Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2024) **22**, pp. 2578–2592 doi: 10.1111/pbi.14370

Evolutionary relationship of moso bamboo forms and a multihormone regulatory cascade involving culm shape variation

Yan Liu[1,2], Chenglei Zhu[1,2], Xianghua Yue[1,3], Zeming Lin[1,2], Hui Li[1,2], Xiaolin Di[1,2], Jiangfei Wang[1,2] and Zhimin Gao[1,2,*]

[1]*Key Laboratory of State Forestry and Grassland Administration/Beijing on Bamboo and Rattan Science and Technology, Beijing, China*
[2]*Institute of Gene Science and Industrialization for Bamboo and Rattan Resources, International Center for Bamboo and Rattan, Beijing, China*
[3]*Sanya Research Base, International Centre for Bamboo and Rattan, Sanya, China*

Received 20 January 2024;
revised 1 April 2024;
accepted 22 April 2024.
**Correspondence* (Tel: 010-84789802, Fax: 010-84789803; email gaozhimin@icbr.ac.cn)

Keywords: *Phyllostachys edulis*, variation, whole-genome resequencing, transcriptome, plant hormone.

Summary

Moso bamboo (*Phyllostachys edulis*) known as Mao Zhu (MZ) in Chinese exhibits various forms with distinct morphological characteristics. However, the evolutionary relationship among MZ forms and the mechanisms of culm shape variation are still lacking. Here, the main differences among MZ forms were identified as culm shape variation, which were confirmed by analysing MZ forms (799 bamboo culms) and MZ (458 bamboo culms) populations. To unravel the genetic basis underlying the morphological variations, 20 MZ forms were subjected to whole-genome resequencing. Further analysis yielded 3 230 107 high-quality SNPs and uncovered low genetic diversity and high genotype heterozygosity associated with MZ forms' formation. By integrating the SNP data of 427 MZ individuals representing 15 geographic regions, the origins of eight MZ forms were successfully traced using the phylogenetic tree and the identified common heterozygous loci. Meanwhile, transcriptomic analysis was performed using shoots from MZ and its two forms with culm shape variation. The results, combined with genomic analyses, demonstrated that hormone signalling related genes played crucial roles in culm variation. Co-expression network analysis uncovered genes associated with multiple plant hormone signal transduction, especially auxin and cytokinin were involved in culm shape variation. Furthermore, the regulatory relationships of a specific transcription factor and their target genes associated with auxin and ethylene signalling were validated by yeast one-hybrid, electrophoretic mobility shift assays, and dual-luciferase reporter. Overall, this study provides important insights into the culm shape variation formation in bamboo, which facilitates to breed new varieties with novel culms.

Introduction

Bamboos, belonging to the grass family Poaceae, subfamily Bambusoideae, comprise 75 genera and approximately 1642 bamboo species, making them important non-timber forest resources globally (Ramakrishnan *et al*., 2020). *Phyllostachys edulis* (Mao Zhu in Chinese, MZ) is one of the most representative and significant bamboo species due to its rapid growth, wide distribution, and exceptional mechanical properties (Chen *et al*., 2022; Peng *et al*., 2013). It is adaptable to various cultivation conditions and covers around 69.78% of the total bamboo-growing area in China (Feng and Li, 2023; Ramakrishnan *et al*., 2020). Through long-term adaptation to different habitats, natural selection, and artificial selection, various forms of MZ with distinct morphological characteristics have been produced over time. For instance, *P. edulis* f. *exaurita* (Ma Yi Zhu in Chinese, MYZ) and *P. edulis* f. *obtusangula* (Fang Mao Zhu in Chinese, FMZ) were first discovered in the wild MZ populations in Fujian Province and Hunan, respectively, and later introduced and cultivated elsewhere (Chen, 2013; Wang, 1984). Previous studies have tried to elucidate the evolutionary relationship between MZ and its forms using molecular markers such as AFLP (Lin *et al*., 2009), SSR (Zhao *et al*., 2015), and SNP (Li, 2020). However, due to the close relationship among these forms and the limited number of markers utilized, the results have not provided conclusive answers to this issue.

The bamboo culm, composed of nodes and internodes, is the most important and utilized part of bamboo. Different forms of MZ exhibit distinct morphological characteristics related to culm features, such as culm shape and culm wall thickness. For example, *P. edulis* f. 'Kikko-chiku' (Gui Jia zhu in Chinese, GJ) and *P. edulis* f. *tubaeformis* (Sheng Yin zhu in Chinese, SY) are characterized by specific culm shape and dwarfism (Zeng *et al*., 2016), while *P. edulis* f. *pachyloen* (Hou Bi mao zhu in Chinese, HB) is renowned for its thickness culm wall (Yang *et al*., 1997). Due to the absence of an interfascicular cambium, the culm shape is primarily determined by the diameter and length of the internodes of bamboo shoot. Previous studies have shown that plant hormones including auxin, cytokinin (CK), gibberellin (GA), abscisic acid (ABA), brassinolide (BR), jasmonic acid (JA), and ethylene (ETH) play crucial roles in controlling cell proliferation and elongation (Bunsick *et al*., 2021; Depuydt and Hardtke, 2011). These hormones often interact and regulate each other (Ross *et al*., 2011). For instance, CK commonly acts in

concert with auxin to regulate cell division and differentiation (Kieber and Schaller, 2018). GA and BR are essential for cell elongation (Daviere and Achard, 2013; Durbak *et al*., 2012), whereas ETH and ABA have been shown to negatively regulate internode elongation (Ross *et al*., 2011).

Studies on MZ forms have revealed plant hormone signalling also plays a significant role in the formation of distinct morphological characteristics (Chen *et al*., 2022; Li *et al*., 2021, 2023). For instance, the pathway of plant hormone signal transduction plays significant role in culm thickening process of HB during shoot development (Li *et al*., 2022). It is suggested that the dwarfing of SY may be related to the phytohormone signalling such as auxin, BR, GA, JA, and ABA (Wang *et al*., 2021). However, further research is still needed to fully understand the mechanisms responsible for the formation of different MZ forms. In this study, a systematic analysis of the morphological characteristics of various MZ forms was conducted to discern their primary differences. The genomic variation dataset obtained from the 20 MZ forms was then integrated with genomic data from 427 MZ individuals collected from 15 different geographic regions to unravel the origins of the MZ forms and investigate their evolutionary relationships. Furthermore, shoot transcriptomic data from MZ and its two representative forms with culm shape variation (GJ and SY) were used to tell the differentially expressed genes associated with culm shape variations. These findings from this study contribute to our knowledge of bamboo diversity and provide valuable information for bamboo utilization and breeding.

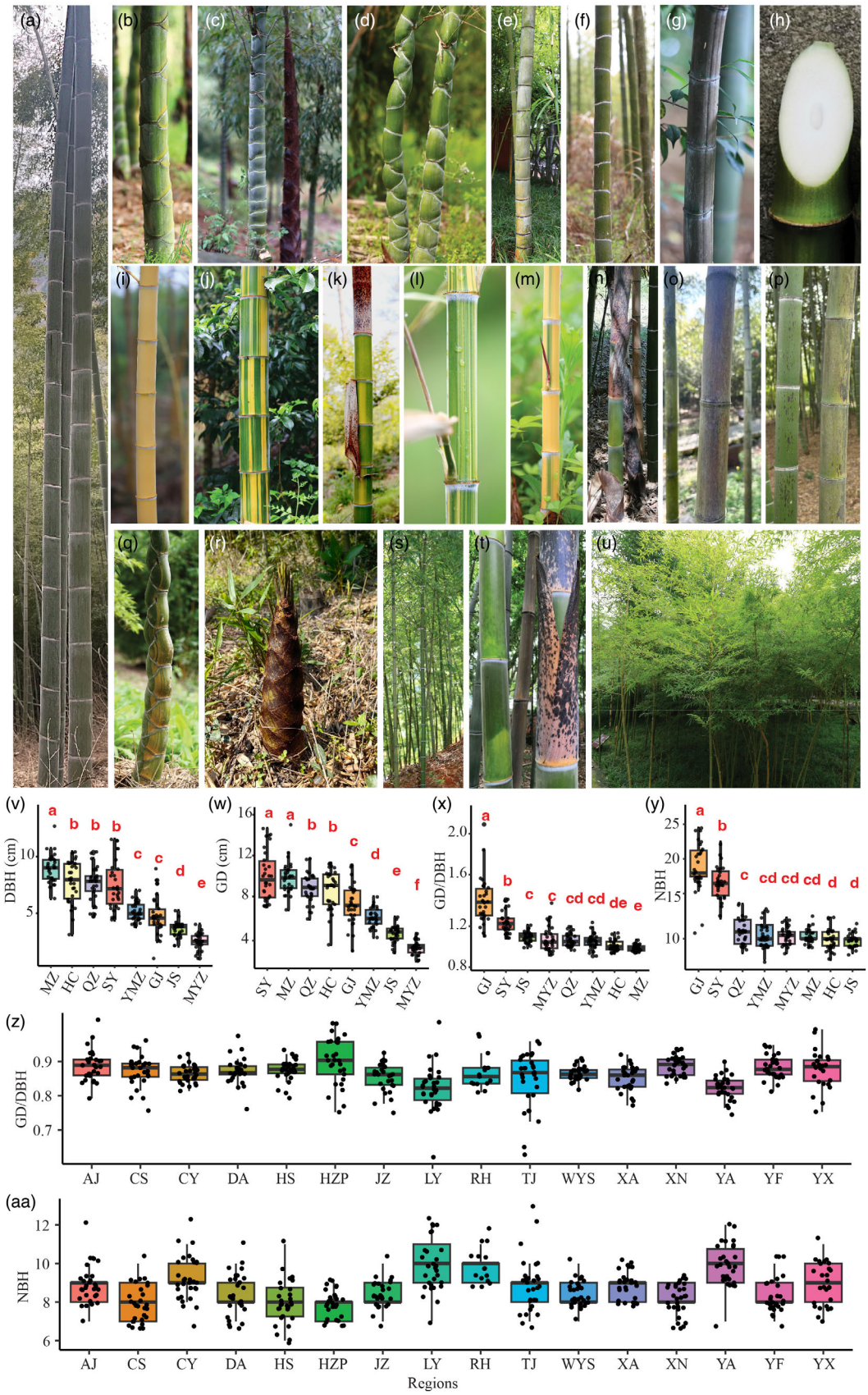
Results

Morphological analysis of *P. Edulis* forms

There are vast morphological differences among MZ (Figure 1a) and its forms (Figure 1b–u), which fall into three main categories of variation, including culm shape (Figure 1b–h), culm colour (Figure 1i–p), and other characteristic variations (Figure 1q–u) (Ma *et al*., 2014). Culm shape variation group contained seven members, including *P. edulis* f. *abbreviata* (Die Mao Zhu in Chinese, DMZ), SY, GJ, *P. edulis* f. *ventricosa* (Fo Du mao zhu in Chinese, FD), *P. edulis* f. *obliquinoda* (Qiang Zhu in Chinese, QZ), FMZ, and HB (Figure 1b–h). Among which, the nodes direction of DMZ, SY, GJ, and QZ are not parallel; the internodes of SY, GJ, and FD are inflation or deflation; the culm shape of HB and FMZ are slightly square. Culm colour variation group contained eight members, including *P. edulis* f. *holochrysa* (Huang Pi mao zhu in Chinese, HP), *P. edulis* f. *huamozhu* (Huang Pi Hua mao zhu in Chinese, HPH), *P. edulis* f. *luteosulcata* (Huang Cao mao zhu in Chinese, HC), *P. edulis* f. *nabeshimana* (Lv Pi Hua mao zhu in Chinese, LPH), *P. edulis* f. *bicolor* (Lv Cao mao zhu in Chinese, LC), *P. edulis* f. *venusta* (Hua Gan Jin Si mao zhu in Chinese, HGJS), *P. edulis* f. *purpureoculmis* (An Ji Zi mao zhu in Chinese, AJZ), and *P. edulis* f. *porphyrosticta* (Ban Mao Zhu in Chinese, BMZ) (Figure 1i–p). Both AJZ and BMZ have light purple spots on the culm, but AJZ covers the entire internodes, while BMZ does not. The remains with other characteristic variation were classified as a third group. Specifically, *P. edulis* 'Mira' (Hua Gui Zhu as Chinese, HGZ) (Figure 1q) has the similar culm shape variation as GJ, but the culm also has yellow and green stripes.

To reveal the morphological difference between MZ and various MZ forms, an investigation and analysis of parameters including diameter at breast height (DBH), ground diameter (GD), and the number of nodes at breast height (NBH) were conducted for MZ and its seven representative forms grown in China Bamboo Expo Park (Zhejiang Province). The results showed that MZ exhibited the largest average DBH, measuring up to 9.24 ± 1.27 cm, followed by HC, and the smallest average DBH was found in MYZ, measuring merely 2.80 ± 0.68 cm (Figure 1v). Consistent with these results, SY had the biggest average GD (10.30 ± 2.42 cm), which was almost identical to that of MZ (10.16 ± 1.57 cm), but approximately three times larger than that of MYZ (3.28 ± 0.69 cm) (Figure 1w). The ratio of GD to DBH was calculated to further reveal the culm shape difference between different MZ forms. The smallest average ratio was observed in MZ (1.10 ± 0.03), which was significantly smaller than the others except HC (1.12 ± 0.05) (Figure 1x). The largest ratio was found in GJ (1.56 ± 0.23), followed by SY (1.34 ± 0.10). Similarly, the average NBH of MZ was 10.30 ± 0.94, which was not different significantly from the others except GJ (20.30 ± 3.55) and SY (18.10 ± 2.31) (Figure 1y). This indicated that the internode length below breast height was drastically shorter in GJ and SY, compared with MZ. Simultaneously, the morphological difference analysis results from Jiangxi Province (349 bamboo culms) and Anhui Province (210 bamboo culms, Table S1 and Figure S1a–h) were highly consistent with that from Zhejiang Province. To determine whether culm morphological exhibited similar characteristics in other MZ populations in China, the GD, DBH, and NBH of 458 MZ culms from 16 different MZ populations (with a range of 15 to 30 culms per population) in 16 regions were calculated and investigated (Table S2). The results showed that the GD and

Figure 1 Morphological characteristic of *Phyllostachys edulis* (Mao Zhu in Chinese, MZ) and 20 MZ forms. (a) MZ. (b) *P. edulis* f. *abbreviata* (Die mao zhu in Chinese, DMZ). (c) *P. edulis* f. *tubaeformis* (Sheng Yin zhu in Chinese, SY). (d) *P. edulis* 'Kikko-chiku' (Gui Jia zhu in Chinese, GJ). (e) *P. edulis* f. *ventricosa* (Fo Du mao zhu in Chinese, FD). (f) *P. edulis* f. *obliquinoda* (Qiang Zhu in Chinese, QZ). (g) *P. edulis* f. *obtusangula* (Fang Mao Zhu in Chinese, FMZ). (h) *P. edulis* f. *pachyloen* (Hou Bi mao zhu in Chinese, HB). (i) *P. edulis* f. *holochrysa* (Huang Pi mao zhu in Chinese, HP). (j) *P. edulis* f. *huamozhu* (Huang Pi Hua mao zhu in Chinese, HPH). (k) *P. edulis* f. *luteosulcata* (Huang Cao mao zhu in Chinese, HC). (l) *P. edulis* f. *nabeshimana* (Lv Pi Hua mao zhu in Chinese, LPH). (m) *P. edulis* f. *bicolor* (Lv Cao mao zhu in Chinese, LC). (n) *P. edulis* f. *venusta* (Hua Gan Jin Si mao zhu in Chinese, HGJS). (o) *P. edulis* f. *purpureoculmis* (An Ji Zi mao zhu in Chinese, AJZ). (p) *P. edulis* f. *porphyrosticta* (Ban Mao Zhu in Chinese, BMZ). (q) *P. edulis* 'Mira' (Hua Gui Zhu in Chinese, HGZ). (r) *P. edulis* f. *anjiensis* (An Ji Jin mao zhu in Chinese, AJJ). (s) *P. edulis* f. *gracilis* (Jin Si mao zhu in Chinese, JS). (t) *P. edulis* f. *epruinosa* (You Mao Zhu in Chinese, YMZ). (u) *P. edulis* f. *exaurita* (Ma Yi Zhu in Chinese, MYZ). (v) The comparison analysis of diameter at breast height (DBH) among MZ and seven MZ forms. (w) The comparison analysis of ground diameter (GD) among MZ and seven MZ forms. (x) The comparison analysis of ratio of GD to DBH (GD/DBH) among MZ and seven MZ forms. (y) The comparison analysis of node number at breast height (NBH) among MZ and seven MZ forms. The different lowercase letters indicate significant difference ($P < 0.05$). (z) Distribution of ratio of GD/DBH among 16 different MZ populations. (aa) Comparison analysis of NBH among 16 different MZ populations.



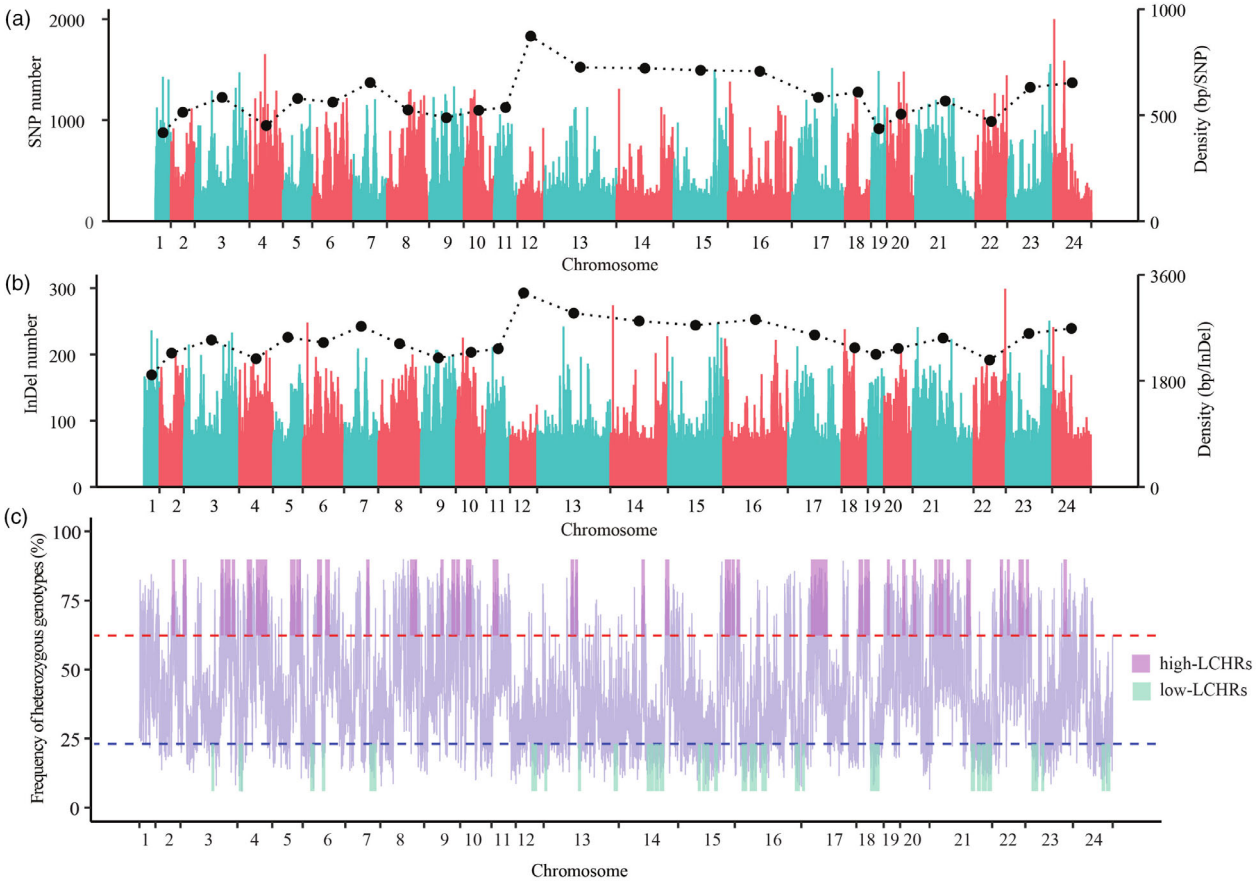


Figure 2 The characteristics of SNPs and InDels identified in 20 forms of MZ. (a) SNP number and density distribution on different chromosomes. (b) InDel number and density distribution on different chromosomes. (c) Heterozygous genotype frequency in non-overlapping 200-kb windows genome wide. The red and blue dashed lines represent the thresholds of high and low heterozygous genotype frequencies, respectively. The long continuous heterozygous SNPs clustered regions of high frequency (high-LCHRs) and long continuous heterozygous SNPs clustered regions of low frequency (low-LCHRs) are shaded in purple and green, respectively.

DBH varied among the 16 different MZ populations (Figure S1i,j), with larger GD corresponding to larger DBH. Moreover, the ratio of GD to DBH in these 16 populations was almost always less than 1 (Figure 1z). The average NBH in these 16 populations ranged from 7.91 to 9.90, with all values being less than 10 (Figure 1aa). This indicated that the characteristics of internode length and coarseness were consistent across MZ population. In conclusion, significant morphological differences were observed between different forms of MZ, with the culm shape variation being the most pronounced in GJ and SY.

Genomic resequencing and variation calling

To explore the genetic differences among the MZ forms, a total of 20 MZ forms were collected for genome sequencing (Table S3). Resequencing on the Illumina platform generated a total of 1.26 Tb data, with an average sequencing depth of 30.04× per accession. The reads were aligned to the MZ reference genome (Zhao *et al*., 2018). The properly mapping rate range from 79.47% to 97.02%, with an average mapping rate of 92.88%. After applying strict quality controls and filters, 3 230 107 high-quality SNPs and 742 564 InDels were identified (Table S4). These SNPs and InDels were distributed non-uniformly across different MZ chromosomes (Figure 2a,b), with Chromosome 21 and Chromosome 13 having the highest number of SNPs and InDels, and Chromosome 12 and Chromosome 19 having the lowest number of SNPs and InDels, respectively. Interestingly, the density of SNPs and InDels on the different chromosomes showed a high level of consistency. For instant, Chromosome 19 had the highest density of SNPs and InDels, while Chromosome 12 had the lowest. The majority of SNPs (72.05%) were found in intergenic regions (Table S5). To investigate the effects of genomic variation, SNP annotations were performed using SnpEff (Cingolani *et al*., 2012) and categorized into three levels based on their impact (Table S5; Figure S2). A total of 4496 SNPs were classified as high-level variation, which could potentially affect or modify coding gene functions, such as stop codon gained, stop codon lost, and start codon lost. These high-effect SNPs were identified in 3676 genes that were unevenly distributed across the chromosomes (Figure S3). KEGG enrichment analysis revealed significant enrichment of these genes in pathway such as ABC transporter, photosynthesis – antenna proteins and carbon fixation in photosynthetic organisms (Figure S4).

The value of transitions relative to transversions (Ts/Tv) is 0.5 in DNA sequence evolution if all types of nucleotide changes have equal rates. In fact, Ts/Tv always exceeds 0.5 or even 1 (Zou and Zhang, 2021). Among the identified variation loci in MZ forms,

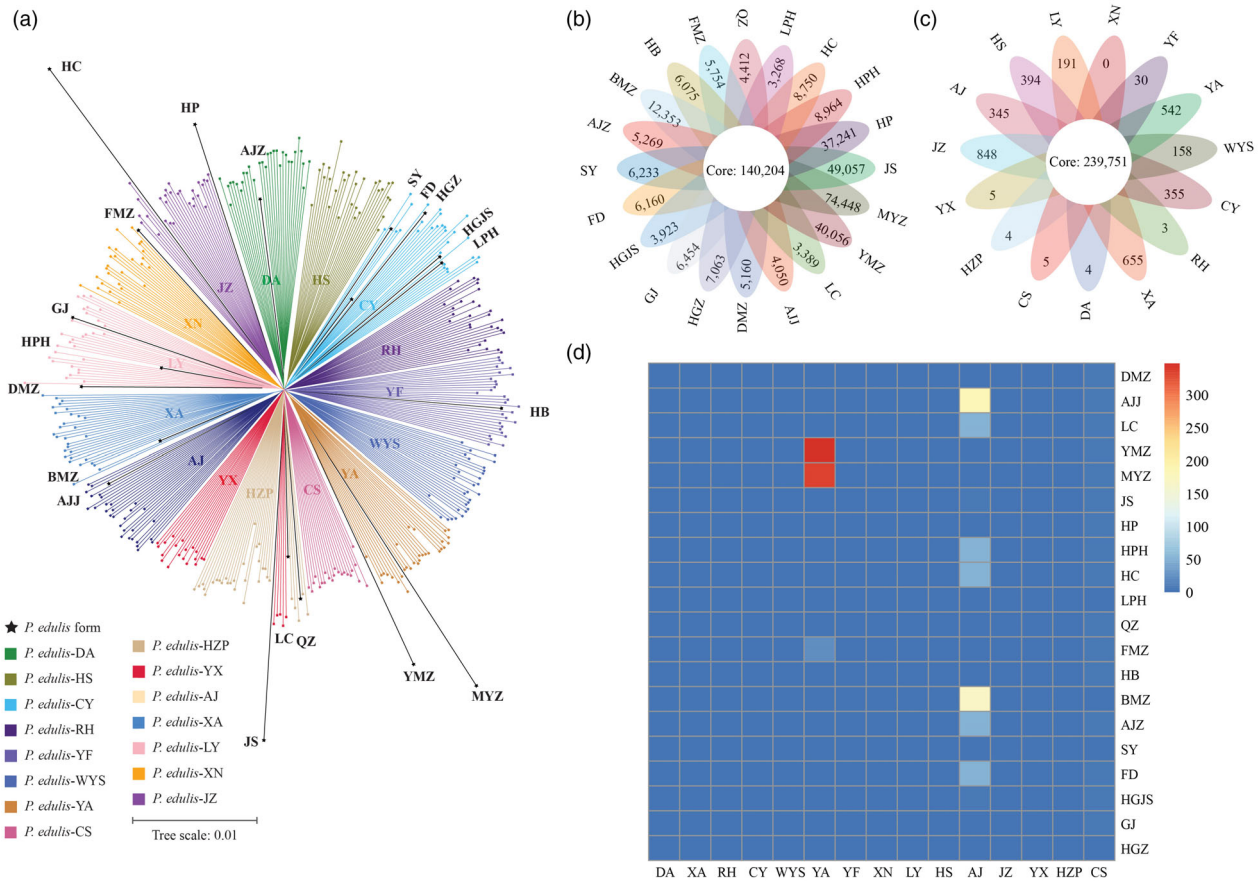


Figure 3 Population and heterozygous SNP locus analyses of MZ forms and MZ populations. (a) Unrooted neighbor-joining phylogenetic tree of 447 individuals consisting 20 MZ forms and 427 MZ individuals from 15 geographic regions. The black lines represent the MZ forms, and differently colored lines represent the individuals from fifteen geographic regions. (b) Venn diagram of heterozygous loci identified among MZ populations from 15 geographic regions. (c) Venn diagram of heterozygous loci identified among 20 MZ forms. (d) Visualization of intersection of total heterozygous loci of 15 MZ geographic regions and specific heterozygous loci of 20 MZ forms.

the Ts/Tv values ranged from 2.52 (MYZ) to 2.95 (LC and AJJ), with an average of 2.87 (Table S6). These results suggest that transitions are less deleterious and less likely to be purged by natural selection than transversions. Further analysis of the genotypes of the detected SNP sites revealed that, on average, 87.49% were heterozygous, and the average genotype heterozygosity ratio was 7.01 at the individual level (Table S6). Low-frequency heterozygous genotypes, defined as those with a frequency of heterozygous genotypes no more than 10% in sequenced samples, were identified in 28.78% of the detected SNP sites. Conversely, high-frequency heterozygous genotypes, with a frequency of heterozygous genotypes no less than 90%, were identified in 36.70% of the detected SNP sites (Figure S5). Compared to high-frequency heterozygous genotypes, low-frequency heterozygous genotypes were evenly distributed in genome regions (Figure S5). Mutation is a random process and is the driving force behind evolution and adaptation (Hershberg and Petrov, 2010). Therefore, long continuous areas with uniform patterns of heterozygous genotypes may reflect subtle gene functions underlying natural selection (Zhao *et al.*, 2021). In total, 40 long continuous regions containing clustered heterozygous SNPs with high frequencies were identified, spanning a total length of 333.8 Mb and containing 10 052 genes (Figure 2c). KEGG enrichment analysis revealed significant enrichment of these genes in pathways such as betalain biosynthesis, isoflavonoid biosynthesis, and diterpenoid biosynthesis (Figure S6a). In addition, 30 long continuous heterozygous SNPs clustered regions of low frequency were identified with a total length of 250.2 Mb and containing 8724 genes. Analysis of KEGG enrichment showed significant enrichment of these genes in pathways such as zeatin biosynthesis, brassinosteroid biosynthesis, and carotenoid biosynthesis (Figure S6b).

Population analysis and diversity

To explore the population relationship between MZ and its forms, the SNPs dataset identified in MZ forms was merged with a larger dataset consisting of 5 446 011 high-quality SNP dataset identified in 427 MZ individuals from 15 geographic regions. The merged SNP dataset contained a total of 6 892 705 SNPs. After applying filters to correct for excess linkage disequilibrium, 1 406 535 SNPs were retained for genetic population analysis (Table S7). The ADMIXTURE program was used for the population structure analysis. The result indicated that the best clustering was observed at $K = 1$, suggesting minimal population differentiation among these samples (Figures S7 and S8). Moreover, distance-based cluster analysis and principal component analysis (PCA) were largely consistent with the geographic distribution of the samples. In these analyses, individuals from the

same region tended to cluster together (Figure 3A; Figure S9). Furthermore, based on the distance-based cluster analysis and PCA, certain MZ forms showed close genetic similarity to individuals from specific regions. For example, SY, FD, HGJ, HGJS, and LPH were found to be genetically closest to individuals from Congyi County, suggesting a possible origin from this County. Similarly, GJ, HPH, and DMZ showed similarity to individuals from Longyou County, as a potential origin region. However, Huangshan City, Renhua County, Wuyishan City, and Chishui City were genetically furthest away from the MZ forms analysed in this study. These findings provide insights into the population relationships and potential origins of the MZ forms based on their genetic profiles compared to individuals from different geographic regions.

To further investigate the genetic relationship among MZ forms and MZ, an alternative method was employed, focusing on the distribution of common heterozygous loci. This method aimed to estimate the origin of MZ forms by identifying specific heterozygous loci that represent characteristics of different geographic regions or MZ forms (Nishiyama *et al*., 2023). To estimate the origin of MZ forms, an MQ value of 60, which corresponds to the highest genotype accuracy in the SNP dataset of MZ forms, was used to identify reliable candidate SNPs. This allowed the detection of total and specific heterozygous loci datasets. Among the MZ forms, MYZ exhibited the highest number of both total and specific heterozygous loci, with 289 791 and 74 448, respectively. Conversely, HC had the lowest number of total heterozygous loci (230132), and LPH had the lowest number of specific heterozygous loci (3268) (Figure 3b; Table S8). Regarding different geographic regions of MZ, when a heterozygous genotype was present in more than 80% of the total individuals at a specific site, it was considered a representative heterozygous locus. The number of total heterozygous loci with an MQ value of 60 was nearly identical among different geographic regions, ranging from 240 601 (Chishui City) to 241 486 (Jinzhai County) (Figure 3c; Table S9). The maximum number of specific heterozygous loci (848) was observed in Jinzhai County, while no specific loci were detected in Xianning City. A heatmap was used to visualize the intersection of the total heterozygous loci from different MZ regions and the specific heterozygous loci dataset from different MZ forms (Figure 3d). AJJ, LC, HPH, HC, BMZ, AJZ, and FD exhibited an 'outlier peak' in the An'ji County, indicating that these forms may have originated from An'ji County. Similarly, *P. edulis* f. *epruinosa* (You Mao Zhu in Chinese, YMZ), MYZ, and FMZ may have derived from Yong'an City, based on the observed patterns in the heatmap. These findings provide additional evidence for the genetic relationships among MZ and its forms, suggesting potential origins based on the distribution of common heterozygous loci.

Candidate genes associated with important morphological differentiation

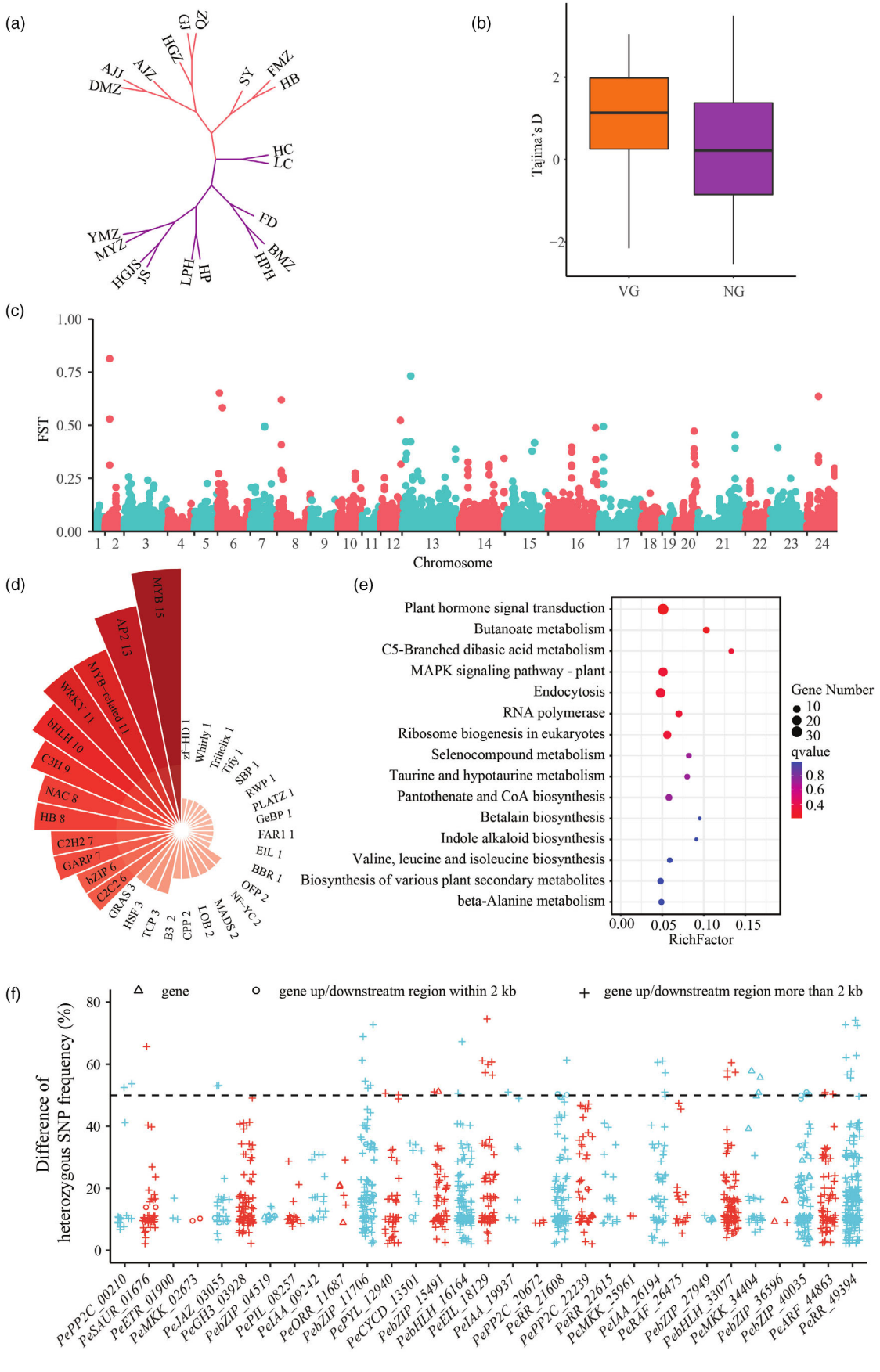
To elucidate the genetic relationships among the sampled MZ forms on a genome-wide scale, a phylogenetic tree was constructed using a pruned linkage disequilibrium SNP dataset from 20 MZ forms. Based on this phylogenetic tree and morphological characters, the 20 MZ forms were divided into two groups (Figure 4a). Group I consisted of nine members, seven of which exhibited distinct culm shape variation, such as DMZ, SY, and GJ. Group II contained 11 members, and interestingly, seven of them displayed distinct culm colour variation, such as HP, LHP, and HPH. To identify the genomic regions and genes responsible for culm shape variation, 16 out of the 20 MZ forms were categorized into culm shape variation group (VG) and culm shape normal group (NG), excluding individuals with strong admixture (Table S3). The Tajima's D value was calculated to assess genetic variation within each group. The VG exhibited higher Tajima's D values (mean = 1.05) and a positive skew, while the NG had lower Tajima's D values (mean = 0.30) (Figure 4b), indicating positive selection acting on the VG. Furthermore, *F*-statistics (F_{st}) values were calculated between the two groups to identify candidate genomic regions associated with culm shape variation differentiation (Figure 4c). Using a threshold of F_{st} value set at 0.07 (top 1% of empirical distribution for F_{st} value), a total of 1512 genes were identified as candidate genes related to culm shape differentiation, among which 143 genes encoded transcription factors (TFs) belonging to 32 families. The most abundant TF families were MYB, AP2, and MYB-related (Figure 4d). Additionally, 24 genes encoding transcription regulators (TRs) and 55 genes encoding protein kinases (PKs) were identified.

KEGG enrichment analysis revealed significant enrichment of these candidate genes in pathways related to plant hormone signal transduction, butanoate metabolism, and MAPK signalling pathway (Figure 4e). Furthermore, we identified SNPs related to plant hormone signal transduction genes and investigated the heterozygous SNP frequency difference between NG and VG. A total of 2717 SNPs were identified, among which 1417 SNPs had heterozygous frequency difference between the two groups, with merely 37 SNPs having differences greater than 50% (Figure 4f; Table S10). In addition, only 38 SNPs were located within 12 genes, which was mainly distributed in *PebZIP_40035* and *PeMKK_34404*. Such findings suggest that those genes are related to plant hormone signal transduction, which may play important roles in culm shape differentiation.

Transcriptome profiles revealed differential regulation in NG and VG shoots

To further investigate the difference in the molecular regulation between NG and VG, a comparative shoot transcriptome analysis was conducted using MZ and its two forms (GJ and SY), which are typical representative forms of NG and VG, respectively. We collected a total of 66 shoot samples at different developmental stages, including 21 GJ samples, 27 SY samples, and 18 MZ samples (Figure S10). Transcriptome sequencing of these 66 samples generated 441.59 Gb of data, with mapping rate ranging from 93.89% to 98.52%, and an average mapping rate of 97.10% on MZ genome (Table S11). For a gene to be

Figure 4 Identification and analysis of candidate genes related to culm shape variation. (a) Unrooted neighbor-joining phylogenetic tree of 20 MZ forms. (b) Comparison analysis of Tajima's D values between culm shape normal group (NG) and culm shape variation group (VG). (c) Distribution of windowed F_{st} values for NG and VG on chromosomes (window size = 50 kb and step = 10 kb). (d) Number of TFs identified in the candidate gene sets. (e) KEGG enrichment analysis of all the candidate genes. (f) Distribution of heterozygous SNP frequency difference associated with plant hormone signal transduction genes.



considered expressed in a given sample, its transcripts per million (TPM) value had to be more than 1. Out of the 45 198 genes detected, 37 546 genes were expressed in at least one sample. A total of 29 528 genes were identified as differentially expressed genes (DEGs), showing substantial differences in at least one comparison. Specifically, there were 26 834 DEGs in comparison of GJ versus MZ and 24 246 DEGs in comparison of SY versus MZ (Figure 5a; Figure S11).

According to the KEGG enrichment analysis, these DEGs (29528) were significantly enriched in the processes such as ribosome (504), plant hormone signal transduction (390), and DNA replication (102) (Figure 5b). Furthermore, we found that 2192 TFs belonging to 52 families were differentially expressed in at least one comparison. The top five families with the most members were bHLH (183), AP2/ERF (170), MYB (157), NAC (147), and HB (137) (Figure S12). To explore the potential TF-mediated transcriptional regulation involved in plant hormone signal transduction, the Pearson correlation coefficient (PCC) between those genes were measured and the threshold was set to 0.70 to filter high correlation gene pairs for co-expression network construction. Ultimately, 169 403 gene pairs containing 2022 TFs were identified (Figure 5c). The highest edge numbers were found in the genes related to auxin signal (25604), followed by ETH signal (16813) and ABA signal (7276). These results provide further evidence that the pathway of plant hormone signal transduction is associated with culm shape variation. As expected, the analysis of hormone levels among variant internodes and normal internodes revealed that the contents of auxin, CK, and GA in the variant internode were significantly higher than those in normal internodes, whereas the contents of JA and ABA in the variant internodes were significantly lower than those in normal internodes (Figure 5d).

Identification and validation of the genetic elements in the regulatory network controlling culm shape variation

Out of the 1512 candidate genes associated with culm shape variation, 908 genes were identified as DEGs (Figure 6a), including 89 TFs, 12 TRs, and 39 PKs (Table S12; Figures S13–15). Ten genes were randomly selected and validated by real-time quantitative PCR (RT-qPCR) (Figures S16 and S17). These genes were significantly differentially expressed between variant and normal internodes, indicating they may play key roles in culm shape variation. KEGG enrichment analysis revealed that all these DEGs were significantly enriched in the pathway of plant hormone signal transduction (Figure 6b). To investigate the potential TF-mediated transcription, the PCC among those DEGs related to hormone signal transduction and the differentially expressed TFs were calculated for co-expression network construction. Ultimately, a total of 122 gene pairs involved in six types of hormone signal, including auxin, CK, and ETH, were identified (Figure 6c). The highest number of gene pairs were found among genes related to auxin signals (37), followed by CK signals (31) and ETH (22), whereas the fewest gene pairs were found in genes related to GA (4), indicating potential mutual regulation between these hormone signals causing culm shape variation. To validate the regulatory network, the expression of 10 randomly selected DEGs related to auxin, CK, and ETH was validated. Similar gene expression trends (upregulation or downregulation) were observed in RT-qPCR results as those of high-throughput sequencing for most of the samples (Figure S18). These results indicated that several hormone signal such as auxin, CK, and ETH, may play important roles in culm shape variation.

To investigate the regulatory interactions among hormones, several TFs, such as PebZIP_27949, PebZIP36596, and PebHLH_16164, were identified to be co-expressed with the genes related to hormone signalling. Through *cis*-element analysis, PebZIP_27949 was a potential regulator of genes associated with auxin, CK, and ETH (Figure 7a; Table S13). For instance, *PeRAF_26475*, a homologue of B4 clade RAF-like kinases in Arabidopsis (*Arabidopsis thaliana*) known to be involved in mediating rapid auxin responses (Kuhn *et al*., 2023), exhibited two G-boxes in its promoter regions (pRAF-1 and pRAF-2) (Figure 7b). Similarly, *PeEIL_18129*, a homologue of ETHYLENE INSENSITIVE3-LIKE factor, which acts as a positive regulator of ETH signalling in Arabidopsis and rice (*Oryza sativa*) (Binder, 2020), displayed one G-box and a putative binding motif in its promoter regions (Figure 7b). These findings suggested that PebZIP_27949 may serve as a potential regulator for both *PeRAF_26475* and *PeEIL_18129*.

To test this possibility, yeast one-hybrid (Y1H) and electrophoretic mobility shift assays (EMSAs) were conducted, confirming that PebZIP_27949 directly bound to the promoters of both *PeRAF_26475* and *PeEIL_18129* (Figure 7c–f). Besides G-box, PebZIP_27949 could bind to the CCGCGTGTCA motif in the *PeEIL_18129* promoter (Figure 7c–f). Dual luciferase reporter assays (DLR) were performed using *35S::PebZIP_27949* as the effector and two promoter fragments containing G-boxes as reporters. The results showed that luciferase activity controlled by both two *PeRAF_26475* promoter fragments was elevated remarkably when PebZIP_27949 was expressed (Figure 7g,h). In addition, the results of DLR assays further supported the expression of *PeEIL_18129* activated by PebZIP_27949 (Figure 7i, j). Taken together, these findings demonstrate that PebZIP_27949 could directly activate the expression of both *PeRAF_26475* and *PeEIL_18129* by binding to their promoters, respectively. Moreover, the expression patterns of *PebZIP_27949*, *PeRAF_26475*, and *PeEIL_18129* were similar in MZ seedlings treated with NAA or ETH (Figure S19), suggesting that PebZIP_27949 may respond to auxin and ETH signalling by regulating the expression of *PeEIL_18129* and *PeRAF_26475* simultaneously (Figure S20).

Discussion

Being one of the most important bamboo species, MZ has undergone a long period of historical evolution, resulting in various forms with distinct morphological variations (Figure 1a–u). In this study, a preliminary investigation involving MZ (458 bamboo culms from 16 geographic regions) and its forms (799 bamboo culms from three geographic regions) confirmed the main morphological differences between the MZ forms and the original MZ (Figure 1v–aa; Figure S1). Whole-genome resequencing analysis revealed significant genetic similarities among them, with an average mapping rate of up to 92.88% on MZ genome. In comparison to Arabidopsis (Ossowski *et al*., 2010) and rice (Wang *et al*., 2018). MZ and its forms exhibited lower SNP density and notably high levels of genotype heterozygosity ratio at the individual level (Figure 2). MZ primarily propagates through a unique rhizome-dependent proliferation system (asexually), which can tolerate more somatic mutations compared to plants propagated purely through sexual reproduction (Ding *et al*., 2022; Huang *et al*., 2018; Nishiyama *et al*., 2023; Zhao *et al*., 2021). Consequently, MZ was characterized by a large quantity of low-frequency and high-frequency heterozygous genotypes (Zhao *et al*., 2021), which were also observed in the MZ forms. Genes

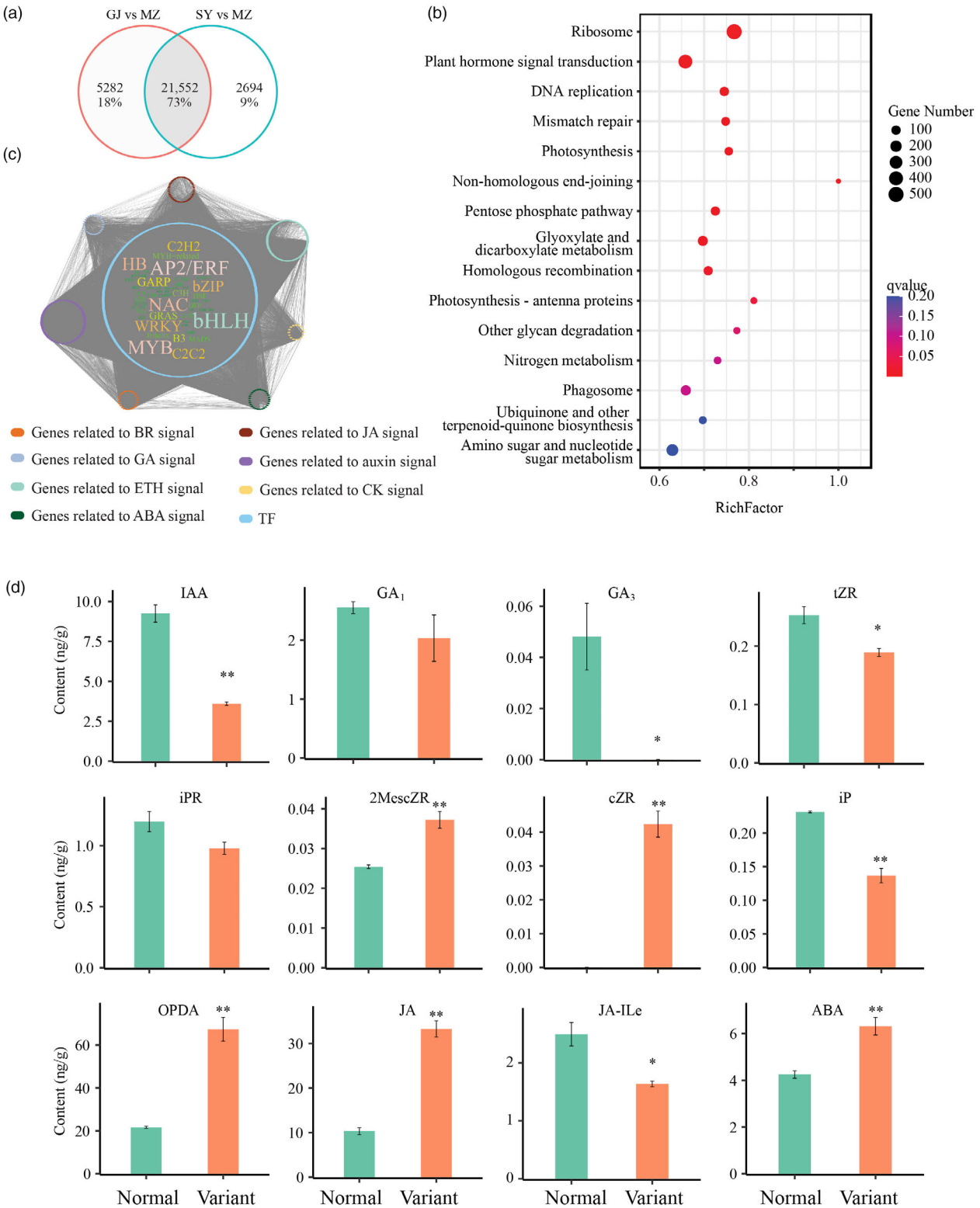


Figure 5 Identification of differentially expressed genes (DEGs) and analysis of endogenous hormone contents in shoots of VGs and NGs. (a) Number of DEGs between different comparison of GJ vs MZ and SY vs MZ. (b) KEGG enrichment analysis of DEGs. (c) Co-expression network of TFs and genes related to plant hormone signal transduction. The number of TFs indicates the font size in the circle. (d) Hormone contents of indole-3-acetic acid (IAA), gibberellin A1 (GA_1), gibberellin A3 (GA_3), *trans*-zeatin riboside (tZR), N6-isopentenyladenosine (iPR), 2-methylthio-*cis*-zeatin riboside (2MescZR), *cis*-zeatin riboside (cZR), N6-isopentenyladenine (iP), *cis*(+)-12-oxophytodienoic acid (OPDA), jasmonic acid (JA), jasmonoyl-L-isoleucine (JA-ILe), and abscisic acid (ABA) in the variant and normal internodes. * and ** represent $P < 0.05$ and $P < 0.01$, respectively.

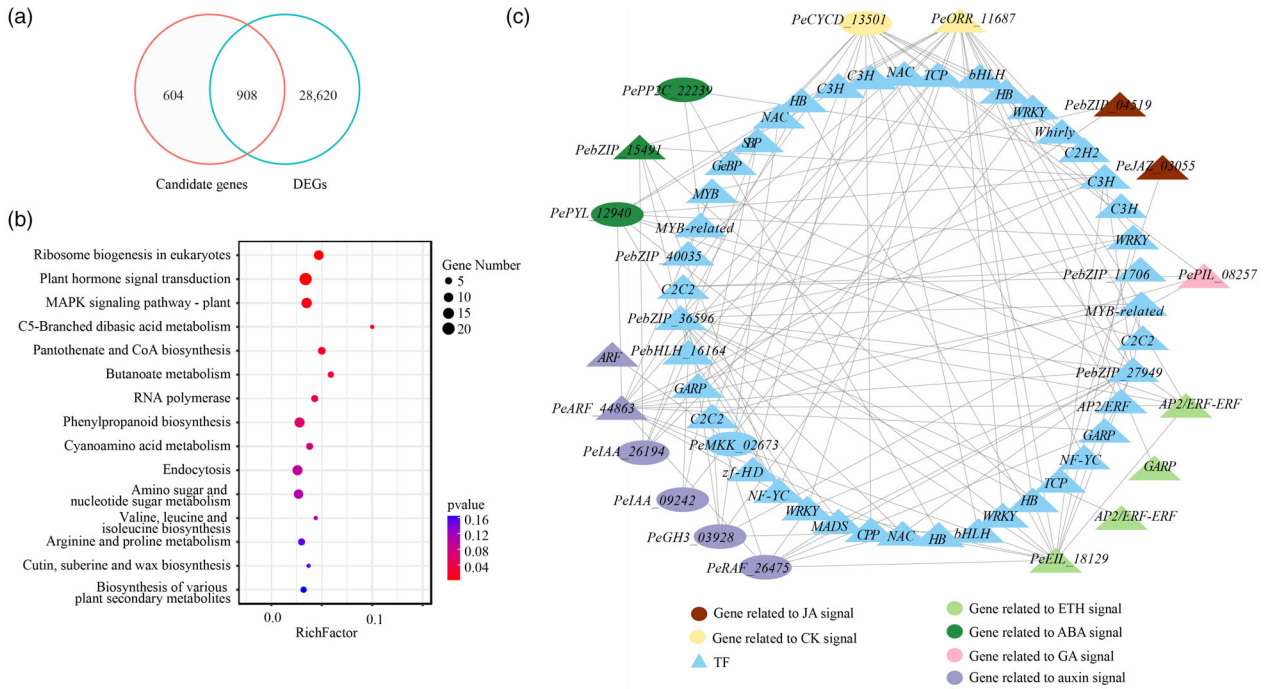


Figure 6 Analysis of candidate genes involved in culm shape variation. (a) Venn diagram of candidate genes and DEGs. (b) KEGG analysis of DEGs in candidate genes. (c) Visualization of co-expression network based on DEGs involved in plant hormone signal transduction.

located in regions of low heterozygosity were found to be enriched in the zeatin biosynthesis and brassinosteroid biosynthesis processes, suggesting that plant hormones may play a vital role in the population evolution and adaptation of the MZ forms.

In this study, a merged dataset of SNPs, derived from 20 MZ forms and 427 MZ individuals of 15 major geographic regions, was employed to construct phylogenetic tree and identify the common heterozygous loci. The results were considered highly reliable when the phylogenetic tree analysis was consistent with the identification of common heterozygous loci. For instance, the reliable candidate origin of AJJ was determined to be An'ji County (Zhang, 2008), and the reliable candidate origin of MYZ and YMZ were found to be Yong'an County. These results were in agreement with previous reports (Chen, 2013). Additionally, the phylogenetic tree analysis revealed that the candidate origin of BMZ was Xing'an County (Guangxi Province), while the identification of common heterozygous loci indicated that it originated from An'ji County (Zhejiang Province). Interestingly, these results were consistent with previous report that BMZ (Hua *et al.*, 2012) was serendipitously discovered in the MZ forestry in Guangxi Province and subsequently introduced to Zhejiang Province. Furthermore, HB (Yang *et al.*, 1997) and *P. edulis* f. *gracilis* (Jin Si mao zhu in Chinese, JS) (Lai, 2012) initially identified in Yifeng and Yixing County, respectively, corresponded with the phylogenetic tree analysis but differed from the results of common heterozygous loci identification. According to the identification of common heterozygous loci, both AJZ and FD originated from Anji County, which aligned with the previous reports (Zhang *et al.*, 2012). Interestingly, these results were consistent with previous reports that BMZ (Hua *et al.*, 2012) was serendipitously discovered in the MZ forestry in Guangxi Province and subsequently introduced to Zhejiang Province. Although both phylogenetic tree construction and identification of common heterozygous loci play important roles in clarifying the origin issue of MZ forms, the results obtained from both methods are not always consistent with those previously reported. Therefore, further studies with a broader range of MZ forms derived from different geographic regions and comprehensive analyses are needed to uncover the origins of these MZ forms and their genetic relationship with MZ.

One of the primary objectives of this study was to identify candidate genes associated with culm shape variation. Through a genome-wide analysis, it was found that the F_{st} values between NG and VG groups were relatively lower compared to other plants, such as garlic (*Allium sativum*) (Wang *et al.*, 2023), orange (*Citrus sinensis*) (Feng *et al.*, 2021), and apple (*Malus domestica*) (Duan *et al.*, 2017). These lower F_{st} values are primarily attributed to the unique flowering characteristics of bamboo. Bamboo has a long flowering cycle and unpredictable flowering periods, which have limited artificial selection and predominantly resulted in natural selection within the population (Ramakrishnan *et al.*, 2020; Zhao *et al.*, 2021). As a result, the breeding processes in bamboo have lagged behind, leading to limited population differentiation (Zhao *et al.*, 2021). Consequently, only the top 1% of the empirical distribution for F_{st} values was ultimately identified as candidate genes involved in culm shape variation. Notably, these candidate genes were significantly enriched in the plant hormone signal transduction pathway. Similar results were obtained from the transcriptome analysis among NG and VG shoots. Therefore, these findings collectively suggest that the genes involved in plant hormone signal transduction pathway may play a significant role in the formation of culm characteristics during evolution.

Lower contents of plant hormones, such as auxin and CK, in the shoot apical meristem have been observed to downregulate the expression of genes associated with hormone signalling. This

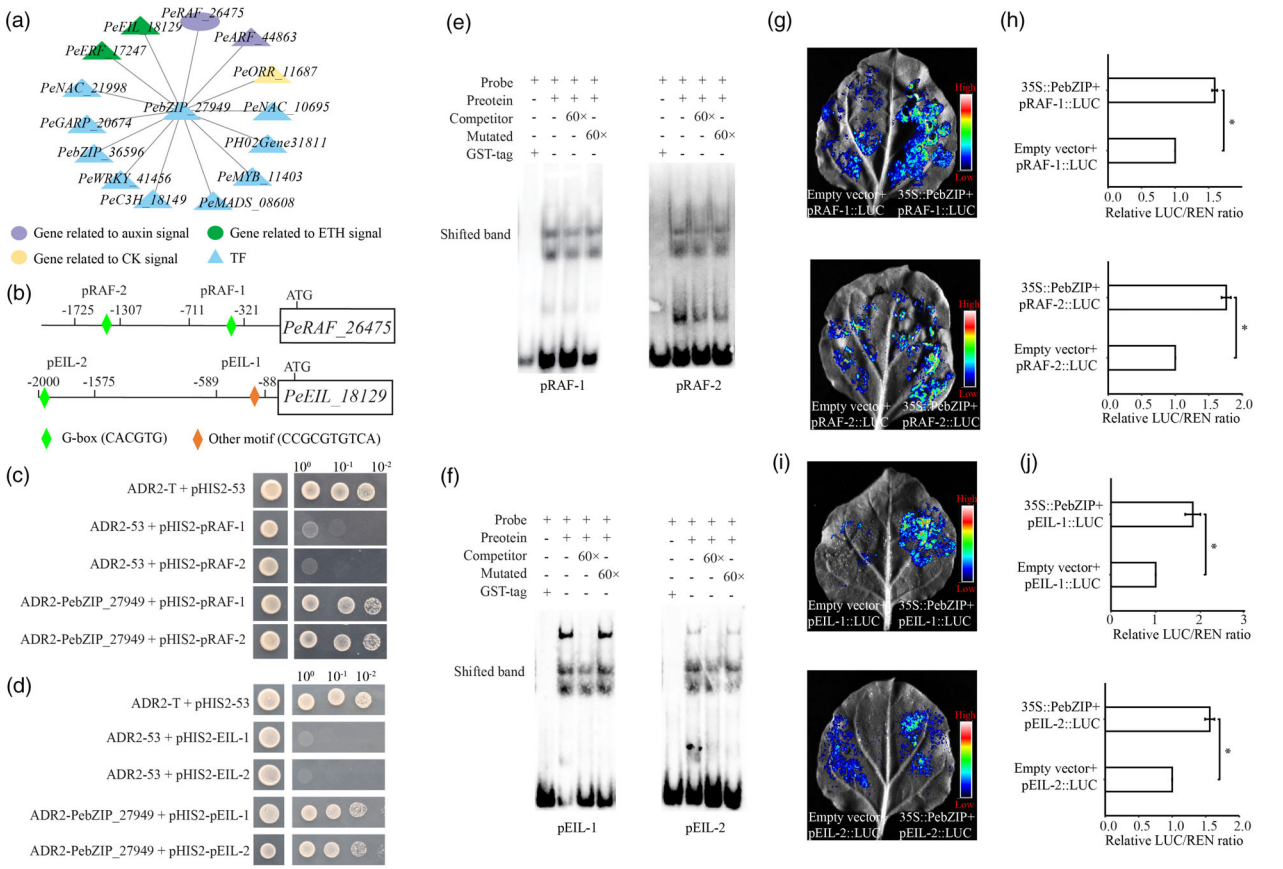


Figure 7 Identification of a key transcription factor regulating plant hormone signal transduction genes. (a) The co-expressed network with *PebZIP_27949* as hub. (b) Diagram of the *PeRAF_26475* and *PeEIL_18129* promoter regions containing potential binding sites for PebZIP_27949. (c) Yeast one-hybrid (Y1H) assays showing the interaction between PebZIP_27949 and two sections of *PeRAF_26475* promoter (pHIS2-pRAF-1 and pHIS2-pRAF-2). The yeast cells were grown on SD/-Leu/-Trp and SD/-Leu/-Trp/-His with 60 mM 3-amino-1,2,4-triazole (3-AT) medium for three days. (d) Y1H assays showing the interaction between PebZIP_27949 and two sections of *PeEIL_18129* promoter (pHIS2-pEIL-1 and pHIS2-pEIL-2). The yeast cells were grown on SD/-Leu/-Trp and SD/-Leu/-Trp/-His with 60 mM 3-amino-1,2,4-triazole (3-AT) medium for three days. (e) PebZIP_27949 binding to two G-box motifs (5′-CACGTG-3′) of the *PeRAF_26475* promoter by electrophoretic mobility shift assay (EMSA). Unlabeled probes were used as competitors. The 60× represents the concentrations of the competitor. Mut-box represents a mutated probe in which the motif is replaced by 5′-AAAAAA-3′. (f) PebZIP_27949 binding to G-box motif and the putative motif (5′-CCGCGTGTCA-3′) of the *PeEIL_18129* promoter by EMSA. Unlabeled probes were used as competitors. The 60× represents the concentrations of the competitor. Mut-boxes represent mutated probes in which the motifs are replaced by 5′-AAAAAA-3′ and 5′-AAAAAAAAA-3′, respectively. (g and h) PebZIP_27949 activating the transcription of both *PeRAF_26475* and *PeEIL_18129* promoters in tobacco leaves by dual luciferase assays, respectively. Representative photographs were taken using a chemiluminescence imaging system. (i and j) LUC/REN activity detection to verify PebZIP_27949 activating the transcription of *PeRAF_26475* and *PeEIL_18129*, respectively. The empty vector was used as a negative control. LUC and REN represent Luciferase and *Renilla* luciferase, respectively. The data are means ± SD (n ≥ 3) (*$P \leq 0.05$).

downregulation of gene expression can lead to dwarfing and reduced diameter of bamboo culms (Wang *et al.*, 2019). Further studies revealed that exogenous auxin affected the plant height of bamboo by promoting internode elongation rather than increasing the number of nodes. Exogenous auxin treatment was found to significantly increase the contents of GA and CK while decreasing the content of ABA and ETH, thereby promoting culm internode elongation (Bai *et al.*, 2023). Similar results were also found in this study, the variant internodes displayed lower contents of auxin, CK, and GA and higher contents of JA and ABA compared with normal internodes (Figure 6d). However, different contents of exogenous auxin had no significant effect on internode cell elongation in MZ seedlings (Chen *et al.*, 2022). A similar phenomenon has been observed in rice, where the accumulation of ETH decreases ABA synthesis but increases GA synthesis and responsiveness, promoting internode elongation in deepwater rice plants (Nagai and Ashikari, 2023). These findings suggest that the interaction between different hormones is really a complex regulatory network in plant growth and development, which is influenced by multiple factors such as hormone contents, environmental conditions, and genotypes.

Based on the transcriptomic analysis, a co-expression network consisting of 122 gene pairs was constructed, in which auxin signalling genes exhibited the highest number of gene pairs, indicating their importance in cell proliferation and elongation processes. Within the identified co-expression network, *PeARF_44863* exhibited the highest number of connections and was co-expressed with genes related to other hormones signalling, such as CK, ETH, and ABA. *PeRAF_26475* was another hub gene identified in the auxin signalling within the co-expressed network, and its homologues have been identified as central mediators for a deeply conserved, rapid auxin response process across land plants

and algae (Kuhn *et al*., 2023). Interestingly, *PeEIL_18129*, a gene related to ETH signalling, was not only co-expressed with above two genes related to auxin signalling but also co-expressed with *PeORR_11687*, a homologue gene to *OsORR2*. Overexpressing *OsORR2* could reduce plant height in rice compared with wild-type rice (Shi *et al*., 2020). In the co-expression network, both four genes above mentioned were co-expressed with TF *PebZIP_27949*, and the putative regulatory relationship between TF PebZIP_27949 and two hormone signalling genes, *PeEIL_18129* and *PeRAF_26475*, was validated by Y1H, EMSA, and DLR. This suggests the existence of an intricate regulatory network involving hormone interactions to regulate the variation of culm shape. Further research is still needed to elucidate the precise mechanisms by which these related genes regulate the variation in culm shape.

Experimental procedures

Morphological investigation and sample collection

MZ and its form populations grow in China Bamboo Expo Park, Huzhou City, Zhejiang Province, the base of Jiangxi Academy of Forestry, Nanchang City, Jiangxi Province, and the test centre of Taiping, Huangshan City, Anhui Province, were selected for morphological investigation. The diameter at breast height (DBH), ground diameter (GD), and the number of nodes at breast height (NBH) were investigated for at least 30 individuals from each form.

Leaf samples were collected from 20 MZ forms at four locations, including the test centre of Taiping, the China Bamboo Expo Park, the base of Zhejiang Academy of Forestry in Hangzhou City, and the base of Jiangxi Academy of Forestry. Table S3 provides detailed information about the 20 MZ forms.

The 4-week-old MZ seedlings were collected after being treated with NAA (5 μM) or ETH (5 ppm) for 0, 1, 2, and 8 h, respectively. The plant materials were snap-frozen in liquid nitrogen and stored at −80 °C for total RNA extraction.

DNA sequencing and variations identification

The total DNA was extracted from the collected leaf samples using the CTAB method. For each sample, a minimum of 5 μg of genomic DNA was utilized to construct paired-end sequencing libraries. The paired-end sequence reads were then aligned to the MZ reference genome (Zhao *et al*., 2018) using the BWA-MEM2 (v 2.2.1) (Li, 2013) software with default parameters. The resulting alignments were subjected to deduplication using SAMtools (v 1.16.1) (Li *et al*., 2009). Next, the HaplotypeCaller function of GATK (v 4.1.8.0) (McKenna *et al*., 2010) was employed to generate GVCF files. Population variant calling was performed using the GenotypeGVCFs function of GATK with default parameters, resulting in variant calls for each variant. To ensure the quality of the variant set, a hard filtering step was applied using GATK. Specifically, SNPs were filtered based on the parameters 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || QUAL < 30.0', while InDels were filtered using the parameter 'QD < 2.0 || FS > 60.0'. To mitigate the possibility of false-positive SNPs due to multiple mapping, regions with a ratio of multiple mapping reads higher than 25% were identified. SNPs within these regions were subsequently removed. The annotation information for the SNPs was obtained using SnpEff (v 5.1d) (Cingolani *et al*., 2012).

Analysis of the genome-wide heterozygous genotype frequency

To investigate the genome-wide pattern of the genomic heterozygosity, the single-site and window levels of heterozygous genotype frequency were calculated using previously published methods (Zhao *et al*., 2021). The average heterozygous genotype frequency at the 200-kb-window level was found to be 42.72%, with a standard deviation of 19.62%. Consequently, regions with heterozygous genotype frequencies higher or lower than the average by one standard deviation were defined as high-frequency (62.34%) and low-frequency heterozygous regions (23.10%), respectively. Furthermore, continuous regions of high-frequency and low-frequency heterozygosity with a length greater than 5 Mb were identified. These regions were referred to as long continuous high-frequency heterozygous regions (high-LCHRs) and low-frequency heterozygous regions (low-LCHRs), respectively.

Population analysis and structure analysis

A SNP dataset obtained from 427 MZ individuals and 20 MZ forms was merged using bcftools (v 1.14) (Danecek *et al*., 2021). The merged SNP dataset was subsequently filtered for the population structure analysis using Plink (v 1.90) (Slifer, 2018) with the parameters '--indep-pairphase 100 10 0.1'. To quantify the correlation between MZ samples, the Identity-By-State genetic distance matrix was computed using Plink with the parameters '-distance 1-ibs flat-missing'. Based on this distance matrix, a neighbour-joining (NJ) phylogenetic tree was constructed using the PHYLIP (v 3.697) (Retief, 2000), employing the 'neighbour' parameter. The resulting tree was visualized using iTOL (Letunic and Bork, 2021). Principal component analysis (PCA) was performed using GCTA (v 1.93.2 beta) (Yang *et al*., 2011), employing the default parameters. The population structures were assessed using ADMIXTURE (v 1.3.0) (Alexander and Lange, 2011) with CV error from $K = 1$ to $K = 6$ and visualization was performed using the R package Pophelper (v 2.3.1) (Francis, 2017). Vcftools (v 0.1.16) (Danecek *et al*., 2011) was used to calculate the population differentiation (F_{st}) and Tajima's D values. F_{st} values were calculated using a 50 kb window with a 10 kb step size and Tajima's D values were calculated using a 50 kb window size.

Common heterozygous loci identification

To further improve the accuracy of the loci information, the MQ values (corresponding to the highest genotyping accuracy) were used to filter the most reliable candidate SNPs. The highest MQ value in 20 MZ forms was 60. Therefore, the heterozygous loci with MQ value equal to 60 were identified as heterozygous loci dataset for each MZ form. Additionally, a heterozygous locus existed in more than 80% of the individuals from the same region and was considered to be representative heterozygous loci for this region. Thus, the representative heterozygous loci with MQ value equal to 60 were identified as common heterozygous loci for 15 different geographic regions.

RNA-Seq, processing, and gene functional annotation

The shoots of MZ, GJ, and SY were used for a transcriptome analyses, and they both grow in the test centre of Anhui Taiping, Huangshan City, in Anhui Province. The top, middle, and basal portions of the SY and MZ bamboo shoots of different heights (0.5 m and 3.0 m) were selected and harvested in April 2023. Four parts of 3.0 m GJ shoot were collected: the basal internode (the first variation internode, GJ1), the end variation internode (GJ3), the end variation internode forward of the fifth internode (variation internode, GJ2), and the end variation internode backward of the fifth internode (normal internode GJ4). In addition, the top, middle, and basal parts of the selected GJ

derived from the same rhizomes with a height of 3.0 m but without internode variation (named GM), as well as SY samples (named SM), were collected. After harvesting, they were immediately frozen in liquid nitrogen and used for RNA-Seq. Each sample was sequenced, which included three biological replicates (Figure S10).

Total RNA isolation was performed using TRIzol reagent. The data for a total of 66 transcriptome profiles were obtained using the Illumina NovaSeq 6000 platform. Hisat2 (v 2.2.0) was used to align the paired-end reads with MZ reference genome (Zhao *et al.*, 2018). The value of transcripts per million mapped reads (TPM) was used to represent the transcript abundances using featureCounts (v 1.6.2) software (Liao *et al.*, 2014). Differential expression analysis was performed using DESeq2 (v 1.38.0) (Love *et al.*, 2014), and genes with Fold Change (FC) ≥ 2 and False Discovery Rate (FDR) < 0.05 were assigned as differentially expressed. The putative functions of the differentially expressed genes (DEGs) were determined based on NR database and Swiss-Prot database.

Determination of endogenous phytohormone contents

The endogenous phytohormone contents of variant internodes (the first variation internode, GJ1) and normal internodes (the basal internode of GM) with three biological replications were measured by high-performance liquid chromatography-mass spectrometry (HPLC-MS) described by (Pan *et al.*, 2010; Simura *et al.*, 2018). A total of five plant hormones were tested, namely, auxin (indole-3-acetic acid, IAA), CK (*trans*-zeatin riboside, tzR; N6-isopentenyladenosine, iPR; 2-methylthio-*cis*-zeatin riboside, 2MeScZR; *cis*-zeatin riboside, cZR, and N6-isopentenyladenine, iP), JA (JA, *cis*(+)-12-oxophytodienoic acid, OPDA and jasmonoyl-L-isoleucine, JA-ILe), ABA, and GA (GA_1 and GA_3).

Real-time quantitative PCR (RT-qPCR)

Primer Premier 5.0 software was used to design the specific primers based on the candidate gene sequences (Table S14). The RT-qPCR reaction was performed with the Roche Light CYCLER 480 SYBR Green I Master Mix (Roche, 04887352001, Germany) in a qTOWER 2.2 system (Analytik Jena, Germany) with three independent experiments, using *PeTIP41* (Fan *et al.*, 2013) as the internal controls. The relative abundance of each gene was calculated from the $2^{-\Delta\Delta CT}$ values (Livak and Schmittgen, 2001) between the target genes with the reference gene.

Yeast one-hybrid (Y1H) assay

The coding sequences (CDS) of *PebZIP_27949* were recombined into the pGADT7-Rec2 vector, and the fragment sequences of *PeEIL_18129* and *PeRAF_26475* promoters were inserted into the pHIS2 vector, respectively. Two recombinant vectors were co-transformed into the Y187 strain. The transformants were screened on SD/-Leu/-Trp/-His medium supplemented with 60 mM of 3-amino-1,2,4-triazole for three days.

Dual-luciferase reporter (DLR) assay

The CDS (without a stop codon) of *PebZIP_27949* was inserted into the pGreenII 62-SK vector and used as effectors, while the fragment sequences of the *PeEIL_18129* and *PeRAF_26475* promoters were inserted into the vector of pGreenII 0800-LUC upstream the LUC gene, respectively. All constructs were transformed into *Agrobacterium* GV3101 (pSoup) (Zomanbio, ZC1406, China). The constructed vectors were infiltrated into tobacco leaves for transient expression. The LUC fluorescence signal was detected using a chemiluminescence imaging system (Tiano4800, China), and LUC and REN luciferase activities were measured using the dual-luciferase® reporter assay system (Vazyme, DL101-01, China) with a single tube luminescent detector (Lumipro, China).

Electrophoretic mobility shift assay (EMSA)

The EMSA was conducted using a chemiluminescent EMSA kit (Beyotime, GS009, China) following the manufacturer's instructions. The CDs of *PebZIP_27949* were cloned into the pGEX4T-1 vectors containing a GST tag. The recombinant PebZIP_27949-GST protein was induced in *Escherichia coli* strain BL21 by 0.6 mM isopropyl β-D-1-thiogalactopyranoside and expressed at 37 °C. The promoter fragments of *PeEIL_18129* and *PeRAF_26475* were labelled with biotin on the ends of the probe (Table S14). The EMSA was performed by incubating the labelled probe with purified PebZIP_27949 proteins at 25 °C for 1.0 h and separating them by 6% native polyacrylamide gel electrophoresis in 0.25× Tris/Borate/EDTA buffer. Non-labelled probes were used as cold competitors.

Acknowledgements

We thank Xiurong Xu (Zhejiang Academy of Forestry), Jinxia Ma (China Bamboo Expo Park), Jinjun Yue (Research Institute of Subtropical Forestry, Chinese Academy of Forestry), and Yongfeng Lou (Jiangxi Academy of Forestry) for the collection of plant materials.

Funding

This research was supported by the National Key Research and Development Program of China [Grant number 2021YFD 2200502] and the National Natural Science Foundation of China [Grant number 31971736].

Conflict of interest

The authors declare no competing interests.

Author contributions

Z. G. and Y. L. designed and supervised this study. Y. L., C. Z., X. Y., Z. L., H. L., X. D., and J. W. performed the experiments, Y. L. and C. Z. performed the data analyses. Y. L. prepared the manuscript, and Z. G. modified the manuscript. All the authors have read and approved the manuscript.

Data availability

The genome resequencing and RNA-Seq data that support the findings of this study have been deposited in the NCBI SRA database under accession numbers PRJNA1065936 and PRJNA1065934, respectively.

References

Alexander, D.H. and Lange, K. (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **18**, 246.

Bai, Y., Cai, M., Mu, C., Zheng, H., Cheng, Z., Xie, Y. and Gao, J. (2023) Integrative analysis of exogenous auxin mediated plant height regulation in moso bamboo (*Phyllostachys edulis*). *Ind. Crop Prod.* **200**, 116852.

Binder, B.M. (2020) Ethylene signaling in plants. *J. Biol. Chem.* **295**, 7710–7725.

Bunsick, M., McCullough, R., McCourt, P. and Lumba, S. (2021) Plant hormone signaling: Is upside down right side up? *Curr. Opin. Plant Biol.* **63**, 102070.

Chen, T. (2013) A new form of *Phyllostachys edulis* (Poaceae: Bambusoideae). *World Bamboo and Rattan* **11**, 25–26 (In Chinese).

Chen, M., Guo, L., Ramakrishnan, M., Fei, Z., Vinod, K.K., Ding, Y., Jiao, C. *et al*. (2022) Rapid growth of Moso bamboo (*Phyllostachys edulis*): cellular roadmaps, transcriptome dynamics, and environmental factors. *Plant Cell* **34**, 3577–3610.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J. *et al*. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. *et al*. (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A. *et al*. (2021) Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008.

Daviere, J.M. and Achard, P. (2013) Gibberellin signaling in plants. *Development* **140**, 1147–1151.

Depuydt, S. and Hardtke, C.S. (2011) Hormone signalling crosstalk in plant growth regulation. *Curr. Biol.* **21**, R365–R373.

Ding, Y., Lin, S., Wei, Q., Yao, W., Que, F. and Li, L. (2022) Advances in developmental biology of bamboo. *J. Nanjing For. Univ.* **46**, 23–40 (In Chinese).

Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X. *et al*. (2017) Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat. Commun.* **8**, 249.

Durbak, A., Yao, H. and McSteen, P. (2012) Hormone signaling in plant development. *Curr. Opin. Plant Biol.* **15**, 92–96.

Fan, C., Ma, J., Guo, Q., Li, X., Wang, H. and Lu, M. (2013) Selection of reference genes for quantitative real-time PCR in bamboo (*Phyllostachys edulis*). *PLoS One* **8**, e56573.

Feng, F. and Li, Y. (2023) China's bamboo resources in 2021. *World Bamboo and Rattan* **21**, 100–103 (In Chinese).

Feng, G., Ai, X., Yi, H., Guo, W. and Wu, J. (2021) Genomic and transcriptomic analyses of Citrus sinensis varieties provide insights into Valencia orange fruit mastication trait formation. *Hortic. Res.* **8**, 218.

Francis, R.M. (2017) Pophelper: an R package and web app to analyse and visualize population structure. *Mol. Ecol. Resour.* **17**, 27–32.

Hershberg, R. and Petrov, D.A. (2010) Evidence that mutation is universally biased towards AT in bacteria. *PLoS Genet.* **6**, e1001115.

Hua, X., Zhou, W. and Lai, G. (2012) *Phyllostachys heterocycla* cv. *Pubescens* f. *porphyrosticta*, a new form from China. *J. Zhejiang Forest. Sci. Technol.* **32**, 73–74 (In Chinese).

Huang, L., Yang, M., Li, L., Li, H., Yang, D., Shi, T. and Yang, P. (2018) Whole genome re-sequencing reveals evolutionary patterns of sacred lotus (*Nelumbo nucifera*). *J. Integr. Plant Biol.* **60**, 2–15.

Kieber, J.J. and Schaller, G.E. (2018) Cytokinin signaling in plant development. *Development* **145**, dev149344.

Kuhn, A., Roosjen, M., Mutte, S., Dubey, S.M., Carrillo Carrasco, V.P., Boeren, S., Monzer, A. *et al*. (2023) RAF-like protein kinases mediate a deeply conserved, rapid auxin response. *Cell* **187**, 130–148.e17.

Lai, G. (2012) A reconsideration of taxonomic position of some infraspecific mutants in the Genus *Phyllostachys* (Bambusoideae). *J. Anhui Agric. Sci.* **40**, 4621–4625 (In Chinese).

Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296.

Li, H. (2013) *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv*, q-bio.GN. 1303.3997.

Li, J. (2020) *Phylogenetic Studies on Phyllostachys Edulis* and its Variation Based on Transcriptome Data. Master Dissertation, Jiangxi Agriculture University.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. *et al*. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Li, L., Shi, Q., Jia, Y., Deng, P. and Gao, J. (2021) Transcriptome and anatomical comparisons reveal the specific characteristics and genes involved in distinct types of growing culms. *Ind. Crop Prod.* **171**, 113865.

Li, Y., Zhang, D., Zhang, S., Lou, Y., An, X., Jiang, Z. and Gao, Z. (2022) Transcriptome and miRNAome analysis reveals components regulating tissue differentiation of bamboo shoots. *Plant Physiol.* **188**, 2182–2198.

Li, L., Zhou, B., Liu, D., Wu, H., Shi, Q., Lin, S. and Yao, W. (2023) Transcriptomic complexity of culm growth and development in different types of moso bamboo. *Int. J. Mol. Sci.* **24**, 7425.

Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

Lin, X., Ruan, X., Lou, Y., Guo, X. and Fang, W. (2009) Genetic similarity among cultivars of *Phyllostachys pubescens*. *Plant Syst Evol* **277**, 67–73.

Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Ma, N., Lai, G., Zhang, P. and Zhang, H. (2014) *The Genus Phyllostachys in China*, pp. 92–107. Hangzhou: Zhejiang Science and Technology Press. (In Chinese).

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. *et al*. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

Nagai, K. and Ashikari, M. (2023) Molecular mechanism of internode elongation in rice. *Breed Sci.* **73**, 108–116.

Nishiyama, N., Shinozawa, A., Matsumoto, T. and Izawa, T. (2023) High genome heterozygosity revealed vegetative propagation over the sea in moso bamboo. *BMC Genomics* **24**, 348.

Ossowski, S., Schneeberger, K., Lucas-Lledo, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D. *et al*. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* **327**, 92–94.

Pan, X., Welti, R. and Wang, X. (2010) Quantitative analysis of major plant hormones in crude plant extracts by high-performance liquid chromatography-mass spectrometry. *Nat. Protoc.* **5**, 986–992.

Peng, Z., Lu, Y., Li, L., Zhao, Q., Feng, Q., Gao, Z., Lu, H. *et al*. (2013) The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* **45**, 456–461, 461e1-2.

Ramakrishnan, M., Yrjälä, K., Vinod, K.K., Sharma, A., Cho, J., Satheesh, V. and Zhou, M. (2020) Genetics and genomics of moso bamboo (*Phyllostachys edulis*): current status, future challenges, and biotechnological opportunities toward a sustainable bamboo industry. *Food Energy Secur.* **9**, e229.

Retief, J.D. (2000) Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.* **132**, 243–258.

Ross, J.J., Weston, D.E., Davidson, S.E. and Reid, J.B. (2011) Plant hormone interactions: how complex are they? *Physiol. Plant.* **141**, 299–309.

Shi, F., Wang, M. and An, Y. (2020) Overexpression of a B-type cytokinin response regulator (*OsORR2*) reduces plant height in rice. *Plant Signal. Behav.* **15**, 1780405.

Simura, J., Antoniadi, I., Siroka, J., Tarkowska, D., Strnad, M., Ljung, K. and Novak, O. (2018) Plant hormonomics: multiple phytohormone profiling by targeted metabolomics. *Plant Physiol.* **177**, 476–489.

Slifer, S.H. (2018) PLINK: key functions for data analysis. *Curr. Protoc. Hum. Genet.* **97**, e59.

Wang, S. (1984) Three new forms of *Phyllostachys pubescens* mazel. *Guihaia* **4**, 319–320 (In Chinese).

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M. *et al*. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49.

Wang, Y., Sun, X., Ding, Y., Fei, Z., Jiao, C., Fan, M., Yao, B. *et al*. (2019) Cellular and molecular characterization of a thick-walled variant reveal a pivotal role of shoot apical meristem in transverse development of bamboo culm. *J. Exp. Bot.* **70**, 3911–3926.

Wang, K.L., Zhang, Y., Zhang, H.M., Lin, X.C., Xia, R., Song, L. and Wu, A.M. (2021) MicroRNAs play important roles in regulating the rapid growth of the *Phyllostachys edulis* culm internode. *New Phytol.* **231**, 2215–2230.

Wang, H., Shemesh-Mayer, E., Zhang, J., Gao, S., Zeng, Z., Yang, Z., Zhang, X. *et al*. (2023) Genome resequencing reveals the evolutionary history of garlic reproduction traits. *Hortic. Res.* **10**, uhad208.

Yang, G., Li, Z., Du, T., Hu, Y., Leng, D. and Zhou, W. (1997) A new cultivated variety of moso bamboo. *Acta Agriculturae Universitatis Jiangxiensis* **19**, 97–98 (In Chinese).

Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82.

Zeng, Q., Wang, H., Yu, L., Cheng, P. and Peng, J. (2016) Ornamental value evaluation of singular ornamental bamboo including *Phyllostachys edulis* cv. Tao kiang. *World Bamboo and Rattan* **14**, 12–13 (In Chinese).

Zhang, P. (2008) New type of moso in Anji County, Zhejiang Province. *World Bamboo and Rattan* **6**, 27 (In Chinese).

Zhang, P., Lai, G. and Zhang, H. (2012) Two new forms of *Phyllostachys edulis* (Gramineae: Bambusoideae) from Anji, Zhejiang. *World Bamboo and Rattan* **10**, 32–33 (In Chinese).

Zhao, H., Yang, L., Peng, Z., Sun, H., Yue, X., Lou, Y., Dong, L. *et al*. (2015) Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rep.* **5**, 8018.

Zhao, H., Gao, Z., Wang, L., Wang, J., Wang, S., Fei, B., Chen, C. *et al*. (2018) Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *Gigascience* **7**, giy115.

Zhao, H., Sun, S., Ding, Y., Wang, Y., Yue, X., Du, X., Wei, Q. *et al*. (2021) Analysis of 427 genomes reveals moso bamboo population structure and genetic basis of property traits. *Nat. Commun.* **12**, 5466.

Zou, Z. and Zhang, J. (2021) Are nonsynonymous transversions generally more deleterious than nonsynonymous transitions? *Mol. Biol. Evol.* **38**, 181–191.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Morphological characteristics comparison of *Phyllostachys edulis* (Mao Zhu in Chinese, MZ) forms and MZ populations grown in different geographic regions.
Figure S2 The distribution of SNPs annotated among 20 MZ forms.
Figure S3 The genes containing SNP with high-level effect distributed on different chromosomes.
Figure S4 Enrichment analysis of genes contained SNP with high-level effect.
Figure S5 The density of two types of heterozygous sites on chromosomes within non-overlapping windows of 200 kb length.
Figure S6 Enrichment analysis of genes located in long continuous heterozygous SNP clustered regions (LCHRs).
Figure S7 The CV error value for each model *K*.
Figure S8 The admixture results using different model *K* values (two to six) for 427 MZ individuals from 15 geographic regions and 20 MZ forms.
Figure S9 The scatter plot of the 1st, 2nd, and 3rd principal components (PCs) in the PC analysis of 427 MZ individuals from 15 geographic regions and 20 MZ forms.
Figure S10 Samples collected from MZ, SY, and GJ at different developmental stages for RNA-Seq.
Figure S11 Number of differentially expressed genes (DEGs) between different samples.
Figure S12 The number of different expressed transcription factors identified in RNA-Seq data of *P. edulis* f. 'Kikko-chiku' (Gui Jia zhu in Chinese, GJ) and *P. edulis* f. *tubaeformis* (Sheng Yin zhu in Chinese, SY) at different developmental stages.
Figure S13 Expression patterns of differentially expressed transcription factors (TFs) in GJ, SY, and MZ shoots at different developmental stages.
Figure S14 Expression patterns of differentially expressed transcription regulators (TRs) in GJ, SY, and MZ shoots at different developmental stages.
Figure S15 Expression patterns of differentially expressed protein kinases (PKs) in GJ, SY, and MZ shoots at different developmental stages.
Figure S16 RT-qPCR analysis of the expression of the randomly selected 10 DEGs in GJ and GM shoots.
Figure S17 RT-qPCR analysis of the expression of the randomly selected 10 DEGs in SY and SM shoots.
Figure S18 Validation of expression levels of the randomly selected 10 DEGs in 48 shoot samples including 21 GJ samples and 27 SY samples.
Figure S19 Expression patterns of *PeEIL_18129*, *PeRAF_26475*, and *PebZIP_27949* in MZ seedlings treated with NAA or ETH.
Figure S20 A proposed schematic model exhibited PebZIP_27949 regulating auxin and ethylene signalling for culm shape variation.
Table S1 The statistical analysis of morphological investigation of 799 bamboo culms from three different geographic regions.
Table S2 The statistical analysis of morphological investigation of 458 MZ individuals from 16 different geographic regions.
Table S3 List of 20 MZ forms used in this study.
Table S4. The statistics of SNPs and InDels identified from 20 MZ forms.
Table S5 The statistical analysis of SNP annotation.
Table S6 The summary of SNP identified from 20 MZ forms.
Table S7 The statistics of total and linkage disequilibrium pruned SNPs in the merged datasets.
Table S8 The statistics of heterozygous SNP identified in 20 MZ forms.
Table S9 The statistics of heterozygous SNP identified in 427 MZ individuals from 15 geographic regions.
Table S10 The information of heterozygous SNPs associated with plant hormone signal transduction genes.
Table S11 The statistical analysis of 66 shoot samples for RNA-Seq.
Table S12 Annotation of differentially expressed TFs, TRs, and PKs.
Table S13 Prediction of bZIP binding sites in the promoters of genes co-expressed with *PebZIP_27949*.
Table S14 List of primers used in this study.