Tree Genetics and Molecular Breeding 2022, Vol.12, No.6, 1-11
http://genbreedpublisher.com/index.php/tgmb

Research Article Open Access

Genomic Sequence Analysis of 4 Culm Shape Variants of Moso Bamboo Based on Re-sequencing

Shaohua Mu [1], Xianghua Yue [2], Juan Li [1], Xueping Li [1], Jian Gao [1]

1 Beijing Key Lab for Bamboo & Rattan Science and Technology, International Center for Bamboo and Rattan, National Forestry and Grassland Administration, Beijing, 100102, China

2 Huangshan Taiping Experimental Station, International Center for Bamboo and Rattan, Huangshan, 245716, China

Corresponding author email: mush@icbr.ac.cn

Tree Genetics and Molecular Breeding, 2022, Vol.12, No.6 doi: 10.5376/tgmb.2022.12.0006

Received: 06 May, 2022

Accepted: 14 May, 2022

Published: 21 May, 2022

Preferred citation for this article:

Mu S.H., Yue X.H., Li J., Li X.P., and Gao J., 2022, Genomic sequence analysis of 4 culm shape variants of moso bamboo based on re-sequencing, Tree Genetics and Molecular Breeding, 12(6): 1-11 (doi: 10.5376/tgmb.2022.12.0006)

Abstract To reveal the whole genome differences of moso bamboo variations on culm shape, *Phyllostachys edulis* f. *obliquinoda*, *Phyllostachys edulis* f. *heterocycla*, *Phyllostachys pubescens* f. *curviculmi*, *Phyllostachys edulis* f. *tubaeformis* were re-sequenced by whole genome re-sequencing, to detect and annotate the single nucleotide polymorphism (SNP), insertion-deletion (InDel) and structure variation (SV). The mutant genes were compared with the functional databases of GO, COG, KEGG, of which 7 477, 7 584, 7 221 and 7 583 mutant genes were annotated respectively. GO annotation classification included 56 functional groups of three functional classification systems: cellular component, molecular function and biological processes. COG classification showed that 369 genes involved in replication, recombination and repair, 327 genes involved in signal transduction mechanism, and 253 genes involved in transcription. The metabolic pathways of flavonoids, carotenoids and other substances involved in the mutant genes were analyzed by KEGG database. In-depth study of the regulatory pathways can provide a theoretical basis for further exploration of the rich polymorphism and genetic variation of moso bamboo, and the development of molecular markers for some significant characters.

Keywords Moso bamboo (*Phyllostachys edulis*) variant; Whole genome re-sequencing; Single nucleotide polymorphism; Small Indel; Structure variation

Moso bamboo (*Phyllostachys edulis*) is a unique economic bamboo species with the widest cultivation area and a long history of utilization in China. Moso bamboo is distributed in 16 provinces and regions in southern China, spanning three subtropical climates: north, middle and south. In the process of long-term adaptation to different growth environments, cultivation and management and genetic drift, moso bamboo has undergone a series of morphological variations, and 21 moso bamboo variants have been found (Ma et al., 2014). In recent years, with the completion of the whole genome sequencing of moso bamboo (Peng et al., 2013) and the application of molecular biology and omics technology in bamboo plants, the research on bamboo plant genome has made rapid progress. At present, there have been reports on the identification and functional verification of gene families such as *SAUR* (Bai et al., 2016), TCP (Liu et al., 2018) and *Hsp* (Xie et al., 2019), but no research on the genome sequence of moso bamboo variants has been found.

The culm shape variants of moso bamboo showed rich polymorphism in culm shape traits. In order to study the causes, it is very necessary to reveal the degree of variation and genetic diversity of moso bamboo at the gene level. High throughput sequencing technology makes it possible to analyze the genome of a species in detail. Genetic variation analysis using this technology has been widely used in millet (Bai et al., 2013; Jia et al., 2019), rice (Takagi et al., 2013), soybean (Qi et al., 2014) and potato (Yin et al., 2020).

Taking the whole genome of moso bamboo as a reference, 4 culm shape variants of moso bamboo were re-sequenced by using economical and mature second-generation sequencing technology, and their single

nucleotide polymorphism (SNP), small Indel (Indel) and structure variation (SV) were deeply excavated to obtain genome sequence data and reveal the mutation types and variation genes of the whole genome, which lays a foundation for developing molecular markers for important traits of moso bamboo and exploring rich genetic polymorphism within species.

1 Results and Analysis

1.1 Genome resequencing analysis

The sequencing data of 4 variants of moso bamboo were obtained by high-throughput sequencing (Table 1). '*Phyllostachys edulis* f. *heterocycla*' has the least Clean reads, with 81 981 057 bp; '*Phyllostachys pubescens* f. *obliquinoda*' has the most Clean reads, with 85 058 968 bp. All sequences were compared with the reference genome of moso bamboo. The percentage of 4 sample data mapped to the reference genome of moso bamboo was more than 98.75% of all Clean reads, and the properly_mapped was more than 86.86%. The average coverage depth of the sample is about 10× (Table 1). The quality of sample sequencing data could meet the needs of subsequent analysis.

Table 1 Sequencing data statistics of 4 *Phyllostachys edulis* variants

Name	Clean_reads	Mapped (%)	Properly_mapped (%)	Ave_ depth	Cov_ratio_ 1X (%)	Cov_ratio_ 5X (%)	Cov_ratio_ 10X (%)
Phyllostachys pubescens f. *obliquinoda*	85 058 968	99.71	86.86	10	98.10	86.67	54.10
Phyllostachys edulis f. *heterocycla*	81 981 057	99.57	89.14	10	97.70	83.42	50.31
Phyllostachys pubescens f.*curviculmi*	83 081 813	99.17	87.63	8	92.17	65.14	42.01
Phyllostachys edulis f. *tubaeformis*	83 065 300	98.75	87.26	10	97.52	82.30	48.82

1.2 Detection and annotation of SNP

1.2.1 SNP detection

SNP detection was carried out on the genomic DNA of the sample, and the SNP data were counted. The results showed that (Table 2), '*Phyllostachys pubescens* f. *obliquinoda*' had the largest number of SNPs, with 1 654 319. '*Phyllostachys pubescens* f. *curviculmi'* had the least number of SNPs, with 1 454 574. Among the 4 samples, the ratio of the number of transition (Ti) SNPs to the number of transition (Tv) SNPs (Ti/Tv) was about 3. Among them, '*Phyllostachys pubescens* f. *curviculmi*' had the highest Ti/Tv value, which was 3.17. '*Phyllostachys pubescens* f. *obliquinoda*' had the lowest Ti/Tv value, which was 2.98. The number of heterozygosity (Het) SNPs was about 10 times that of homozygosity (Homo) SNPs, and the Het-ratio were 91.51% (R01)>91.06% (R02)>90.86% (R04)>87% (R03) respectively. The higher the Het-ratio, the higher the degree of heterozygosity of the samples.

Table 2 Statistics of SNP data of 4 *phyllostachys edulis* variants

Name	ID	SNP	Ti	Tv	Ti/Tv	Het	Homo	Het-ratio (%)
Phyllostachys pubescens f. *obliquinoda*	R01	1 654 319	1 239 666	414 653	2.98	1 513 942	140 377	91.51
Phyllostachys edulis f. *heterocycla*	R02	1 640 317	1 233 040	407 277	3.02	1 493 683	146 634	91.06%
Phyllostachys pubescens f.*curviculmi*	R03	1 454 574	1 106 090	348 484	3.17	1 265 561	189 013	87.00%
Phyllostachys edulis f. *tubaeformis*	R04	1 591 930	1 196 090	395 840	3.02	1 443 633	148 297	90.68%

1.2.2 SNP annotation

SnpEff software was used to annotate the genome-wide SNP of moso bamboo variant to obtain the region or type of its variant ectopic site in the genome (Figure 1). The number of SNPs in the coding region of the 4 moso bamboo variants accounted for about 2%, and the proportion of non-synonymous mutations in '*Phyllostachys edulis* f. *heterocycla*' was the highest, about 51.64%; Followed by '*Phyllostachys pubescens* f. *obliquinoda*' and '*Phyllostachys edulis* f. *tubaeformis*', the proportion of non-synonymous mutations were 51.51% and 51.44% respectively. The proportion of non-synonymous mutations in '*Phyllostachys pubescens* f. *curviculmi*' was the lowest, which was 50.78%. SNPs with non-synonymous mutations in the coding region were easy to lead to changes in gene function and morphological traits. Among the 4 moso bamboo variants, the culm shape variation

of *Phyllostachys edulis* f. *heterocycla* was the largest, and the proportion of non-synonymous mutations was consistent with the change degree of bamboo culm shape.

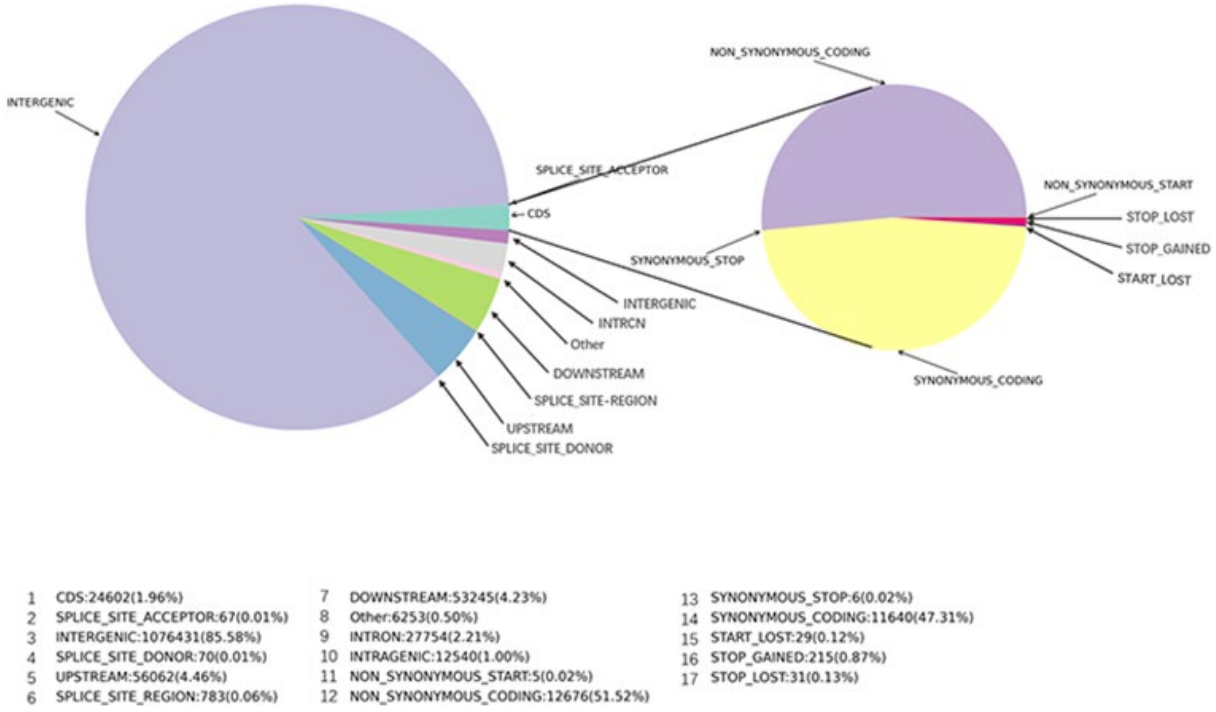

Figure 1 SNP annotations pie of *Phyllostachys edulis* f. *obliquinoda*

Note: 1: CDS; 2: Splice_site_acceptor; 3: Intergenic; 4: Splice_site_donor; 5: Upstream; 6: Splice_site_region; 7: Downstream; 8: Other; 9: Intron; 10: Intragenic; 11: Non_synonymous_start; 12: Non_synonymous_coding; 13: Synonymous_stop; 14: Synonymous_coding; 15: Start_lost; 16: Stop_gained; 17: Stop_lost

1.3 Detection and annotation of Indel

1.3.1 Indel detection

The total number of Indels of the four samples detected in the whole genome ranged from 256 254 to 286 888 (Table 3), in which the total number of insertion mutations was slightly lower than that of deletion mutations; The total number of Indels detected in the coding region was 4 575~4 744, of which the total number of insertion mutations was about 2/3 of the deletion mutations. In each sample, the number of homozygous mutations in the whole genome was about twice that of heterozygous mutations, and the number of homozygous mutations in the coding region was slightly lower than that of heterozygous mutations. The number of Indel in the coding region of the 4 samples were: '*Phyllostachys edulis* f. *heterocycla*' (4744) > '*Phyllostachys edulis* f. *tubaeformis*' (4706) > '*Phyllostachys pubescens* f. *curviculmi*' (4618) > '*Phyllostachys pubescens* f. *obliquinoda*' (4575), and the change law was consistent with the number of SNPs.

Table 3 Total genome and coding region InDel statistics of 4 phyllostachymus edulis variants

Name	CDS					Genome				
	Insertion	Deletion	Het	Homo	Total	Insertion	Deletion	Het	Homo	Total
Phyllostachys pubescens f. *obliquinoda*	2 873	1 702	2 487	2 088	4 575	140 920	145 968	9 1670	195 218	286 888
Phyllostachys edulis f. *heterocycla*	2 997	1 747	2 551	2 193	4 744	141 252	145 000	93 712	192 540	286 252
Phyllostachys pubescens f. *curviculmi*	2 973	1 645	2 584	2 034	4 618	127 110	129 144	95 846	160 408	256 254
Phyllostachys edulis f. *tubaeformis*	2 966	1 740	2 567	2 139	4 706	136 531	140 228	93 358	183 401	276 759

1.3.2 Indel annotation

By comparing with the gene and location information of moso bamboo genome, the specific location of Indel sites in four samples was annotated, the problem that whether they occur in intergenic region, gene region or CDs region was analyzed, the information whether the mutation type is frameshift mutation was judged, and the Indels of frameshift mutation that may lead to the change of gene function was focused on. The annotation results showed (Figure 2) that the Indels of the frameshift mutation in the CDs region of the 4 samples accounts for about 54%, of which, '*Phyllostachys edulis* f. *heterocycla*' (54.87%) > '*Phyllostachys pubescens* f. *obliquinoda*' (54.63%) > '*Phyllostachys edulis* f. *tubaeformis*' (54.56%) > '*Phyllostachys pubescens* f. *curviculmi*' (53.97%). The change law of Indels of the frameshift mutation in the CDs region of the 4 samples was consistent with the proportion of SNP with non-synonymous mutation.

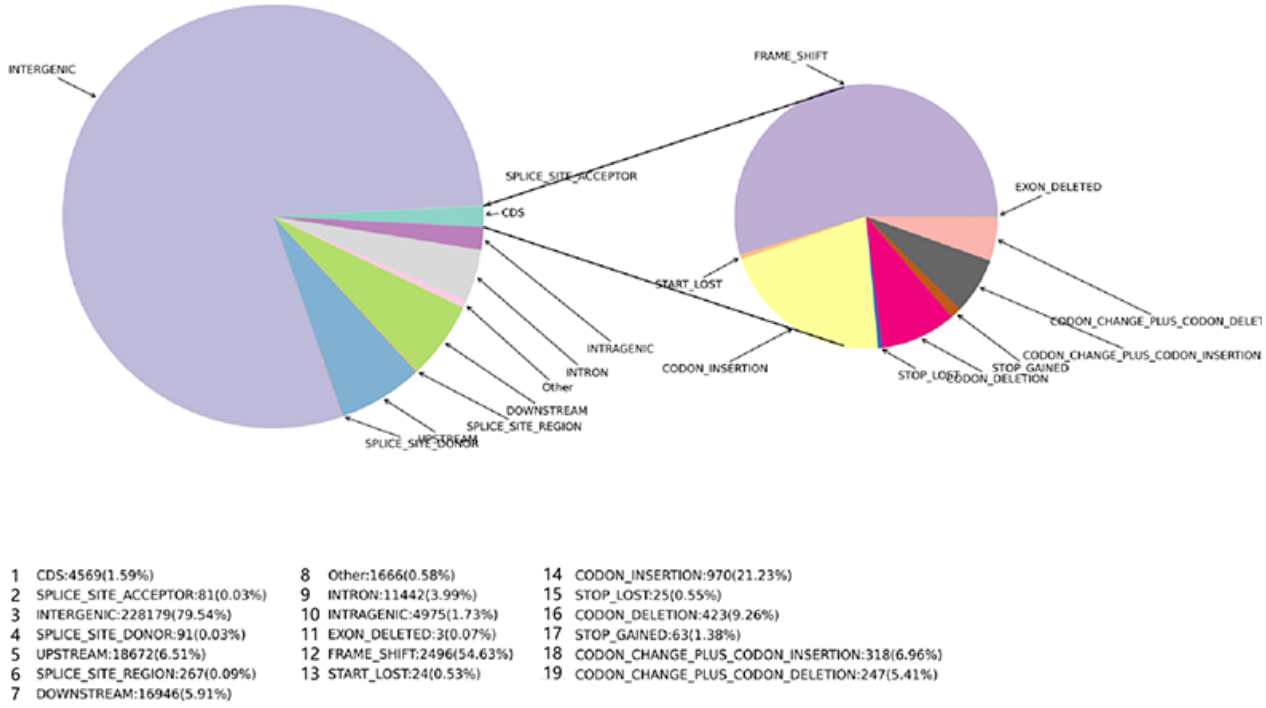

Figure 2 InDel annotations pie of *Phyllostachys edulis* f. *obliquinoda*

Note: 1: Cds; 2: Splice_site_acceptor; 3: Intergenic; 4: Splice_site_region; 5: Upstream; 6: Splice_site_region; 7: Downstream; 8: Other; 9: Intron; 10: Intragenic; 11: Exon_ delected; 12: Frame_shift; 13: Start_lost; 14: Codon_insertion; 15: Stop_lost; 16: Codon_deletion; 17: Stop_gained; 18: Codon_change_plus_insertion; 19: Codon_change_plus_deletion

1.4 Detection and annotation of SV

1.4.1 SV detection

According to the statistics of the number of SV of each type in the 4 samples detected by BreakDancer software (Table 4), the number of SV loci of '*Phyllostachys pubescens* f.*curviculmi*' was the largest, with 79 276, of which the proportion of insertion and deletion variation in the total variation was 59.98%, and the proportion of inter-chromosomal translocation variation was 31.12%. The number of SV loci of '*Phyllostachys edulis* f. *tubaeformis*' was the least, with was 68 377, of which the proportion of insertion and deletion variation in the total variation was 65.38% and the proportion of inter-chromosomal translocation variation was 26.63%. The number of SV loci in '*Phyllostachys pubescens* f. *obliquinoda*' was 78 394, and the proportion of insertion and deletion variation in the total variation was 76.01%, of which the number of insertion types was the largest, with 26 456. The number of SV loci in '*Phyllostachys edulis* f. *heterocycla*' was 75 866, and the proportion of insertion and deletion variation in the total variation was 61.88%, of which the number of insertion types was the least, with 13084. The variation law of SV of 4 moso bamboo variants was not inconsistent with SNP and Indel.

Table 4 Statistics on the number of structural variations of 4 *phyllostachys edulis*

Name	SV total	Insertion	Deletion	Inversion	Intra-chromosomal translocation	Inter-chromosomal translocation	Other
Phyllostachys pubescens f. *obliquinoda*	78 394	26 456	33 132	486	4 537	13 681	102
Phyllostachys edulis f. *heterocycla*	75 866	13 084	33 864	685	4 707	23 417	109
Phyllostachys pubescens f.*curviculmi*	79 276	17 224	30 324	2551	4 392	24 668	117
Phyllostachys edulis f. *tubaeformis*	68 377	13 707	30 999	602	4 731	18 210	128

1.4.2 SV annotation

By comparing with the gene location information of the reference genome, the information of SV occurring in exon region, intron region and intergenic region was annotated, and the structural variation annotation of deletion, insertion and inversion was annotated. The statistical results showed that (Table 5), the SV variation of the 4 samples mainly occured in the intergenic region, followed by the exon region, and the number in intron regions was the least; Among the 3 types of structural variation annotation, the number of deletion was much higher than the other two types, followed by insertion, and the number of inversion was the least. Among the 4 moso bamboo variants, the number of deletion of *Phyllostachys edulis* f. *heterocycla* was the most, the number of insertion of *Phyllostachys pubescens* f. *obliquinoda* was the most, and the number of deletion of *Phyllostachys pubescens* f. *curviculmi* was the least.

Table 5 Statistics of the annotation results of structural variation of 4 *Phyllostachys edulis*

Name	Exon			Intron			Intergenic		
	Deletion	Insertion	Inversion	Deletion	Insertion	Inversion	Deletion	Insertion	Inversion
Phyllostachys pubescens f. *obliquinoda*	2 361	1 670	127	903	489	15	29 868	24 297	344
Phyllostachys edulis f. *heterocycla*	2 322	563	138	987	278	15	30 555	12 243	532
Phyllostachys pubescens f.*curviculmi*	1 865	831	236	879	394	66	27 580	15 999	2249
Phyllostachys edulis f. *tubaeformis*	2 210	640	132	895	250	20	27 894	12 817	450

1.5 Mining variant genes and annotating their functions

1.5.1 Excavating variant genes

SNPs with non-synonymous mutations and genes with Indel and SV in CDs region in 4 variants of moso bamboo were extracted and statistically analyzed (Table 6). The results showed that there were 12 921 variant genes in '*Phyllostachys pubescens* f. *obliquinoda*', including 5 349 non-synonymous mutant SNP genes, 3 832 Indel genes and 3 740 SV mutant genes. The total number of differential genes was the largest. The number of variant genes in '*Phyllostachys edulis* f. *heterocycla*' and '*Phyllostachys edulis* f. *tubaeformis*' was close, which were 11 967 and 11 916 respectively. Among them, there were 5 419 and 5 406 non-synonymous mutant SNP genes, 3 938 and 3 901 Indel genes, and 2 610 and 2 609 SV mutant genes, respectively. There were 11 467 mutated genes in '*Phyllostachys pubescens* f. *curviculmi*', including 5 078 non-synonymous mutated SNP genes, 3 797 Indel genes and 2 592 SV mutated genes. The total number of differential genes was the least.

Table 6 Statistical analysis of variation genes of 4 *Phyllostachys edulis variants*

Name	Genes with non-synonymous SNP	Genes with InDel	Genes with SV	Total
Phyllostachys pubescens f. *obliquinoda*	5 349	3 832	3 740	12 921
Phyllostachys edulis f. *heterocycla*	5 419	3 938	2 610	11 967
Phyllostachys pubescens f.*curviculmi*	5 078	3 797	2 592	11 467
Phyllostachys edulis f. *tubaeformis*	5 406	3 901	2 609	11 916

1.5.2 Annotating the function of variation genes

The variation genes of the 4 moso bamboo variants were compared with the functional databases such as GO, KEGG and COG respectively. It was found that the number of variation genes annotated into the database were 7 477, 7 584, 7 221 and 7 583 respectively. The GO classification statistical results of variation genes showed the number of genes corresponding to 56 functional groups of the three gene function classification systems, namely biological processes, cellular component, and molecular function, and the percentage of gene number. It could be seen from the GO annotated classification map of '*Phyllostachys pubescens* f. *obliquinoda*' variant genes that in the cell component classification system, there were 444 genes related to cell wall and 217 genes related to cytoskeleton (Figure 3); In the biological process classification system, there were 103 genes related to carotenoid synthesis; In the functional gene classification system, there were 546 genes related to signal transduction, 25 transcription factors and 104 genes related to hormone synthesis. There were 686 genes related to cell wall and 76 genes related to cytoskeleton related genes in '*Phyllostachys edulis* f. *heterocycla*'. There were 671 genes related to cell wall and 72 genes related to cytoskeleton in '*Phyllostachys pubescens* f.*curviculmi*'. There were 259 genes related to cell wall and 186 genes related to cytoskeleton in '*Phyllostachys edulis* f. *tubaeformis*'. The genes closely related to culm shape, such as cell wall and cytoskeleton, were quite different among the 4 moso bamboo variants.

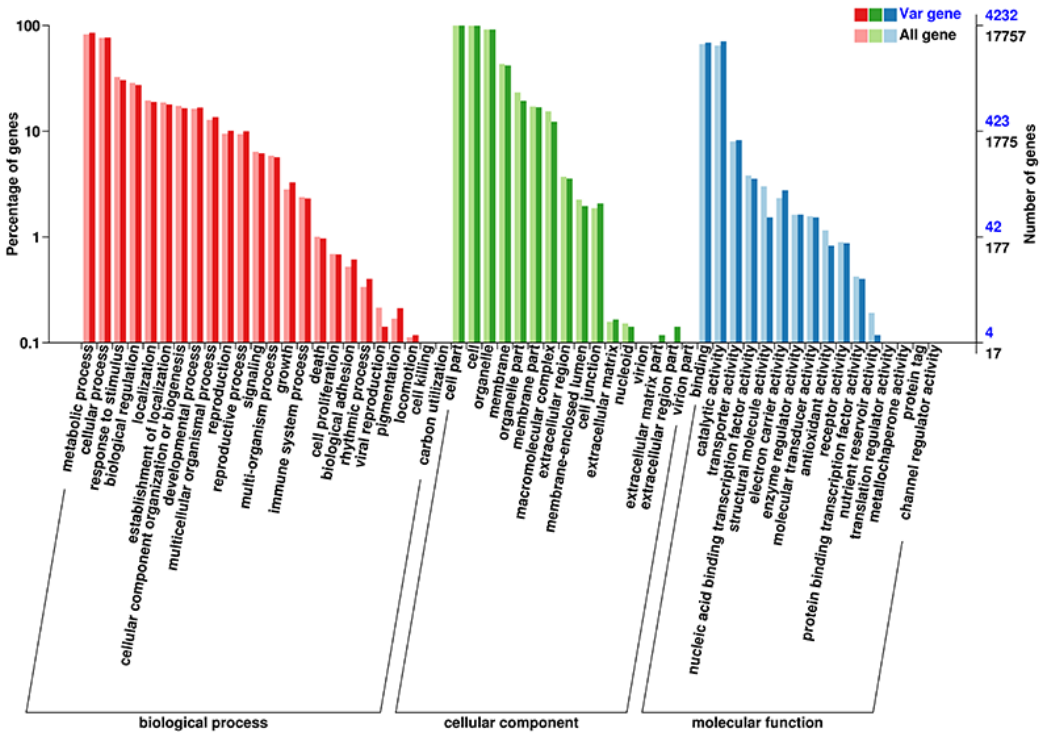

Figure 3 Classification of *Phyllostachys edulis* f. *obliquinoda* gene variations compared with GO database by blast

Note: 1: Metabolic process; 2: Cellular process; 3: Response to stimulus; 4: Biological regulation; 5: Localization; 6: Establishment of localization; 7: Cellular component organization or biogenesis; 8: Developmental process; 9: Multicellular organismal process; 10: Reproduction; 11: Reproductive process; 12: Signaling; 13: Multi-organism process; 14: Growth; 15: Immune system process; 16: Death; 17: Cell proliferation; 18: Biological adhesion; 19: Rhythmic process; 20: Viral reproduction; 21: Pigmentation; 22: Locomotion; 23: Cell killing; 24: Carbon utilization; 25: Cell part; 26: Cell; 27: Organelle; 28: Membrane; 29: Organelle part; 30: Membrane part; 31: Macromolecular complex; 32: Extracellular region; 33: Membrane-enclosed lumen; 34: Cell junction; 35: Extracellular matrix; 36: Nucleoid; 37: Virion; 38: Extracellular matrix part; 39: Extracellular region part; 40: Virion part; 41: Binding; 42: Catalytic activity; 43: Transporter activity; 44: Nucleic acid binding transcription factor activity; 45: Structural molecule activity; 46: Electron carrier activity; 47: Enzyme regulator activity; 48: Molecular transducer activity; 49: Antioxidant activity; 50: Receptor activity; 51: Protein binding transcription factor activity; 52: Nutrient reservoir activity; 53: Translation regulator activity; 54: Metallochaperone activity; 55: protein tag; 56: Channel regulator activity

It could be seen from the COG annotation classification map of variant genes that the number of genes corresponding to different functional classifications varied greatly, and the number of genes corresponding to functional classifications such as transcription, replication, recombination and repair and signal transduction mechanism was large (Figure 4). The number of genes involved in replication, recombination and repair obtained by COG annotation was 369, the number of genes related to transcription was 253, and the number of genes involved in signal transduction mechanism was 327.

Through KEGG database, we could study the genes and expression information of 4 moso bamboo variants from an overall network, and systematically analyze various metabolic pathways and functions of gene products in cells. Taking the plant hormone signal transduction pathway of the variant genes of '*Phyllostachys pubescens* f. *obliquinoda*' as an example (Figure 5), it was noted that 239 genes were involved in this pathway, including 51 variant genes. The whole pathway was formed by different enzymes through complex biochemical reactions. The red box represented that the variant genes were related to this pathway.

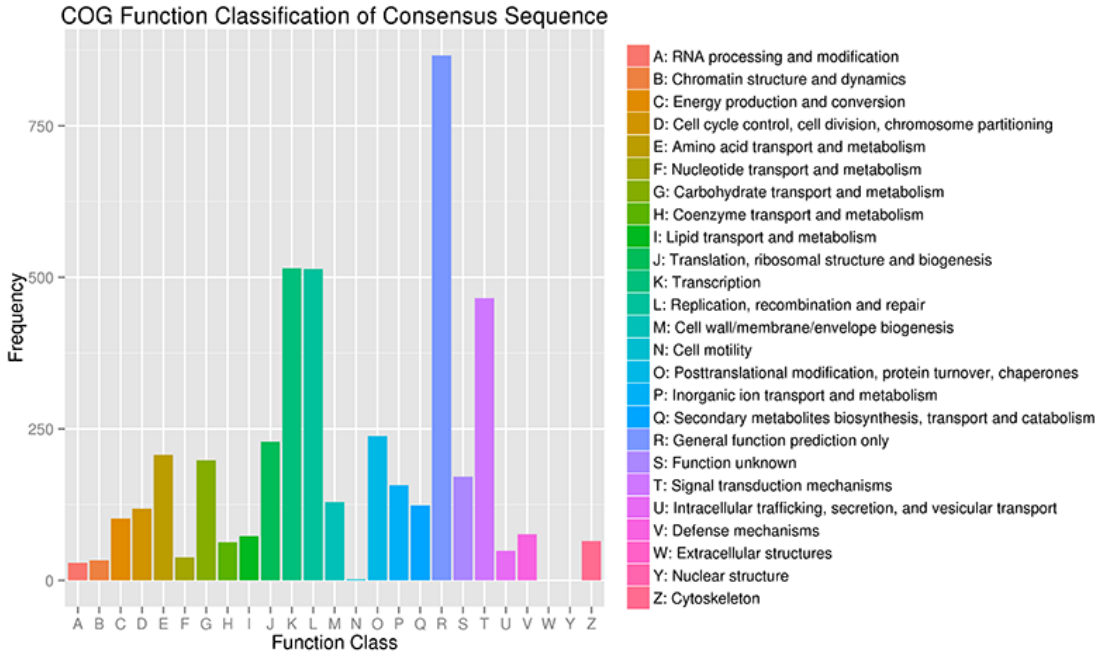

Figure 4 Classification of *Phyllostachys edulis* f.*obliquinoda* gene variations compared with COG database

Note: A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; Y: Nuclear structure; Z: Cytoskeleton

2 Discussion

Moso bamboo was an important economic bamboo species in China. It has many uses, such as timber, edible, ornamental and so on. Its cultivation and management has a long history. Due to long-term natural evolution, geographical isolation and differentiation and artificial cultivation, moso bamboo species produce rich genetic variation, and the morphology or color of the culms or branches and other organs change, thus showing unique characteristics (Cheng, 2006, Science Times, (11): 76-77). These moso bamboo variant types have high ornamental value and are more suitable for landscaping, and become new species resources in landscape garden construction (Shi et al., 2006, Journal of Sichuan Forestry Science and Technology, 27(1): 67-71).

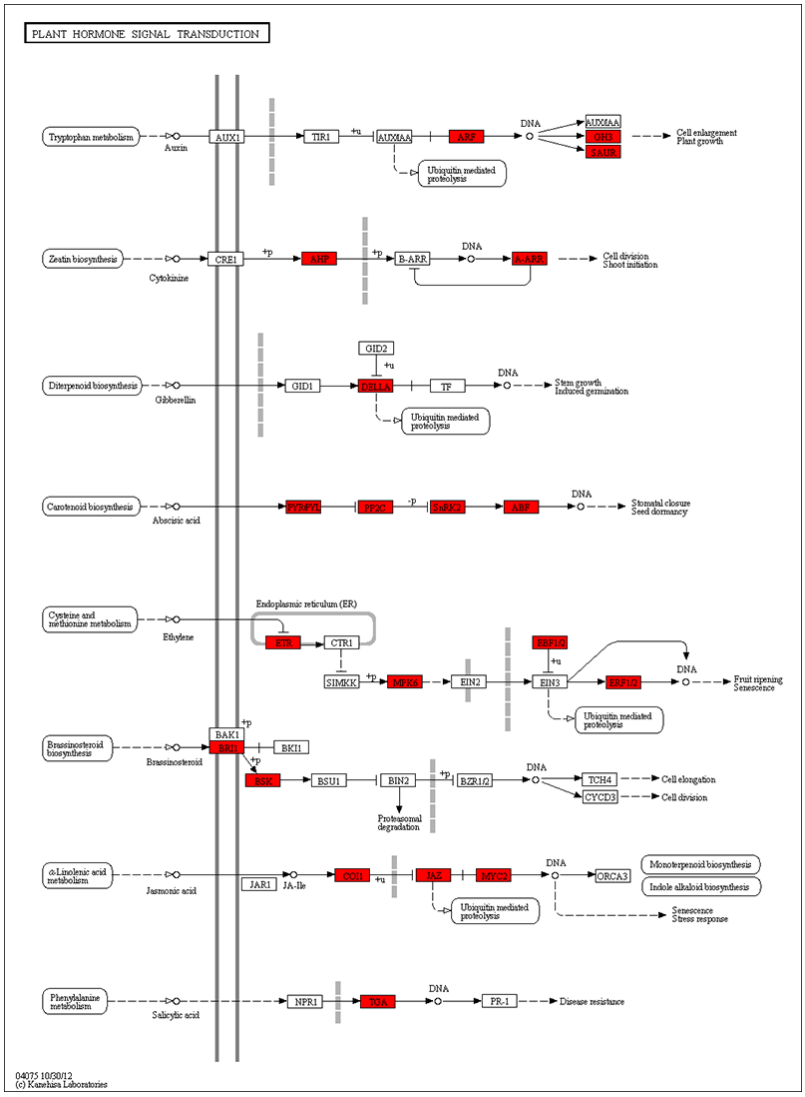

Figure 5 Flavonoid biosynthesis pathway of *Phyllostachys edulis* f. *obliquinoda* gene variations compared with KEGG database by blast

The 4 variant types of moso bamboo were re-sequenced, the molecular data of culm shape variation were statistically analyzed, and the SNP, InDel and SV were detected and annotated. The variation of SNP type could be divided into transformation (Ti) and transversion (Tv). The ratio of transformation/transversion (Ti/Tv) of the samples was about 3, indicating that the transformation type was more likely to occur than the transversion type. The proportion of heterozygosity SNP sites in the total number of SNP sites (heterozygosity ratio) was about 90%, indicating that the degree of heterozygosity of the sample was very high, and the proportion of SNP sites on homologous chromosomes containing different types of bases was high. The number of InDel loci reflected the difference between the samples and the reference genome of moso bamboo. The InDel in the coding region could cause frameshift mutation, resulting in changes in gene function. The number of structure variation of SV deletion, insertion and inversion reflected the deletion, insertion and inversion of large fragments at the genomic level.

The variation in CDS region may cause the change of gene function. By comparing and annotating the non-synonymous mutation SNP and the variation genes with InDel and SV in CDs region in 4 samples with GO, COG, KEGG and other functional databases, we could analyze the function of the gene. Go database annotation clustering reflected 56 functional groups of the three gene function classification systems, namely biological processes, cellular component, and molecular function, and comprehensively described the number of genes and the attributes of gene products in the samples. The COG database annotated the direct homologous classification

of gene products. The number of genes corresponding to different functional classifications varied greatly, reflecting the differences of metabolic or physiological bias corresponding to various moso bamboo variants. KEGG database could form a whole network of genes and expression information. The whole pathway was formed by many different enzymes through complex biochemical reactions. By marking the enzyme number in the box, it could be explained that the corresponding variant genes were related to this pathway.

At present, there are some research advances in the biological, physiological characteristics and molecular markers of different variants of moso bamboo. The results showed that there were significant differences in five physiological indexes such as the net photosynthetic rate, chlorophyll content, β carotene content, nitrate reductase activity and free amino acid content (Chen et al., 2011), and the physiological index value of *Phyllostachys pubescens* f. *obliquinoda* was larger than that of *Phyllostachys edulis* f. *heterocycla* (Yan, 2011). Molecular genetic analysis found that the degree of genetic variation between variants was small (Ruan, 2008). In addition, the research results of culm shape variation of other bamboo species such as *Bambusa ventricosa cv. Nana* can provide a useful reference for the study of culm shape variation of moso bamboo. It was found that the main causes of culm shape variation of *Bambusa ventricosa cv. Nana* were the effects of environmental factors and plant auxin, in which environmental factors can affect the synthesis and distribution of plant auxin (Chen et al., 2008). The research on the culm of *Pseudosasa japonica* showed that the growth and development of its underground culm is regulated by the molecular network composed of hormones, transcription factors and their downstream functional genes; In the initial stage of internode elongation, the genes related to gibberellin synthesis and signal transduction were the highest; In the rapid growth stage of internodes, the expression of downstream function related genes of cell wall and cytoskeleton was the highest (Wei and Ding, 2017). Therefore, based on the research results of other bamboo species, further analysis of the genes related to variation of moso bamboo variants involved in cellular component, biological processes and molecular function was helpful to study the molecular regulation mechanism of the culm shape variation of moso bamboo. In addition, it should be pointed out that the phenotypic variation of moso bamboo culm was a very complex problem. For example, the variation of *Phyllostachys edulis* f. *heterocycla* usually occurs only in the internodes about 2 meters below the bamboo culm (some individuals can exceed 4 meters), and the development of the upper internodes is normal, which indicates that the morphological variation of bamboo culm is genetically unstable. Therefore, the culm shape variation of moso bamboo is a complex biological process, and its molecular regulation mechanism needs to be further studied.

3 Materials and Methods

3.1 Materials

The experimental materials were 4 variants or varieties of moso bamboo (*Phyllostachys edulis*) (Table 7), which were collected from Anhui Taiping Experimental Center of International Center for Bamboo and Rattan (118.13°E, 30.28°N). The young leaves that have just grown and have not yet expanded were selected, and stored at ultra-low temperature after quick freezing with liquid nitrogen.

Table 7 Sample information of 4 variations of moso bamboo

Variation	Morphological characteristics
Phyllostachys edulis f. *obliquinoda*	Adjacent nodes askew or lower middle culm curved
Phyllostachys edulis f. *heterocycla*	Adjacent nodes of the lower middle culm connected one side, opposite tortoiseshell
Phyllostachys edulis f. *curviculmi*	Culm S curved
Phyllostachys edulis f. *tubaeformis*	Culm gradually bulged into a horn shape toward base, internodes shortened

3.2 DNA extraction and genome sequencing

Genomic DNA was extracted by the improved CTAB method (Zidani, 2005). After passing the test, it was broken by ultrasonic, the fragmented DNA was purified, the end was repaired, A and a sequencing connector were added at the 3' end, the fragment size and PCR enrichment were selected, and the sequencing library was constructed. After passing the quality inspection, it was sequenced on the Illunima Hiseq 2500 platform to obtain the original

sequencing sequence. The data was filtered and the reads with connector were removed. Clean Reads were obtained by filtering reads with high N content (more than 10%) and low mass value (base number below 10 exceeding 50%).

3.3 Comparison and statistics with reference genome

By using BWA software (Li and Durbin, 2009), the short sequence obtained by sequencing was compared with the reference genome of moso bamboo (version: Bamboo; download website: http://www.bamboogdb.org/page/download.jsp). The location ratio, sequencing depth and coverage of the 4 samples were counted.

3.4 Detection of genomic variation

According to the location results of Clean Reads, SNP and InDel variation were detected by de duplication with Picard software (Gordon et al., 2012) and pretreatment such as local weight comparison and correction of base mass value with GATK software (McKenna, 2010). SV was detected by using BreakDancer software (Chen et al., 2009).

3.5 Annotation of SNP, InDel and SV

SNP, InDel and SV were annotated by SnpEff software (Cingolani et al., 2012). According to the detection, the position of the variation site compared to the reference genome, the occurrence area of the variation site in the genome and the synonymous and non-synonymous mutations were obtained.

3.6 Excavating variant genes and functional annotation

The InDel genes, SV genes and non-synonymous mutant SNP genes were searched in the CDS region of the samples and the reference genome, and the possible functional variation genes were excavated. The variant genes were compared with functional databases such as GO (Ashburner et al., 2000), COG (Tatusov et al., 2000) and KEGG (Minoru et al., 2004) by BLAST software to obtain gene annotation and analyze gene function.

Authors' Contributions

MSH was the main executor of experimental design and research, completing some data analysis and writing the first draft of the manuscript; YXH and LJ were responsible for sample collection and laboratory work; LXP was responsible for sequencing and data analysis; GJ participated in the design and implementation of the experiment. All authors read and approved the final manuscript.

Acknowledgments

This study was jointly funded by the "948" Project of the National Forestry and Grassland Administration (2015-4-13) and the Fundamental Scientific Research Funds of the International Center for Bamboo and Rattan (1632016003).

References

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T., Harris M.A., Hill D.P., Issel-Tarver L., Kasarskis A., Lewis S., Matese J.C., Richardson J.E., Ringwald M., Rubin G. M. and Sherlock G., 2000, Gene Ontology: tool for the unification of biology, Nat. Genet., (25): 25-29
https://doi.org/10.1038/75556
PMid:10802651 PMCid:PMC3037419

Bai H., Cao Y.H., Quan J.Z., Li Z.Y., Zhu Y.B., Zhu L.H., Dong Z.P., and Li D.Y., 2013, Identifying the genome-wide sequence variations and developing new molecular markers for genetics research by resequencing a landrace cultivar of foxtail millet, PLoS One, 8(9): 274-275
https://doi.org/10.1371/journal.pone.0073514
PMid:24039970 PMCid:PMC3769310

Bai Q., Hou D., Li L., Cheng Z.C., Wei Ge W., Liu J., Li X.P., Mu S.H., and Gao J., 2016, Genome-wide analysis and expression characteristics of small auxin-up RNA (*SAUR*) genes in moso bamboo (*Phyllostachys edulis*), Genome, 60(4): 325-336
https://doi.org/10.1139/gen-2016-0097

Chen J.P., Yan C.Y., Wang Y.M., Zhang X.M., Zeng X., Cao X.Y., and He Q., 2011, Physiological characteristics of 6 different variation type bamboo's leaves, Zhongnan Linye Keji Daxue Xuebao (Journal of Central South University of Forestry&Technology), 31(4): 70-73

Chen K., Wallis J.W., Mclellan M.D., Larson D.E., Kalicki J.M., and Pohl C.S., 2009, Breakdancer: an algorithm for high-resolution mapping of genomic structural variation, Nat. Methods, 6(9): 677-681
https://doi.org/10.1038/nmeth.1363

Chen S.L., Yang Q.P., Guo Z.W., 2008, Study on variation of culm morphology in *Bambusa ventricosa* with different cultivation measures, Linye Kexue Yanjiu (Forest Research), 21(1): 118-121

Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., and Wang L., 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, 6(2): 80-92
https://doi.org/10.4161/fly.19695
PMid:22728672 PMCid:PMC3679285

Gordon P.M.K., Dimnik.L., and Parboosingh J.S., 2012, Optimizing genotype quality metrics for individual exomes and cohort analysis, BMC Proceedings, 6: 21-22
https://doi.org/10.1186/1753-6561-6-S6-P42
PMCid:PMC3467544

Jia X.P., Zhang B., Quan J.Z., Li J.F., Wang Y.F., and Yuan X.L., 2019, Genome-wide association analysis of plant height in foxtail millet under different photoperiod conditions, Huabei Nongxuebao (Acta Agriculturae Boreali-Sinica), 34(4): 16-23

Li H., and Durbin R., 2009, Fast and accurate short read alignment with Burrows-Wheeler transform, Bioinformatics, 25(14): 1754-1760
https://doi.org/10.1093/bioinformatics/btp324
PMid:19451168 PMCid:PMC2705234

Liu J., Huang R., Cheng Z.C., and Li X.P., 2018, Genome-wide identification and the whole analysis of TCP gene family in moso bamboo (*Phyllostachys edulis*), Jiyinzuxue yu Yingyong Shengwuxue (Genomics and Applied Biology), 37(12): 5388-5397

Ma N.X., Lai G.H., and Zhang P.X., 2014, The genus phyllostachys in China, Zhejiang Science and Technology Press, Hangzhou, pp.92-107

Mckenna A, Hanna M., and Banks E., 2010, The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data, Genome Res., 20(9): 1297-1303
https://doi.org/10.1101/gr.107524.110
PMid:20644199 PMCid:PMC2928508

Minoru K., Susumu G., and Shuichi K., 2004, The KEGG resource for deciphering the genome, Nucl. Acid Res., 32(22): 277-280
https://doi.org/10.1093/nar/gkh063
PMid:14681412 PMCid:PMC308797

Peng Z.H., Lu Y., and Li L.B., 2013, The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*), Nat. Genet., 45(4): 456-461
https://doi.org/10.1038/ng.2569
PMid:23435089

Qi X.P., Li M.W., Xie M., Liu X., Ni M., Shao G.H., Song C., Yim A.K., Wong F.L., Isobe S., Wong C.F., Wong K.S., Xu C.Y., Li C.Q., Wang Y., Guan R., Sun F.M., Fan G.Y., Xiao Z.X., Zhou F., Phang T.H., Liu X., Tong S.W., Chan T.F., Yiu S.M., Tabata S., Wang J., Xu X., and Lam H.M., 2014, Identification of a novel salt tolerance gene in wild soybean by whole- genome sequencing, Nat. Communications, 5: 4340-4342
https://doi.org/10.1038/ncomms5340
PMid:25004933 PMCid:PMC4104456

Takagi H., Abe A., and Yoshida K., 2013, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, Plant J., 74(1): 174-183
https://doi.org/10.1111/tpj.12105
PMid:23289725

Tatusov R.L., Galperin M.Y., and Natale D.A., 2000, The COG database: a tool for genome-scale analysis of protein functions and evolution, Nucl. Acid Res., 28(1): 33-36
https://doi.org/10.1093/nar/28.1.33
PMid:10592175 PMCid:PMC102395

Wei Q., and Ding Y L., 2017, Transcriptome sequencing, de novo assembly and expression analysis of several genes related to internode development in *Pseudosasa japonica*, Nanjing Linye Daxue Xuebao (Journal of Nanjing Forestry University (Natural Science Edition)), 41(5): 42-48

Xie L.H., Li X.Y., and Hou D., 2019, Genome-wide analysis and expression profiling of the heat shock factor gene family in *Phyllostachys edulis* during development and in response to abiotic stresses, Forests, 10(2): 100-105
https://doi.org/10.3390/f10020100

Yan C.Y., 2011, Different variation type's bamboo with bamboo's biology physiology characteristic, Thesis for M.S., Central South University of Forestry&Technology, Supervisor: Chen J.H., pp.39-43

Yin M.H., Wang Q., Zhang H.L., Cai X.H., Xu C.Q., Chen F.L., Liu S.Y., Zhang Q.W., Cai H., and Chen R.H., 2020, Whole genome re-sequencing analysis of alpine potato and local farm potato in Huaiyu Mountain under high altitude habitats, Jiyinzuxue yu Yingyong Shengwuxue, 39(3): 1198-1207

Ruan X.S., 2008, Genetic variations of *Phyllostachys edulis* provenance and cultivars by ISSR and AFLP, Thesis for M.S, Zhejiang Forestry College, Supervisor: Fang W., and Lin X.C., pp.30-31

Zidani S., Ferchichi A., and Chaieb M., 2005, Genomic DNA extraction method from pearl millet (*Pennisetum glaucum*) leaves, Afr. J. Biotechnol., 4: 862-866