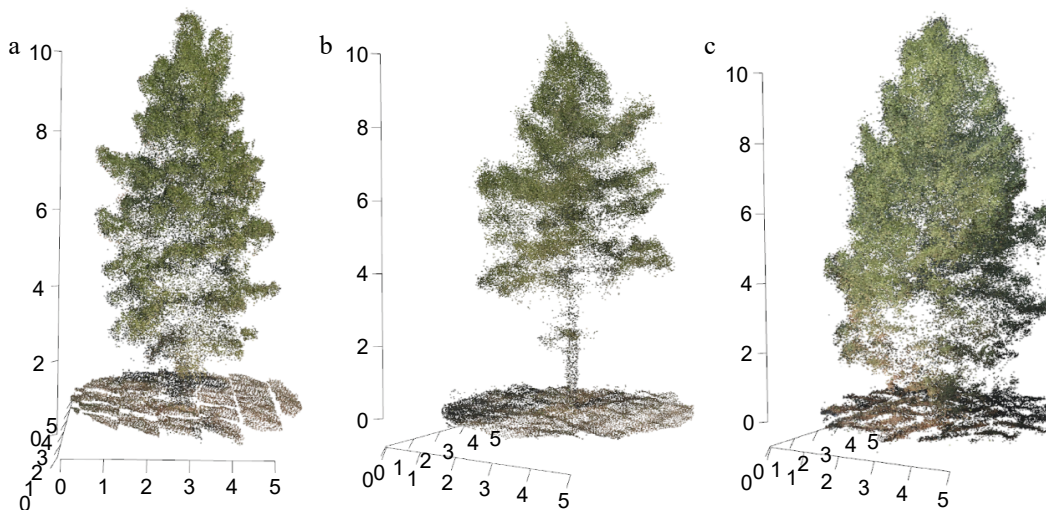
In this study, the statgenGWAS package<sup>[35]</sup> was utilized for GWAS analysis, focusing on the association between individual SNPs and specific morphological traits of slash pine, namely crown area and tree height. Each SNP locus was independently assessed through a generalized least squares (GLS) model, where the correlations with the morphological traits were determined. This approach was critical for revealing genetic factors influencing the growth characteristics of slash pine.

A kinship matrix, calculated using the 'Yang' method from the Genome-wide Complex Trait Analysis software<sup>[36]</sup>, was incorporated as a covariate in the model to account for fine-scale genetic differentiation. This inclusion allowed for a comprehensive analysis of genetic relationships and variations among the slash pine samples.

The GWAS results were visualized using the CMplot R package<sup>[37]</sup>. The kinship heatmap was illustrated with the ggplot2 package. Detection thresholds were established at 4 and 5 for significance levels of  $\alpha < 0.05$  and  $\alpha < 0.01$ , respectively. SNP annotation was conducted using snpEff sv4.5<sup>[38]</sup>. This process included generating a Generic Feature Format (GFF) file of slash pine probe sequences, constructing a specific snpEff database for slash pine, and annotating the SNP array for the 210 analyzed slash pine samples.

## Results

The UAV imaging provided detailed three-dimensional point clouds of slash pine trees, capturing significant variations in canopy architecture and trunk visibility, as illustrated in Fig. 1. The figure shows three distinct structural forms of the trees. In Fig. 1a, the tree features a lower canopy structure, allowing for clear visibility of the trunk. This structural form can be critical for studies focused on trunk accessibility and lower canopy dynamics. Figure 1b depicts a tree with an elevated canopy, characterized by a pronounced lower crown and visible understory. The separation between the crown base and the trunk is more distinct in this structure, which can be useful for analyzing crown-base height and understory interactions. Figure 1c shows a tree with a notably low crown base, resulting in minimal trunk visibility and a denser understory. This

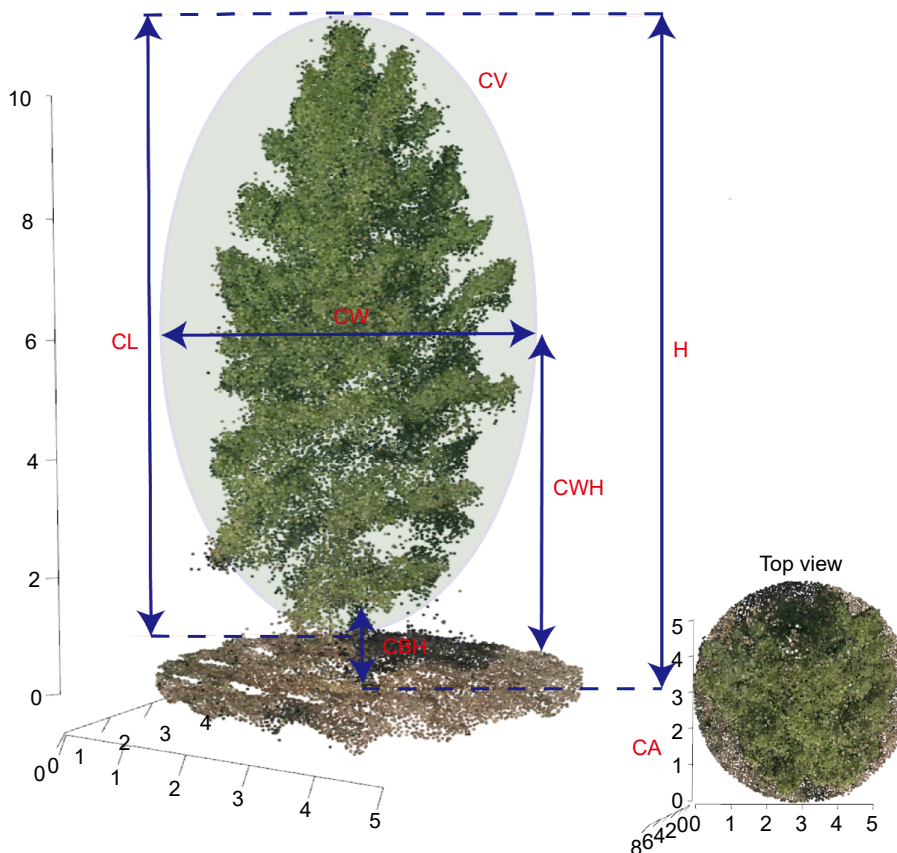


**Fig. 1** Three-dimensional point cloud visualizations of slash pine trees derived from UAV imaging. (a) Tree with a lower canopy structure, (b) elevated canopy structure, and (c) lower crown with understory.

configuration might affect light penetration and understory growth, making it relevant for ecological stratification studies. These phenotypic variations among the trees highlight the diversity within the slash pine plantation. Such diversity is pertinent to understanding adaptive traits, photosynthetic efficiency, and the overall ecological dynamics of the forest canopy. The precise UAV imagery facilitates detailed analysis of

structural complexity and functional dynamics in forest environments. This tripartite representation underscores the phenotypic diversity observed within this slash pine plantation, offering insights into the structural complexity and functional dynamics of forest canopies.

Figure 2 provides a detailed illustration of the methodologies used to measure morphological traits in slash pine,



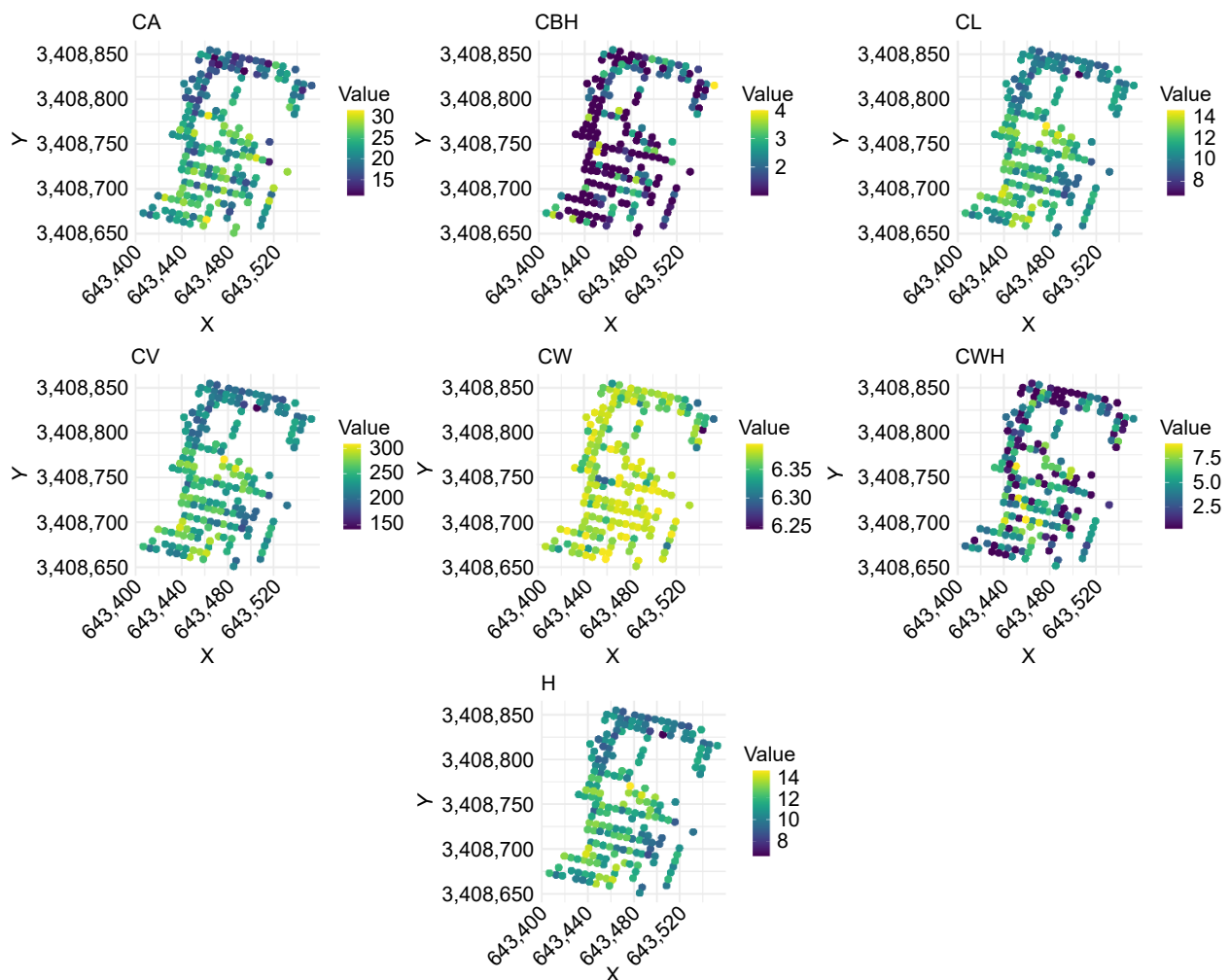
**Fig. 2** Diagrammatic representation of morphological trait measurement methods in slash pine using UAV-derived 3D point clouds. CA: canopy area, CBH: crown base height, CL: crown length, CV: canopy volume, CW: crown width, CWH: crown width height, H: tree height.

utilizing three-dimensional point cloud data acquired from UAV imagery. This figure delineates key morphological features such as canopy area (CA), crown base height (CBH), crown length (CL), canopy volume (CV), crown width (CW), crown width height (CWH), and tree height (H). The top view inset offers a comprehensive perspective of the crown area, allowing for accurate measurement of the canopy's spatial dimensions. These measurements are critical for understanding the physical structure and potential growth patterns of the trees. To evaluate the accuracy of the present model, correlation analysis was conducted between the measured and predicted values of these morphological traits. The results, shown in [Supplemental Fig. S1](#), illustrate the relationship between the actual and predicted values through scatter plots and calculate the coefficient of determination ( $R^2$ ). The  $R^2$  values for each morphological trait were all above 0.9, indicating the high predictive accuracy of the present model. This high level of accuracy demonstrates the reliability of using UAV-derived 3D point clouds for detailed phenotypic analysis in forestry research.

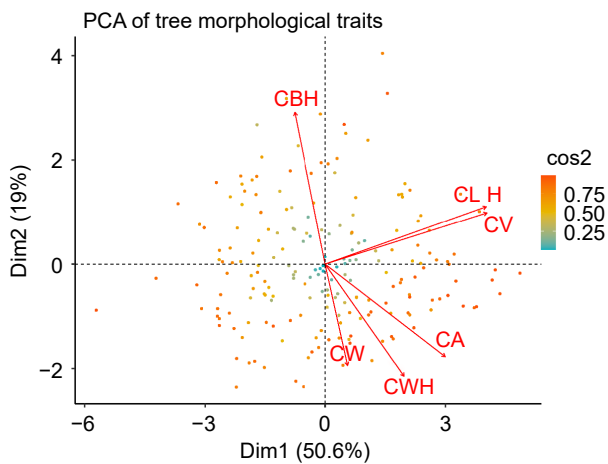
[Figure 3](#) illustrates the spatial distribution of various morphological traits in slash pine as assessed by UAV-derived phenotyping. The collection of scatterplots shows individual phenotypic traits of the trees, with each subplot representing a

different trait: CA, CBH, CL, CV, CW, CWH, H. CW values are predominantly high across the sampled population, suggesting either trait-wide robustness or environmental conditions that favor wide crowns. Conversely, the majority of CBH measurements are significantly low, indicating a commonality in lower crown initiation across trees. Both CL and H show similar ranges in value distribution, suggesting a correlative growth pattern between vertical crown extent and total tree height. These spatial distributions provide valuable insight into the phenotypic variability and potential environmental influences on morphological traits of lodgepole pine. Mapping these traits to geographic coordinates provides a comprehensive understanding of their growth dynamics and structural diversity across the sampled terrain.

Principal component analysis (PCA), shown in [Fig. 4](#), elucidates the underlying structure of variability in morphological traits within a slash pine population. This analysis distills a multidimensional data set into principal components, highlighting the primary axes of variance and covariation among key traits. Dimension 1 (Dim1) encapsulates 50.6% of the trait variability, suggesting a strong gradient of correlated morphological features within the dataset. Dimension 2 (Dim2) accounts for an additional 19%, capturing additional variability



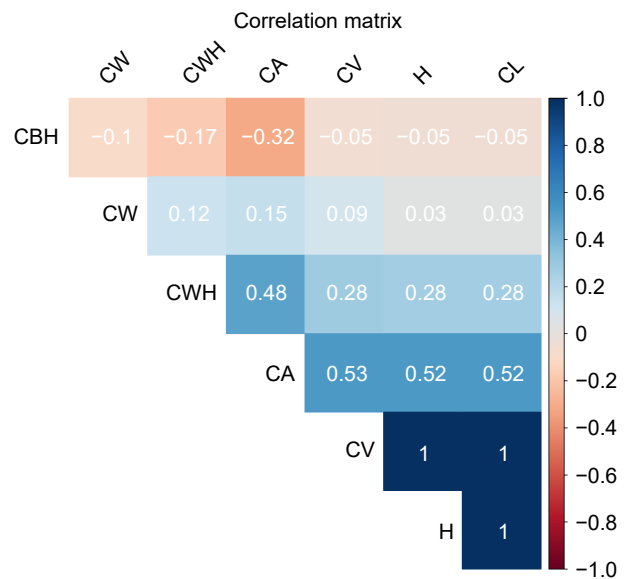
**Fig. 3** Spatial distribution of morphological traits in slash pine assessed by UAV-derived phenotyping. CA: canopy area, CBH: crown base height, CL: crown length, CV: canopy volume, CW: crown width, CWH: crown width height, H: tree height.



**Fig. 4** Principal component analysis of morphological trait covariation in slash pine. Each point represents an individual tree, color-coded according to the  $\cos^2$  value, which indicates the quality of representation of the tree in the PCA plot. Each arrow indicates the direction and strength of the correlation between the respective morphological trait and the principal components.

related to specific morphological traits. Vector orientations in the biplot, depicted as arrows, manifests the loadings of individual traits on the principal components. The proximity and orientation of the vectors corresponding to crown width (CW), canopy area (CA), and crown width height (CWH) suggest a notable positive correlation among these traits. Similarly, the orientation of the vectors for crown length (CL), canopy volume (CV), and tree height (H) also indicate a strong positive correlation, demonstrating that these traits tend to vary in concert across the studied tree individuals. In contrast, crown base height (CBH) is predominantly associated with Dim2 and shows a distinct pattern of variation orthogonal to the main trait cluster, possibly reflecting divergent growth dynamics or adaptive responses to environmental heterogeneity. The color intensity of each data point representing individual trees is calibrated to its  $\cos^2$  value, a measure of how well the data are represented on the principal components. Data points shaded toward the red end of the spectrum denote a higher  $\cos^2$  value, indicating a more reliable representation on the PCA plot, while those shaded toward green indicate a lower correspondence. This analysis highlights the major patterns of trait variation and their interrelationships within the population, providing insight into the genetic and environmental factors influencing these morphological traits.

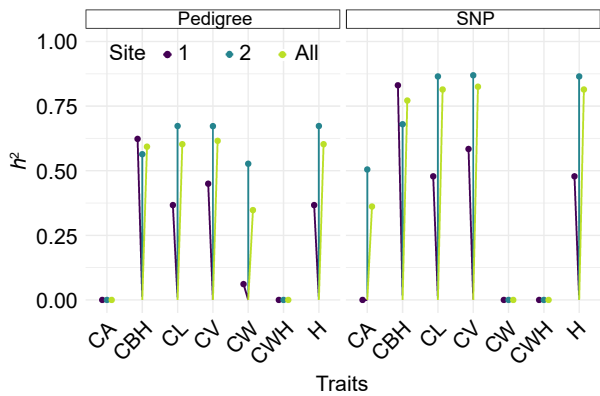
Figure 5 presents a correlation matrix heatmap showing the pairwise correlation coefficients between different morphological traits of slash pine trees. The color intensity and sign of the correlation coefficient values range from  $-1$  to  $+1$ , with blue shades representing positive correlations and red shades representing negative correlations. A correlation value of  $1$  indicates a perfect positive relationship,  $-1$  indicates a perfect negative relationship, and  $0$  indicates no correlation. The heatmap provides an immediate visual interpretation of the strength and direction of relationships between traits. Crown Base Height (CBH) shows slightly negative correlations with all other traits, suggesting that as CBH increases, other traits tend to decrease slightly. However, the strength of these relationships is weak, as the values are close to  $0$ . Canopy area (CA) and crown width



**Fig. 5** Correlation heatmap of morphological traits in slash pine. The color intensity and sign of the correlation coefficient values range from  $-1$  to  $+1$ , where shades of blue represent positive correlations and shades of red indicate negative correlations. A correlation value of  $1$  implies a perfect positive relationship,  $-1$  indicates a perfect negative relationship, and  $0$  denotes no correlation.

height (CWH) show a moderate positive correlation, indicated by a lighter blue, meaning that trees with wider crowns generally have greater width at a given height. The strongest positive correlations are observed between canopy area (CA) and canopy volume (CV), CV and tree height (H), and crown length (CL). This means that increases in canopy area are strongly related to increases in crown volume, and similarly, crown volume is closely related to tree height and crown length. These results highlight the interrelated growth patterns of different morphological traits in lodgepole pine and provide insight into their structural and functional relationships.

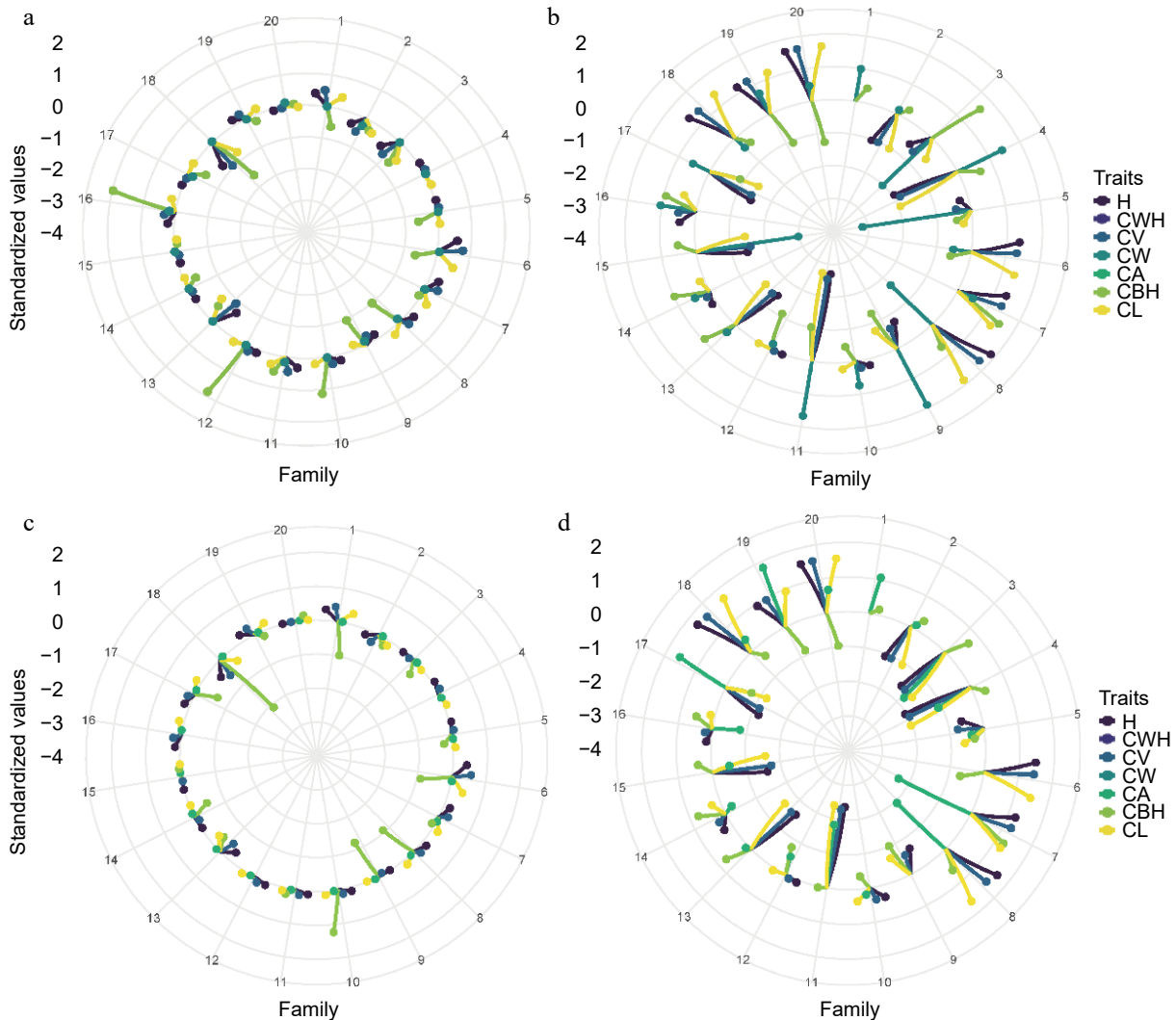
Figure 6 illustrates a lollipop plot showing heritability estimates for several morphological traits in slash pine, comparing heritabilities derived from pedigree-based analyses with those obtained from SNP-based approaches. The vertical lines with colored dots at the end represent the heritability estimates for each trait at different locations. The color of the dots corresponds to the different sites, with dark purple representing site 1, teal representing site 2, and yellow-green representing the pooled data across all sites. The left side of the graph, labeled 'Pedigree', shows the heritability estimates for traits such as CA, CBH, CL, CV, CW, CWH, and H as calculated using traditional pedigree information. The right side, labeled 'SNP', shows the heritability estimates for the same set of traits, but derived from SNP marker data. A comparison between the pedigree and SNP approaches shows that the SNP-based estimates tend to show similar or slightly higher heritabilities for traits such as CBH, CL, and CV. This suggests that SNP markers may capture more of the additive genetic variance for these traits. Conversely, for traits such as CW, pedigree-based estimates appear to be more conservative, which may reflect the different efficiency of each method in capturing the underlying genetic architecture. Notably, the heritability for traits such as CBH, CL, CV, and H are



**Fig. 6** Comparative heritability estimates of slash pine morphological traits using pedigree and SNP data. The vertical lines with colored dots at the end represent the heritability estimates for each trait at different sites. The color of the dots corresponds to different sites, with dark purple indicating Site 1, teal representing Site 2, and yellow green denoting the pooled data across all sites.

relatively high for both methods, indicating that these traits are potentially more influenced by additive genetic factors and could be effectively selected in breeding programs. In contrast, traits such as CWH have lower heritabilities, suggesting a greater influence of environmental factors or non-additive genetic variance. This comparative analysis highlights the strengths and limitations of each approach to estimating heritability, providing valuable insights for slash pine selection and breeding.

Figure 7 consists of four radar plots showing the standardized breeding values of morphological traits for twenty family lines of slash pine. Panels A and C correspond to Site 1, while panels B and D correspond to Site 2. The spokes of the radar charts represent different families, and their radial extent from the center indicates the magnitude of the standardized breeding values for each trait. Panels A and B contrast the pedigree-derived breeding values for Site 1 and Site 2. Families 16 and 12 at Site 1 (Panel A) are notable for their high CBH values, indicating their genetic advantage for this trait. In contrast, at Site 2 (Panel B), families 11 and 9 excel in CW, while families 5 and 15 show negative values for the same trait. Families 18, 19, and 20



**Fig. 7** Comparative analysis of pedigree and SNP-based breeding values for morphological traits in slash pine across two sites. (a) Pedigree-based in Site 1, (b) pedigree-based in Site 2, (c) SNP-based in Site 1, (d) SNP-based in Site 2.



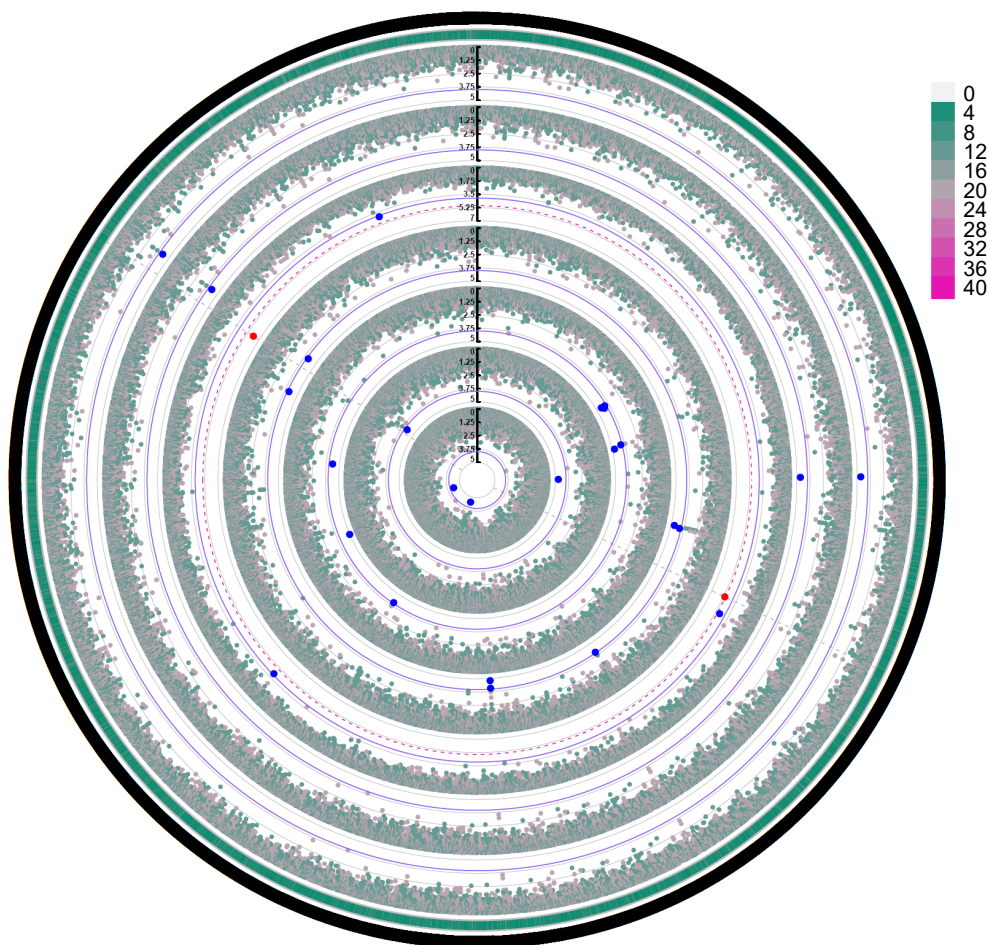
## UAV-based phenotyping in slash pine genomics

at Site 2 show strong breeding values for H, CWH, and CL, indicating their overall genetic superiority for these traits. Panels C and D, which show the SNP-derived breeding values, present a more nuanced picture. At Site 1 (Panel C), there is a general trend toward lower breeding values for most traits, with families 18, 1, 8, and 9 showing negative values for CBH, while family 10 shows a significantly positive value. Conversely, Site 2 (panel D) is more consistent with the pedigree-derived data, with families 18, 19, and 20 maintaining high breeding values. Notably, several families (2, 3, 4, 11, and 15) have negative breeding values for multiple traits, indicating potential genetic limitations. By juxtaposing pedigree and SNP-based analyses, Fig. 7 underscores the complexity of genetic evaluation and the importance of a multifaceted approach to breeding value estimation. The integration of both types of data is critical for a comprehensive understanding of the genetic potential within the breeding program. This comparative analysis highlights the strengths and limitations of each approach and provides valuable insights for the selection and breeding of slash pine.

Figure 8 shows a circular Manhattan plot visualizing the results of a genome-wide association study (GWAS) for seven

growth traits in slash pine throughout 2023. Each concentric circle in the plot corresponds to one of the traits, ordered from the innermost circle to the outermost as follows: CWH, CV, CW, CA, CBH, CL, and H.

The angular coordinates are mapped to the genomic locations of the SNPs, and the radial distance from the center quantifies the significance level of each SNP association. The outermost circle provides a visual representation of SNP density across the genome, with the significance scale detailed in the legend. Blue and red markers highlight SNPs that reach significant ( $\alpha < 0.05$ ) and highly significant ( $\alpha < 0.01$ ) association levels, respectively. The thin purple solid and red dashed circle lines indicate threshold lines for significant associations with different phenotypes at  $\alpha < 0.05$  and  $\alpha < 0.01$  levels, respectively. The gray solid circle lines correspond to the tick marks on the vertical axis. Key SNPs identified for CWH include 'scaffold22007\_82420' and 'scaffold44702\_108111', suggesting a robust genetic association with this trait. Similarly, for CV, SNPs such as 'scaffold114212\_305642' and 'scaffold78970\_13765' have been pinpointed as significant. The SNPs associated with CW, such as 'Gene.216560\_1274' and 'Gene.216560\_1286'



**Fig. 8** GWAS of the performance of the seven growth traits (from inner to outer: CWH, CV, CW, CA, CBH, CL, H) of slash pine in 2023. The outermost circle represents SNP density, and the grade division is indicated in the legend. Blue dots represent SNP loci reaching a significant association level ( $\alpha < 0.05$ ), and red dots represent SNP loci reaching a highly significant association level ( $\alpha < 0.01$ ). The thin purple solid circle lines and red dashed circle lines indicate threshold lines for significant associations with different phenotypes at  $\alpha < 0.05$  and  $\alpha < 0.01$  levels, respectively. The gray solid circle lines correspond to the tick marks of the vertical axis. Only the red significant SNPs identified at the  $\alpha < 0.01$  level are marked by gray dotted lines distributed vertically.

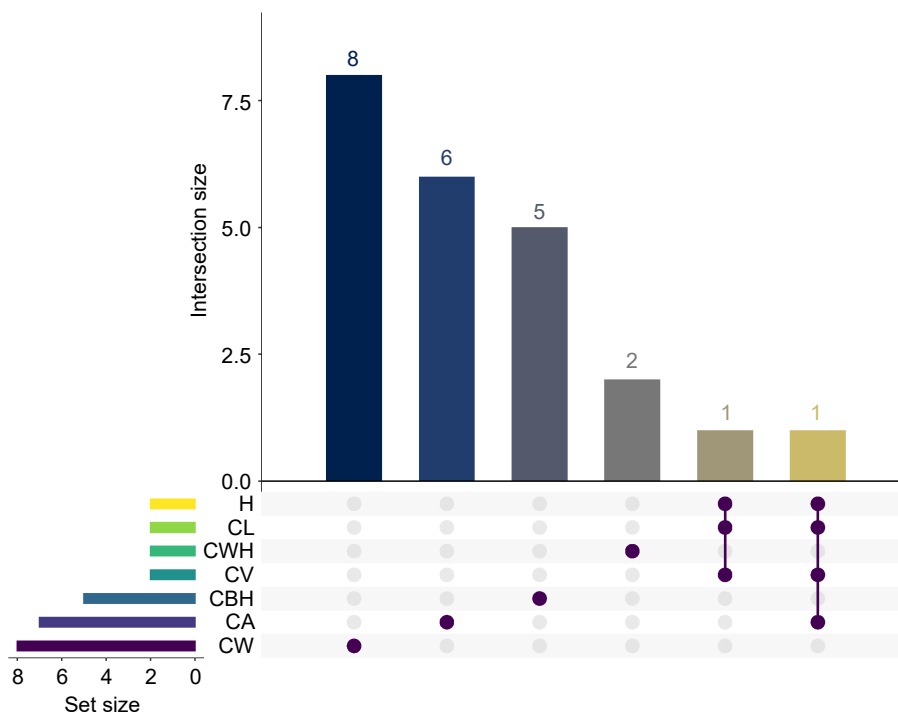
indicate important loci for this trait. The SNPs associated with CA and CBH are also mapped accordingly, with SNPs such as 'scaffold124904\_157642' and 'scaffold136313\_20762' of note. Significant SNPs affecting both CL and H, such as 'scaffold114212\_305642' and 'scaffold78970\_13765', are found on the outer rings, indicating possible pleiotropic effects or genetic linkage. The plot is highlighted with threshold lines for the  $\alpha < 0.05$  and  $\alpha < 0.01$  levels to help distinguish different levels of significance. The identification of these significant SNPs provides a foundation for further genetic research and potential breeding program improvements in slash pine. This analysis highlights the genetic architecture underlying important growth traits and helps to identify loci that could be targeted for selection in breeding programs.

Figure 9 presents an UpSet plot showing the overlap of significant genetic loci for seven growth traits in slash pine. The horizontal bars represent the number of significant loci associated with each trait, with CW having the largest number of unique loci (8), followed by CA with 7, and CWH with 2. The vertical bars represent the size of the intersection, which represents the number of loci shared between trait combinations. The plot highlights an intersection between H and CL, where all identified loci are shared, suggesting a potential genetic correlation. In contrast, the CW trait is distinguished by its unique loci, highlighting a specific genetic influence that is distinct from the other traits. The Viridis color scheme enhances the visualization by distinguishing the sets by the number of loci they share. At the base of the matrix, filled circles indicate a trait's participation in an intersection, with lines connecting loci shared by different traits, delineating their commonality. The

intersection involving traits H, CL, CA, and CV is particularly highlighted, as indicated by the connected solid circles, indicating a shared genetic basis. In Table 1, a substantial number of genes and SNPs have been identified, each associated with major morphological traits in slash pine. For CWH, two significant SNPs were associated with one gene, PITA\_34108. CV was characterized by associations with two genes and two SNPs. CA had a more complex pattern, with seven SNPs associated with six different genes, suggesting a diverse genetic influence. CW was associated with eight SNPs and six genes. For CBH, five SNPs and four genes were identified. Similarly, H and CL were each associated with two genes and two SNPs. These results provide valuable insights into the genetic architecture of important growth traits in slash pine and highlight both unique and common genetic influences that can be targeted for selection in breeding programs.

## Discussion

Understanding the nuances of phenotypic trait variations is fundamental in the realm of genetic breeding research. Such insights are instrumental in optimizing the planning and execution of field trials, as well as in the meticulous selection of germplasm resources, thereby elevating the precision and efficacy of the breeding selection process<sup>[39–41]</sup>. In the present study, the application of UAV technology played a pivotal role in accurately capturing spatial phenotypic traits. This approach, having demonstrated its effectiveness in agricultural contexts for growth estimation and yield prediction, is equally valuable in forestry research<sup>[20,42,43]</sup>. The integration of detailed drone-derived phenotypic data with comprehensive GWAS analysis in



**Fig. 9** The UpSet plot shows the number of significant GWAS SNP loci for seven traits (CWH, CV, CW, CA, CBH, CL, and H). The matrix plot at the bottom of the UpSet plot shows the intersection conditions between different sets, with each row representing a set and each column representing an intersection combination. The brown dots indicate that the intersection combination includes the set in the corresponding row. The bar graph above the matrix plot shows the number of SNP loci in each trait combination. Set Size refers to the number of SNP loci in each individual group being represented.

**Table 1.** Information for 23 candidate genes with 28 significant SNPs that were annotated as growth-related.

Trait	SNP <sup>a</sup>	Gene <sup>b</sup>	Allele <sup>c</sup>	Effect <sup>d</sup>	Gene function
CWH	scaffold22007_82420	PITA_34108	C > T	missense	Neurofilament heavy polypeptide-like
	scaffold44702_108111	–	–	–	–
CV	scaffold114212_305642	PITA_24813	A > T	missense	Ubiquitin family
	scaffold78970_13765	PITA_16597	G > A	downstream_gene	PAR1 protein
CA	scaffold124904_157642	PITA_04981	A > G	synonymous	Fe <sub>2</sub> OG dioxygenase domain-containing protein
	scaffold125333_242817	PITA_17436	A > C	downstream_gene	KH domain
	scaffold165935_108445	PITA_23370	T > C	downstream_gene	–
	scaffold200020_15423	PITA_24554	C > G	missense	Ethylene-responsive element binding factor
	scaffold200020_15465	PITA_24554	G > A	missense	Ethylene-responsive element binding factor
	scaffold71840_181523	–	G > A	intergenic	–
	scaffold78970_13765	PITA_16597	G > A	downstream	PAR1 protein
	Gene.216560_1274	Gene.216560	G > A	missense	Putative DEAD-like helicase
CW	Gene.216560_1286	Gene.216560	A > G	missense	Putative DEAD-like helicase
	Gene.22986_174	Gene.22986	T > C	synonymous	Twinkle homolog protein, chloroplastic/mitochondrial isoform X2
	scaffold10289_45294	PITA_04230	A > C	synonymous	Peptide N-acetyl-beta-D-glucosaminyl asparaginase amidase A
	scaffold103964_286446	–	C > T	intergenic	–
	scaffold230610_125905	PITA_01330	G > A	downstream	Hsp20/alpha crystallin family
	scaffold41862_112441	PITA_00750	G > T	intron	Zinc finger protein CONSTANS-LIKE 6
	scaffold59196_160287	PITA_40818	T > C	intergenic	Arogenate dehydrogenase 2
	scaffold136313_20762	PITA_21834	G > A	missense	Two-component response regulator ORR22
CBH	scaffold139188_163347	PITA_38848	G > C	upstream	Annexin
	scaffold29915_108204	–	T > C	intergenic	–
	scaffold76835_129711	–	A > G	intergenic	–
	super3003_478531	PITA_04633	G > A	downstream	MFS domain-containing protein
H	scaffold114212_305642	PITA_24813	A > C	missense	Ubiquitin family
	scaffold78970_13765	PITA_16597	G > A	downstream	PAR1 protein
CL	scaffold114212_305642	PITA_24813	A > C	missense	Ubiquitin family
	scaffold78970_13765	PITA_16597	G > A	downstream	PAR1 protein

<sup>a</sup>The number after '.' represents the SNP position on the corresponding unigene or scaffold; <sup>b</sup>depending on the location of the SNP, it is annotated with the corresponding unigene or the gene corresponding to the scaffold; <sup>c</sup>the first nucleotide is the reference nucleotide, and the second nucleotide is the nucleotide following the substitution event. <sup>d</sup>the annotation of the SNP, i.e. where the mutation occurred.

our slash pine study offers a novel perspective in understanding the genetic underpinnings of these traits.

### Enhanced extraction of tree morphological traits using UAV-derived point cloud data

The extraction of tree morphological traits like CBH, CL, H, CA, CW, and CWH from UAV-derived point cloud data represents a significant leap in forest phenomics. This method, as seen in the present study, harnesses the potential of advanced UAV technology to capture these complex traits accurately. The same study was found in eucalyptus trees<sup>[44]</sup>, where UAV LIDAR and RGB images were used for competitive extraction of morphological structural traits of eucalyptus trees, such as tree height, diameter at breast height (Dbh), crown, trunk, and branches and they demonstrated that these morphological traits obtained based on a 3D point cloud were better than those measured manually. Liu et al.<sup>[45]</sup> found that point cloud data is outperforms in coniferous tree (Yunnan pine) than other morphological forest species in terms of applicability and accuracy (DBH: Root Mean Square Error (RMSE) = 1.17 cm, Tree Height: RMSE = 0.54 m). Furthermore, the present methodology aligns with the findings of Thiel & Schullius<sup>[46]</sup> in terms of the reliability of UAV-derived point clouds, which were shown to be comparable to lidar point clouds. This suggests that the UAV imagery-based point cloud is not only a valid approach but also a potentially more accessible and cost-effective method for extracting tree structural traits. The comparative ease of data acquisition and processing with UAVs could significantly streamline forest phenomics, making it an attractive tool for researchers and forest managers alike.

This side by side proves that 3D point cloud-based morphological features of conifers can be better extracted. In the present study, it was found that many of the under-branch heights were lower than 1.3 m, so the information of the diameter at breast height (DBH) was not obvious, and the DBH metrics were not extracted.

The present approach utilized a series of custom-developed functions, each tailored to measure specific aspects of slash pine morphology. The calculate\_CBH\_LAS function, for instance, identified the branching point above the trunk using the height thresholds and standard deviation calculations. The calculate\_crown\_width function and calculate\_canopy\_height\_outside\_trunk function provided a comprehensive view of the tree's overall height, mirroring the precision noted in Ghanbari Parmehr & Amati's<sup>[47]</sup> study on forest canopy traits.

Moreover, the calculate\_extreme\_points\_height function determined the maximum height reached by the canopy, offering insights into the vertical growth extent. The calculate\_crown\_volume\_ellipsoid computed the crown volume, applying an ellipsoidal model to the canopy dimensions derived from the point cloud data. This comprehensive approach to morphological assessment not only highlights the efficiency of UAV technology but also its accuracy, as validated against manually annotated field data, with precision levels exceeding 90% (Supplemental Fig. S1).

The integration of these functions with UAV-derived point cloud data overcomes the limitations of traditional field measurement methods, as noted by Dandois & Ellis<sup>[48]</sup>, who discussed the challenges of manual tree trait measurement in

dense forest environments. UAV technology, in contrast, provides a non-invasive, high-throughput, and highly accurate alternative, critical for both growth studies and genetic breeding in trees. While the high accuracy and efficiency of UAV-based methods are clear, challenges remain in data processing and expertise requirements for handling large datasets. Future research should focus on refining data analysis techniques and developing more user-friendly tools to enhance the application of UAV technology in forest phenomics and ecosystem management.

### Genetic variation and GWAS analysis

The present study's heritability estimates and breeding values provide pivotal insights into slash pine's genetic architecture, revealing a spectrum of heritability for various morphological traits. Notably, the findings on tree height heritability surpass those in previous slash pine studies<sup>[49]</sup>. This divergence highlights the comprehensive nature of our phenotypic data and suggests a broader genetic influence on these traits.

The results of this study demonstrate significant disparities in heritability estimates for various morphological traits in slash pine when comparing SNP-based and pedigree-based approaches<sup>[50]</sup>. Specifically, the SNP method exhibits a trend of similar or slightly higher heritability estimates compared to the traditional pedigree approach for traits such as CBH, CL, and CV, suggesting that SNP markers may be more effective in capturing additive genetic variance for these traits, similar results were found by Soleimani et al.<sup>[51]</sup>.

In the context of breeding programs, the relatively high heritability estimates for traits like CBH, CL, CV, and H imply a substantial influence of additive genetic factors<sup>[52]</sup>, making them viable candidates for selection. Conversely, traits with lower heritability, such as CWH, suggest a greater impact of environmental factors or non-additive genetic variance<sup>[53]</sup>.

The breeding value analysis presented in Fig. 7 further underscores the complexity of genetic evaluation. While both pedigree and SNP data highlight genetic advantages for specific families across traits, SNP-derived data introduces nuances, occasionally deviating from pedigree-based results<sup>[54]</sup>, which calls attention to potential genetic limitations. Consequently, a comprehensive understanding of genetic potential within breeding programs necessitates the integration of both pedigree and SNP data<sup>[55]</sup>.

A total of 28 associations were identified by GWAS, including 22 different SNPs localized to 16 different candidate genes. Two of these candidate genes were present at two different mutation sites and were significantly associated with CW and CA traits, respectively, and were annotated as putative DEAD-like helicase and ethylene-responsive ERF, respectively. DEAD-like helicase belongs to the largest subfamily of RNA deconjugating enzymes SF2 and regulates various aspects of plant growth by participating in all biological processes of RNA metabolism in the plant kingdom<sup>[56]</sup>, such as seed development and seedling growth<sup>[57]</sup>, pollen tube formation<sup>[58]</sup>, and post-maturation processing of 23S ribosomes<sup>[59]</sup>. The gene has also been found to play an important role in somatic embryonic processes in *Lobelia*, where it is not only significantly expressed at the syncytial embryo stage but also induced in response to abiotic stress signals such as JA<sup>[60]</sup>. ERF has also been reported to play an important role in biological processes such as agronomic traits of wheat<sup>[61]</sup>, flower and seed development in *Arabidopsis*<sup>[62]</sup>, and floral control in maize<sup>[63]</sup>.

Of interest, the mutant locus scaffold114212\_305642 was detected in CV, H, and CL traits, respectively. While scaffold78970\_13765 was recognized in CV, CA, H, and CL, respectively. The former corresponds to a candidate gene belonging to the ubiquitin family, which modifies the plant proteome to differentiate into appropriate cell and tissue types for maximum adaptation to the environment<sup>[64]</sup>. The latter is annotated as PAR1 protein, which is able to influence photomorphogenesis in plants and can form a network of heterodimeric complexes with the flavonoid sterol-activated HLH transcription factor PRE1 and phytochrome-interacting factor (PIF) to regulate cell elongation and growth<sup>[65]</sup>. The candidate gene significantly associated with CA is also annotated as Fe2OG dioxygenase domain-containing protein, which has been shown to interact with gibberellins and participate in plant growth<sup>[66]</sup>.

The regulation of gene expression at the post-transcriptional level is mainly achieved by proteins that bind RNA recognition motifs (RRMs) and K-homology structural domains (KHs)<sup>[67]</sup>. In this study, the KH domain was recognized in CA, which may be involved in the important transcriptional level regulation during the growth of slash pine. Twinkle homolog protein, chloroplastic/mitochondrial isoform X2 was significantly associated with CW, which is involved in the regulation of maize ectodermal development<sup>[68]</sup>. Peptide N-acetyl-beta-D-glucosaminyl asparaginase amidase A is involved in regulating the activity of macromolecules such as sugars and reducing misfolded proteins in plants<sup>[69]</sup>. Hsp20/alpha crystallin family as an important chaperone protein under heat stress can ensure high productivity of rice<sup>[70]</sup>. The CO (CONSTANS) genes of *Arabidopsis*, rice and barley, which play an important role in the regulation of flowering by photoperiod<sup>[71]</sup>, were also identified as CW significantly related genes. For CBH trait, three candidate genes were identified, namely Two-component response regulator ORR22, Annexin, and MFS domain-containing protein. Among them, Annexin binds to membrane phospholipids in a calcium-dependent manner and is widely involved in the regulation of plant growth and MFS proteins are the largest group of secondary membrane transporter proteins in the cell, and can function as transporter proteins with a variety of substrates in plant growth and development<sup>[72]</sup>. These associations highlight the link between significant SNPs and known functional genes or biological pathways, providing an intuitive understanding of their biological significance. Consequently, the above results indicate that the GWAS in this study has effectively revealed the genetic basis of growth trait variations in slash pine.

### Limitations and future directions

In our study on slash pine morphological traits and genetic breeding, several limitations and future research directions merit consideration.

Firstly, the sample size, while substantial, may not fully represent the genetic diversity within slash pine populations. Expanding the sample size and including individuals from diverse geographical regions could enhance the robustness of the present genetic findings.

Secondly, environmental factors such as soil composition, climate, and inter-tree competition are known to influence tree growth. Integrating environmental data into the analyses

## UAV-based phenotyping in slash pine genomics

would enable a more comprehensive understanding of gene-environment interactions.

Furthermore, while the GWAS analysis identified candidate genes associated with morphological traits, further functional studies and validation experiments are needed to confirm the roles of these genes in slash pine growth and development.

Lastly, future research directions may involve the integration of multi-omics data, including transcriptomics and proteomics, to unravel the intricate regulatory networks involved in tree growth and morphology.

## Conclusions

In summary, the present study delves into the realm of slash pine genetics and morphology, shedding light on key aspects of phenotypic variation and genetic breeding. Through the innovative use of UAV technology, efficient and highly accurate measurements of critical morphological traits were achieved, revolutionizing the way we collect data in forestry research. The present results underscore the importance of morphological traits, such as tree height and crown dimensions, in understanding growth patterns and structural variation in slash pine. UAV technology has proven to be a game changer, providing rapid and accurate data collection compared to traditional labor-intensive methods. In addition, the integration of detailed drone-derived phenotypic data with comprehensive GWAS provided a new perspective on the genetic basis of these traits. The identification of candidate genes associated with specific morphological traits paves the way for more precise genetic improvement and breeding strategies in slash pine.

In addition, the practical implications of the present findings for forest management and breeding programs are profound. Forest inventory processes and monitoring efforts can be greatly enhanced by the ability to quickly and accurately assess morphological traits using UAV technology. This technological advance will allow foresters to make more informed decisions regarding forest health, growth patterns, and resource allocation. In addition, the identification of candidate genes associated with key morphological traits provides valuable markers for selective breeding programs aimed at improving growth rates, disease resistance, and overall productivity of slash pine. Implementation of these genetic findings in breeding programs may lead to the development of superior slash pine cultivars that are better able to withstand environmental stresses and meet the demands of sustainable forestry practices.

While the present study represents a significant leap forward, it is important to acknowledge the limitations, such as sample size constraints and the need for further validation of candidate genes. These limitations guide us toward future research directions that include expanding sample diversity, refining data processing methods, considering environmental factors, and integrating multi-omics data for a more comprehensive understanding.

## Author contributions

The authors confirm contribution to the paper as follows: conducting experiment and writing manuscript: Yan R, Dong Y; study design and writing manuscript: Li Y; supporting data collection and field experiments: Diao S; supervising experiments and performing revisions of the manuscript: Wu C, Xu C,

Luan Q. All authors reviewed the results and approved the final version of the manuscript.

## Data availability

The data mentioned in this paper are available upon request from the corresponding author. The codes and examples of point cloud data used in this study have been backed up to a cloud drive, which can be accessed at: <https://pan.quark.cn/s/1dbde2d1f7fc>.

## Acknowledgments

This work was funded by the Biological Breeding-Major Projects (Grant no. 2023ZD040580105) and the Fundamental Research Funds of CAF (Grant no. CAFYBB2022QA001).

## Conflict of interest

The authors declare that they have no conflict of interest.

**Supplementary Information** accompanies this paper at (<https://www.maxapress.com/article/doi/10.48130/forres-0024-0022>)

## Dates

Received 9 March 2024; Revised 24 June 2024; Accepted 10 July 2024; Published online 25 July 2024

## References

- Valladares F. 1999. Architecture, ecology and evolution of plant crowns. *Handbook of Functional Plant Ecology* 121–94
- Niinemets Ü. 2010. A review of light interception in plant stands from leaf to canopy in different plant functional types and in species with varying shade tolerance. *Ecological Research* 25:693–714
- Thompson ID, Okabe K, Tylianakis JM, Kumar P, Brockerhoff EG, et al. 2011. Forest biodiversity and the delivery of ecosystem goods and services: translating science into policy. *BioScience* 61:972–81
- Zieschank V, Junker RR. 2023. Digital whole-community phenotyping: tracking morphological and physiological responses of plant communities to environmental changes in the field. *Frontiers in Plant Science* 14:1141554
- Yang J, Swenson NG. 2023. Height and crown allometries and their relationship with functional traits: an example from a subtropical wet forest. *Ecology and Evolution* 13:e9804
- Estopa RA, Paludeto JGZ, Müller BSF, de Oliveira RA, Azevedo CF, et al. 2023. Genomic prediction of growth and wood quality traits in *Eucalyptus benthamii* using different genomic models and variable SNP genotyping density. *New Forests* 54:343–62
- Tao S, Xie Y, Luo J, Wang J, Zhang L, et al. 2023. Superior clone selection in a *Eucalyptus* trial using forest phenotyping technology via UAV-based DAP point clouds and multispectral images. *Remote Sensing* 15:899
- Bian L, Zhang H, Ge Y, Čepl J, Stejskal J, et al. 2022. Closing the gap between phenotyping and genotyping: review of advanced, image-based phenotyping technologies in forestry. *Annals of Forest Science* 79:22
- Wei M, Li H, Wang Q, Liu R, Yang L, et al. 2023. Genome-wide identification and expression profiling of B3 transcription factor genes in *Populus alba* × *Populus glandulosa*. *Frontiers in Plant Science* 14:1193065

10. Ray D, Berlin M, Alia R, Sanchez L, Hynynen J, et al. 2022. Transformative changes in tree breeding for resilient forest restoration. *Frontiers in Forests and Global Change* 5:1005761
11. Thapa S, Zhu F, Walia H, Yu H, Ge Y. 2018. A novel LiDAR-based instrument for high-throughput, 3D measurement of morphological traits in maize and sorghum. *Sensors* 18:1187
12. Alani AM, Giannakis I, Zou L, Lantini L, Tosti F. 2020. Reverse-time migration for evaluating the internal structure of tree-trunks using ground-penetrating radar. *NDT & E International* 115:102294
13. Li H, Yin S, Wang L, Xu N, Liu L. 2022. Transcription factor PagLBD21 functions as a repressor of secondary xylem development in *Populus*. *Forestry Research* 2:19
14. Mori AS, Suzuki KF, Hori M, Kadoya T, Okano K, et al. 2023. Perspective: sustainability challenges, opportunities and solutions for long-term ecosystem observations. *Philosophical Transactions of the Royal Society B* 378:20220192
15. Ehrlich-Sommer F, Hoenigsberger F, Gollob C, Nothdurft A, Stampfer K, et al. 2024. Sensors for Digital Transformation in Smart Forestry. *Sensors* 24:798
16. Brockerhoff EG, Corley JC, Jactel H, Miller DR, Rabaglia RJ, et al. 2023. Monitoring and surveillance of forest insects. In *Forest Entomology and Pathology*, eds Allison D, Paine J, Slippers TD, Wingfield B, Entomology MJ. Cham: Springer International Publishing pp. 669–705. [https://doi.org/10.1007/978-3-031-11553-0\\_19](https://doi.org/10.1007/978-3-031-11553-0_19)
17. Ge Y, Atefi A, Zhang H, Miao C, Ramamurthy RK, et al. 2019. High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel. *Plant Methods* 15:66
18. Iglhaut J, Cabo C, Puliti S, Piermattei L, O'Connor J, et al. 2019. Structure from motion photogrammetry in forestry: a review. *Current Forestry Reports* 5:155–68
19. Wallace L, Lucieer A, Malenovsky Z, Turner D, Vopěnka P. 2016. Assessment of forest structure using two UAV techniques: a comparison of airborne laser scanning and structure from motion (SfM) point clouds. *Forests* 7:62
20. Guimarães N, Pádua L, Marques P, Silva N, Peres E, et al. 2020. Forestry remote sensing from unmanned aerial vehicles: a review focusing on the data, processing and potentialities. *Remote Sensing* 12:1046
21. Quan M, Liu X, Du Q, Xiao L, Lu W, et al. 2021. Genome-wide association studies reveal the coordinated regulatory networks underlying photosynthesis and wood formation in *Populus*. *Journal of Experimental Botany* 72:5372–89
22. Ashwath MN, Lavale SA, Santhoshkumar AV, Mohapatra SR, Bhardwaj A, et al. 2023. Genome-wide association studies: an intuitive solution for SNP identification and gene mapping in trees. *Functional & Integrative Genomics* 23:297
23. Yang A, Ding X, Feng Y, Zhao R, Ye J. 2023. Genetic diversity and genome-wide association analysis of pine wood nematode populations in different regions of China. *Frontiers in Plant Science* 14:1183772
24. Cappa EP, Chen C, Klutsch JG, Sebastian-Azcona J, Ratcliffe B, et al. 2022. Multiple-trait analyses improved the accuracy of genomic prediction and the power of genome-wide association of productivity and climate change-adaptive traits in lodgepole pine. *BMC Genomics* 23:536
25. Li Y, Yang X, Tong L, Wang L, Xue L, et al. 2023. Phenomic selection in slash pine multi-temporally using UAV-multispectral imagery. *Frontiers in Plant Science* 14:1156430
26. Moreira FF, Oliveira HR, Volenec JJ, Rainey KM, Brito LF. 2020. Integrating high-throughput phenotyping and statistical genomic methods to genetically improve longitudinal traits in crops. *Frontiers in Plant Science* 11:681
27. White JW, Andrade-Sanchez P, Gore MA, Bronson KF, Coffelt TA, et al. 2012. Field-based phenomics for plant genetics research. *Field Crops Research* 133:101–12
28. Diao S, Ding X, Luan Q, Chen Z, Wu H, et al. 2024. Development of 51 K liquid-phased probe array for Loblolly and Slash pines and its application to GWAS of Slash pine breeding population. *Industrial Crops and Products* 216:118777
29. Song Z, Tomasetto F, Niu X, Yan W, Jiang J, et al. 2022. Enabling breeding selection for biomass in slash pine Using UAV-based imaging. *Plant Phenomics* 2022:9783785
30. Niu X, Song Z, Xu C, Wu H, Luan Q, et al. 2023. Prediction of needle physiological traits using UAV imagery for breeding selection of slash pine. *Plant Phenomics* 5:0028
31. Tao X, Li Y, Yan W, Wang M, Tan Z, et al. 2021. Heritable variation in tree growth and needle vegetation indices of slash pine (*Pinus elliottii*) using unmanned aerial vehicles (UAVs). *Industrial Crops and Products* 173:114073
32. Zhang W, Qi J, Wan P, Wang H, Xie D, et al. 2016. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. *Remote Sensing* 8:501
33. Covarrubias-Pazarán G. 2016. Genome-assisted prediction of quantitative traits using the R package *sommer*. *PLoS One* 11:e0156744
34. Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Cham: Springer. xvi, 260 pp. <https://doi.org/10.1007/978-3-319-24277-4>
35. van Rossum BJ, Kruijer W, van Eeuwijk F, Boer M, Malosetti M, et al. 2020. Package 'statgenGWAS'. R package version 1
36. Yang J, Lee SH, Wray NR, Goddard ME, Visscher PM. 2016. GCTA-GREML accounts for linkage disequilibrium when estimating genetic variance from genome-wide SNPs. *Proceedings of the National Academy of Sciences of the United States of America* 113:E4579–E4580
37. Yin L. 2023. *CMplot: circle manhattan plot*. R package version 4.4.1. <https://CRAN.R-project.org/package=CMplot>
38. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3. *Fly* 6:80–92
39. Reynolds M, Chapman S, Crespo-Herrera L, Molero G, Mondal S, et al. 2020. Breeder friendly phenotyping. *Plant Science* 295:110396
40. Strauss S, Lande R, Namkoong G. 1992. Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* 22:1050–61
41. Iwata H, Hayashi T, Tsumura Y. 2011. Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genetics & Genomes* 7:747–58
42. Torresan C, Berton A, Carotenuto F, Di Gennaro SF, Gioli B, et al. 2017. Forestry applications of UAVs in Europe: a review. *International Journal of Remote Sensing* 38:2427–47
43. Liu Q, Li S, Li Z, Fu L, Hu K. 2017. Review on the applications of UAV-based LiDAR and photogrammetry in forestry. *Scientia Silvae Sinicae* 53:134–48
44. Liao L, Cao L, Xie Y, Luo J, Wang G. 2022. Phenotypic traits extraction and genetic characteristics assessment of *Eucalyptus* Trials based on UAV-borne LiDAR and RGB images. *Remote Sensing* 14:765
45. Liu G, Wang J, Dong P, Chen Y, Liu Z. 2018. Estimating individual tree height and diameter at breast height (DBH) from terrestrial laser scanning (TLS) data at plot level. *Forests* 9:398
46. Thiel C, Schmuilius C. 2017. Comparison of UAV photograph-based and airborne lidar-based point clouds over forest from a forestry application perspective. *International Journal of Remote Sensing* 38:2411–26
47. Ghanbari Parmehr E, Amati M. 2021. Individual tree canopy parameters estimation using UAV-based photogrammetric and LiDAR point clouds in an urban park. *Remote Sensing* 13:2062
48. Dandois JP, Ellis EC. 2013. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment* 136:259–76

## UAV-based phenotyping in slash pine genomics

49. Coble DW, Lee YJ. 2011. A mixed-effects height—diameter model for individual loblolly and slash pine trees in east Texas. *Southern Journal of Applied Forestry* 35:12–17
50. Beaulieu J, Lenz P, Bousquet J. 2022. Metadata analysis indicates biased estimation of genetic parameters and gains using conventional pedigree information instead of genomic-based approaches in tree breeding. *Scientific Reports* 12:3933
51. Soleimani VD, Baum BR, Johnson DA. 2007. Analysis of genetic diversity in barley cultivars reveals incongruence between S-SAP, SNP and pedigree data. *Genetic Resources and Crop Evolution* 54:83–97
52. Berlin M, Jansson G, Högberg KA, Helmersson A. 2019. Analysis of non-additive genetic effects in Norway spruce. *Tree Genetics & Genomes* 15:42
53. Berguson WE, McMahon BG, Riemenschneider DE. 2017. Additive and non-additive genetic variances for tree growth in several hybrid poplar populations and implications regarding breeding strategy. *Silvae Genetica* 66:33–39
54. Xia C, Amador C, Huffman J, Trochet H, Campbell A, et al. 2016. Pedigree-and SNP-associated genetics and recent environment are the major contributors to anthropometric and cardiometabolic trait variation. *PLoS Genetics* 12:e1005804
55. Misztal I, Lourenco D, Legarra A. 2020. Current status of genomic evaluation. *Journal of Animal Science* 98:skaa101
56. de la Cruz J, Kressler D, Linder P. 1999. Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends in Biochemical Sciences* 24:192–98
57. Kanai M, Hayashi M, Kondo M, Nishimura M. 2013. The plastidic DEAD-box RNA helicase 22, H53, is essential for plastid functions both in seed development and in seedling growth. *Plant and Cell Physiology* 54:1431–40
58. Shimizu KK, Ito T, Ishiguro S, Okada K. 2008. MAA3 (MAGATAMA3) helicase gene is required for female gametophyte development and pollen tube guidance in *Arabidopsis thaliana*. *Plant and Cell Physiology* 49:1478–83
59. Nishimura K, Ashida H, Ogawa T, Yokota A. 2010. A DEAD box protein is required for formation of a hidden break in *Arabidopsis* chloroplast 23S rRNA. *The Plant Journal* 63:766–77
60. Xu X, Chen X, Shen X, Chen R, Zhu C, et al. 2021. Genome-wide identification and characterization of DEAD-box helicase family associated with early somatic embryogenesis in *Dimocarpus longan* Lour. *Journal of Plant Physiology* 258–259:153364
61. Zhang L, Liu P, Wu J, Qiao L, Zhao G, et al. 2020. Identification of a novel ERF gene, *TaERF8*, associated with plant height and yield in wheat. *BMC Plant Biology* 20:263
62. Jofuku KD, Den Boer BG, Van Montagu M, Okamoto JK. 1994. Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *The Plant Cell* 6:1211–25
63. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences of the United States of America* 104:11376–81
64. Stone SL, Callis J. 2007. Ubiquitin ligases mediate growth and development by promoting protein death. *Current Opinion in Plant Biology* 10:624–32
65. Galstyan A, Bou-Torrent J, Roig-Villanova I, Martínez-García JF. 2012. A dual mechanism controls nuclear localization in the atypical basic-helix-loop-helix protein PAR1 of *Arabidopsis thaliana*. *Molecular Plant* 5:669–77
66. Sun J, Zhang X, Fu C, Ahmad N, Zhao C, et al. 2023. Genome-wide identification and expression analysis of GA20ox and GA3ox genes during pod development in peanut. *PeerJ* 11:e16279
67. Chen T, Cui P, Chen H, Ali S, Zhang S, Xiong L. 2013. A KH-domain RNA-binding protein interacts with FIERY2/CTD phosphatase-like 1 and splicing factors and is important for pre-mRNA splicing in *Arabidopsis*. *PLoS Genetics* 9:e1003875
68. Zhang J, Zhang F, Tian L, Ding Y, Qi J, et al. 2022. Molecular mapping of quantitative trait loci for 3 husk traits using genotyping by sequencing in maize (*Zea mays* L.). *G3 Genes[Genomes] Genetics* 12:jkac198
69. Karamanos Y. 2013. Endo-N-acetyl- $\beta$ -D-glucosaminidases and peptide-N4-(N-acetyl- $\beta$ -D-glucosaminyl) asparagine amidases: more than just tools. *Advances in Biochemistry* 1:81–99
70. Kilasi NL, Singh J, Vallejos CE, Ye C, Jagadish SVK, et al. 2018. Heat stress tolerance in rice (*Oryza sativa* L.): identification of quantitative trait loci and candidate genes for seedling growth under heat stress. *Frontiers in Plant Science* 9:1578
71. Griffiths S, Dunford RP, Coupland G, Laurie DA. 2003. The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*. *Plant Physiology* 131:1855–67
72. Wang C, Huang W, Ying Y, Li S, Secco D, et al. 2012. Functional characterization of the rice *SPX-MFS* family reveals a key role of *OsSPX-MFS1* in controlling phosphate homeostasis in leaves. *New Phytologist* 196:139–48



Copyright: © 2024 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.