Large Scale Profiling of Protein Isoforms Using Label-Free Quantitative Proteomics Revealed the Regulation of Nonsense-Mediated Decay in Moso Bamboo (*Phyllostachys edulis*)

Xiaolan Yu [1], **Yongsheng Wang** [1], **Markus V. Kohnen** [1], **Mingxin Piao** [1,2], **Min Tu** [1], **Yubang Gao** [1], **Chentao Lin** [1,3], **Zecheng Zuo** [1,2,*] **and Lianfeng Gu** [4,*]

[1] Basic Forestry and Proteomics Research Center, College of Life Science, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[2] Jilin Province Engineering Laboratory of Plant Genetic Improvement, College of Plant Science, Jilin University, 5333 Xi'an Road, Changchun 130062, China

[3] Department of Molecular, Cell & Developmental Biology, University of California, Los Angeles, CA 90095, USA

[4] Basic Forestry and Proteomics Research Center, College of forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China

[*] Correspondence: zuozecheng@fafu.edu.cn (Z.Z.); lfgu@fafu.edu.cn (L.G.); Tel.: +86-591-83590305 (L.G.)

Abstract: Moso bamboo is an important forest species with a variety of ecological, economic, and cultural values. However, the gene annotation information of moso bamboo is only based on the transcriptome sequencing, lacking the evidence of proteome. The lignification and fiber in moso bamboo leads to a difficulty in the extraction of protein using conventional methods, which seriously hinders research on the proteomics of moso bamboo. The purpose of this study is to establish efficient methods for extracting the total proteins from moso bamboo for following mass spectrometry-based quantitative proteome identification. Here, we have successfully established a set of efficient methods for extracting total proteins of moso bamboo followed by mass spectrometry-based label-free quantitative proteome identification, which further improved the protein annotation of moso bamboo genes. In this study, 10,376 predicted coding genes were confirmed by quantitative proteomics, accounting for 35.8% of all annotated protein-coding genes. Proteome analysis also revealed the protein-coding potential of 1015 predicted long noncoding RNA (lncRNA), accounting for 51.03% of annotated lncRNAs. Thus, mass spectrometry-based proteomics provides a reliable method for gene annotation. Especially, quantitative proteomics revealed the translation patterns of proteins in moso bamboo. In addition, the 3284 transcript isoforms from 2663 genes identified by Pacific BioSciences (PacBio) single-molecule real-time long-read isoform sequencing (Iso-Seq) was confirmed on the protein level by mass spectrometry. Furthermore, domain analysis of mass spectrometry-identified proteins encoded in the same genomic locus revealed variations in domain composition pointing towards a functional diversification of protein isoform. Finally, we found that part transcripts targeted by nonsense-mediated mRNA decay (NMD) could also be translated into proteins. In summary, proteomic analysis in this study improves the proteomics-assisted genome annotation of moso bamboo and is valuable to the large-scale research of functional genomics in moso bamboo. In summary, this study provided a theoretical basis and technical support for directional gene function analysis at the proteomics level in moso bamboo.

Keywords: *Phyllostachys edulis*; label-free quantitative proteomics; long noncoding RNA; alternative splicing; protein isoform; nonsense-mediated mRNA decay

1. Introduction

Bamboo belongs to gramineous plants with about 90 genera and 1200 species [1]. It is a major non-wood forest species and widely distributed in tropical, subtropical and temperate zones [2]. Moso bamboo has been cultivated for a long time and is one of the main bamboo species in China, and is rich in lignocellulose [3]. Its planting area accounts for 70% of Chinese bamboo and 80% of total bamboo timber [1]. Moso bamboo characteristically experiences rapid growth [1–4]. In addition, moso bamboo has many important economic, cultural and ecological values, such as edible bamboo shoot, timber, ornamental, and greening values [5–7]. The bamboo forest has an important ecological value for the terrestrial carbon cycle and for ecosystem protection [5–7]. Because of its important economic value and ecological perspectives, it has attracted the attention of community forestry, and has been regarded as the main model plant in the study of plant biology of bambusoideae.

The completed genome sequences of moso bamboo in 2013 have shown that 2.05 Gb assembled sequences included annotated 31,987 genes [4]. The annotated genes of the moso bamboo were based on prediction algorithms, and were further improved by the long-read sequencing technology of Pacific Biosciences (PacBio) [8,9]. In recent years, large-scale transcriptome data have also been available for moso bamboo [9–16], which greatly improved the annotation of protein-coding genes in moso bamboo. In additional to protein-coding genes, long noncoding RNA (lncRNA), plays an important regulatory role in the biochemical processes of organisms, although the function of this has not yet been identified in organisms [17,18]. In total, 1989 long non-coding RNAs (lncRNA) were predicted in moso bamboo [9]. Although the transcriptome greatly promotes the annotation of the moso bamboo genes, these hypothetical genes are based on gene prediction algorithms. Gene prediction algorithms are based on universal translation rules. The genome annotation process usually includes many errors and is generally inaccurate [19–23]. Recently, proteomic data from mass spectrometry experiments have been commonly used to correct and validate protein coding genes and identify new coding genes [21–26]. Proteomics also can effectively quantify the most valuable protein participants in the organism [21–23,27]. However, the high content of lignification and fiber in moso bamboo has caused difficulties in protein extraction, which limits the research progress of quantitative proteomics in moso bamboo [4,28,29]. Thus, the annotated gene according to the genome and transcriptome has not been validated from the protein level, and the function of these genes in the growth and development of moso bamboo remains unknown.

In eukaryotes, a single gene can produce multiple transcriptional isoforms through variable post-transcription regulation [30–32]. Next-generation sequencing (NGS) methods revealed multiple splicing isoforms, which were generated from alternative splicing (AS) [30,31]. In recent years, the PacBio sequencing platform, a single-molecule sequencing technology, can produce longer reads, greatly helping to recognize multiple splicing isoforms produced by AS in plant [33]. Full-length splice isoform sequencings have been applied to the discovery of multiple transcriptional isoforms produced by AS in a variety of species [15,34–36]. Similarly, previous studies have shown that AS and alternative polyadenylation (APA) were ubiquitous in moso bamboo [8,9,37]. The functional characteristics of protein isoforms translated by different splicing isoforms in eukaryotic organisms have been shown to be significantly diverse [38]. However, it remained unknown that if these splicing isoforms in moso bamboo could be translated into different protein isoforms because alternative transcripts might be degraded by nonsense-mediated mRNA decay (NMD) in mammalian [39] and plant [40–42]. In vitro expression experiments provide evidence for whether different splicing isoforms produced by the same gene encode proteins, and protein-protein experiments demonstrate a significant divergence in function between different isoforms [43]. Correspondingly, proteomics is also commonly used to identify whether different splicing isoforms encode proteins [44–52]. Compared to in vitro expression assays, evidence from the proteomics can more directly and strongly demonstrate whether different isoforms naturally encode proteins in vivo. However, it is unclear whether all isoforms in moso bamboo are identified by PacBio sequencing encode proteins, since there is a lack of evidence of protein levels. Proteome experiments are urgently needed to prove whether isoforms encode proteins.

Moso bamboo is an important forest species with a variety of ecological, economic and cultural values. However, the gene annotation information of moso bamboo is based on the transcriptome sequencing, lacking the evidence of proteome. In this study, we successfully optimized the extraction efficiency of moso bamboo protein and identified the total protein of moso bamboo by using label-free quantitative proteomics. These studies improved the annotation of the protein isoforms of moso bamboo based on proteomic analysis and revealed the differences in physiological process of moso bamboo by analyzing the protein expression profile of moso bamboo seedlings and leaves using label-free quantitative proteomics. Also, we proposed that 18% unique transcript isoforms included premature termination codons could also be translated into proteins. In summary, proteomic analysis in this study improves the proteomics-assisted genome annotation of moso bamboo and is valuable for the large-scale research of functional genomics in moso bamboo.

2. Materials and Methods

2.1. Plant Materials and Growth Conditions

Moso bamboo (*Phyllostachys edulis*) seeds were treated with a shelling machine and soaked in water for 24 h. Seeds were sown on soil and grown in the greenhouse at a temperature of 25 °C and 16 h light/8 h dark photoperiod, with a light intensity of 80 µmol m^{-2} s^{-1} and 50% relative humidity. All the materials were in the same condition and were placed in the greenhouse for cultivation. The leaves, roots and stems of the moso bamboo were collected from 30- and 60-days period, respectively. All collected samples were rapidly frozen in liquid nitrogen and then stored at −80 °C. We collected a group of leaves with same sizes from multiple planting pots and pooled them together. The leaves collected during the 30-days and 60-days period were fully developed leaves (Figure 1A). In the process of material selection, we randomly selected each seedling. Multiple leaves with the same height and size were pooled together, which could ensure the reliability of the experimental results and enough material for protein extraction.
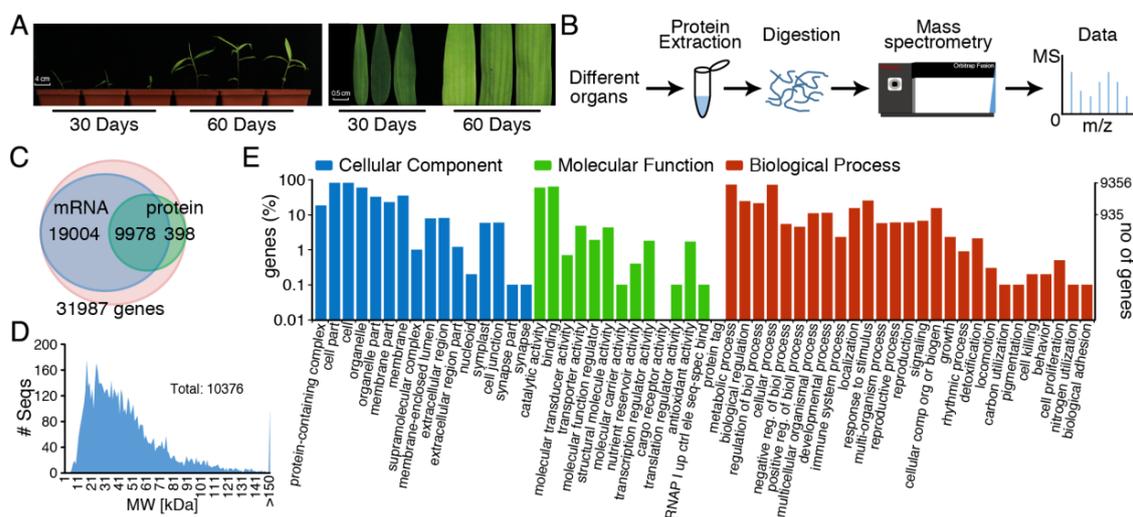
Figure 1. The flowchart of high-throughput label-free quantitative proteomics in moso bamboo. (**A**) The fully developed leaves of 30-days-old seedlings and 60-days-old moso bamboo were collected to detect the concentration and efficiency of protein extraction among different periods; (**B**) The pipeline of label-free quantitative proteomics; (**C**) Venn diagram of moso bamboo proteins and the mRNA transcription; (**D**) The distribution of molecular weight of moso bamboo proteins; (**E**) GO functional annotation of total proteins.

2.2. Extraction of Proteins from Moso Bamboo

Moso bamboo materials were frozen in liquid nitrogen and were extracted from ground material with lysis buffer (10 mM Tris-HCl pH 8.0 5 mM EDTA 1% SDS 8 M Urea 20 mM Dithiothreitol). The lysis buffer was supplemented with EDTA-free Protease Inhibitor Cocktail Tablets. After incubation on ice for 30 min, the lysate was centrifuged at 12,000× *g* for 15 min. After adding phenol extraction buffer, the supernatants were precipitated overnight at −20 °C using six volumes of 10% TCA/acetone (10% trichloroacetic acid dissolved in acetone) or 100% acetonitrile (ACN). After centrifugation at 12,000× *g* for 30 min at 4 °C, pellets were washed three times with pre-cooled acetone and dried for 5 min by vacuum freeze-drying. Protein samples were resuspended in buffer (8 M urea 0.1 M Tris-HCl pH 8.0). The concentration of purified proteins was measured by BCA protein assay (Thermo Fisher Scientific, Waltham, MA, USA) at the absorbance 562 nm, using bovine serum albumin as the standard.

2.3. Protein Digestion and Peptide Fraction

A total of 500 μg proteins from seedlings and leaves were digested using a filter-aided sample preparation (FASP) method following DTT reduction, iodoacetamide (IAA) alkylation and overnight digestion with trypsin at a ratio of 1:100. Digested peptides were dried using a CentriVap (Thermo Fisher) and pre-fractionated with Ultimate 3000 (Thermo Fisher scientific, Waltham, MA, USA). The peptide mixture was re-dissolved in buffer A (buffer A: 20 mM ammonium formate in water, pH 10.0, adjusted with ammonium hydroxide), and then fractionated by high pH separation using Ultimate 3000 system connected to a reverse phase column (XBridge C18 column, 4.6 mm × 250 mm, 5 μm (Waters Corporation, Milford, MA, USA)). High pH separation was performed using a linear gradient with 5% buffer B up to 45% buffer B over 40 min (buffer B: 20 mM ammonium formate in 80% ACN, pH 10.0, adjusted with ammonium hydroxide). The column was re-equilibrated to initial conditions for 15 min. The column flow rate was maintained at 0.8 mL/min and column temperature was maintained at 30 °C. In total, twelve fractions were collected. Each fraction was dried and desalted using a C18 Ziptip (Thermo fisher scientific).

2.4. Mass Spectrometry Analysis

Peptide fractions were resuspended with 30 μL solvent C (C: water with 0.1% formic acid) and analyzed by on-line nanospray LC-MS/MS on an Orbitrap Fusion coupled to an EASY-nano-LC system (Thermo Scientific, MA, USA). For one LC-MS/MS run, 5 μL (1 μg) peptide sample was loaded onto the trap column (Thermo Scientific Acclaim PepMap C18, 100 μm × 2 cm), with a flow of 10 μL/min and subsequently separated on the analytical column (Acclaim PepMap C18, 75 μm × 15 cm) with a 60-min linear gradient of 3–32% solvent D (D: ACN with 0.1% formic acid). The column flow rate was maintained at 300 nL/min. The electrospray voltage of 2 kV versus the inlet of the mass spectrometer was used.

The mass spectrometer was run under data dependent acquisition mode, and automatically switched under MS and MS/MS mode in 3 s cycles. MS1 mass resolution was set as 60 K with *m/z* 350–1550 and MS/MS resolution was set as 50 K under HCD mode. The dynamic exclusion was set to n = 1, and the dynamic exclusion time was 45 s. AGC target was 8×10^4 and the max injection time was 120 ms.

2.5. Protein Identification Using Mass Spectrometry

Open reading frames (ORFs) were extracted and translated into amino acid sequences using TransDecoder software (v.5.5.0) with the default option [53]. This software is also used to build the potential amino acid sequences for 1989 long chain non-coding sequence and 93,481 unique transcript isoforms [9]. All MS/MS raw data were analyzed using Proteome Discoverer 2.1 (Thermo Fisher Scientific, San Jose, CA, USA; version 2.1) with the SEQUEST algorithm. A total of 31,987 protein-coding genes have been annotated in the moso bamboo genome [4]. In this study, we used annotated protein

sequence from previous published paper [4], which is a combination of genome and transcriptome sequencing analysis of the encoded information of the moso bamboo. That protein database provided one most reliable and unique protein annotation for each locus. All MS/MS spectra searches were conducted against a concatenated forward/reversed version of *P. heterocycla-v1.0* sequence database [4] and filtered using a false discovery rate (FDR) < 0.01 at both the peptide and protein level. Carbamidomethyl of cysteine were selected as the fixed modification while oxidation (M), acetylation (protein N-term) and phosphorylation (STY) were allowed for variable modifications. The database search was performed with an initial precursor ion tolerance of 10 p.p.m. Fragment ion mass tolerance to 0.02 Da and two missed cleavages are allowed. Target decoy searches strategy provided a substantial improvement in large-scale protein identification by mass spectrometry [54]. For label-free procedures, protein abundance quantification was carried out based on the identified peptides using the scaffold software [55]. The relative expression abundance of proteins were calculated and normalized as spectral abundance factors (NSAF) label-free quantitative algorithm which based on the spectral counts (SpC) to define the relative expression of the proteins as SpC divided by the protein length (L), and divided by the total SpC/L across all identified proteins [56–58]. The protein quantitation and confidence were performed using a relatively conservative threshold (fold change ≥ 4 for up-regulation or ≤ 0.25 for down-regulation), false discovery rate < 0.01 with at least 2 peptides matched were considered for further analysis).

2.6. GO Enrichment Analysis and Pathway Analysis

Blast2GO software optimizes the BLAST algorithm to identify homologous sequences and then assigns uncharacterized novel sequences in non-model species with functional annotations, which is represented via the gene ontology (GO) vocabulary [59]. In this study, GO terms of each gene in moso bamboo were assigned using BLAST2GO with default option, and GO terms was assessed with InterProScan [59]. The enrichment of GO terms was assessed with the BiNGO plugin for Cytoscape, using hypergeometric test for statistical analysis. The cutoff *p value* is set to 0.05 and all terms with a *p* value < 0.05 were considered statistically as over-represented GO and shown as colored nodes in the enrichment network and histogram. Pathway analysis were performed using KEGG online sites (https://www.kegg.jp/blastkoala/) [60].

2.7. Data Availability

The original LC-MS/MS file was available at http://forestry.fafu.edu.cn/pub/cells and could be analyzed by protein retrieval software such as Proteome Discoverer.

3. Results

3.1. Extraction and Identification of Total Protein Using Phenol Extraction Combined with Acetonitrile Precipitation for Orbitrap Fusion Tribrid Mass Spectrometry

In this study, proteins of 30-days-old seedlings and 60-days-old leaves were extracted, and their concentration was analyzed with the BCA assay. The fully developed leaves of 30-days-old seedlings and 60-days-old moso bamboo were collected to detect the concentration and efficiency of protein extraction among different periods (Figure 1A). Phenol extraction combined with acetonitrile precipitation or TCA/acetone precipitation were used to extract bamboo protein. For stem and root sample harvested at 60-days old, phenol extraction combined with acetonitrile precipitation resulted in only low protein concentrations due to serious lignification and increased fiber content at this developmental stage. Nevertheless, this method could effectively extract total proteins from 30-days-old seedlings and 60-days-old leaves (Figure 1A). We did not carry out mass spectrometry analysis for the roots and stems of 60-days-old moso bamboo because of the low protein concentration. Protein samples extracted from seedlings and leaves of moso bamboo were analyzed using Orbitrap Fusion Tribrid mass spectrometry (Figure 1B). In total, 1,020,880 tandem mass spectra were generated (Table 1).

Table 1. The comparative analysis of the efficiency of the protein extraction methods in seedlings and leaves of moso bamboo.

	Phenol Extraction			
	Acetonitrile		Tca/Acetone	
	Seedling	Leaf	Seedling	Leaf
Spectra	255470	295637	243561	226212
Peptide	19470	30881	3155	1694
Protein	7192	9336	1937	1365

Comparing the observed spectra with the protein database derived from prediction of genome and transcriptome data, 55,200 peptides were identified from these highly trusted mass spectrometry data. In total, mass spectrometry peptides matching identified 10,376 annotated genes, which further verified the gene models on a translational level in moso bamboo (Figure 1C). The protein molecular weight map showed that the 10,376 detected proteins ranged from 10 kDa to 150 kDa. The protein content was relatively high in the range of 15 kDa to 90 kDa, which covered a wide protein size distribution (Figure 1D). Phenol extraction combined with acetonitrile precipitation method could extract 7192 and 9336 proteins from seedlings and leaves of moso bamboo, respectively. However, phenol extraction combined with TCA/acetone precipitation only extracted 1937 and 1635 proteins from seedlings and leaves, respectively. The results further showed that phenol extraction combined with acetonitrile precipitation method was efficient and feasible for extracting total protein from moso bamboo (Table 1).

3.2. Analysis of Physiological Process and Metabolic Pathway for All Identified Protein Data

GO functional annotation and classification were performed for all identified protein data of moso bamboo by using Blast2GO software [59]. Metabolic process, cellular process and response to stimulus occupied the first three positions in the biological process, indicating that the proteins involved in these physiological metabolic processes were relatively high in the expression of protein (Figure 1E). GO terms from the molecular function ontology shown that most genes were involved in binding, catalytic activity, transporter activity and structural molecule activity. The results of GO annotation analysis showed a general physiological state of moso bamboo.

3.3. Differential Expression Profiles of Proteomics in Seedlings and Leaves of Moso Bamboo using High-Throughput Label-Free Quantitative Proteomics

A total of 7192 and 9336 proteins were identified in 30-days-old seedlings and 60-days-old leaves, respectively. 1040 proteins were specific to the seedling stage and 3184 proteins were specifically identified in the 60-days-old leaves (Figure 2A). In addition, 6152 proteins were identified in both tissues, accounting for almost 60% of all identified proteins, 85% of the identified proteins in 30-days-old seedlings and 66% of the identified proteins in 60-days-old leaves. The results showed that the physiological and biochemical processes were similar in the leaves and seedlings of moso bamboo.

According to all the identified proteomic data of moso bamboo, the functional enzymes were classified and the data on the seedlings and leaves of moso bamboo were compared. The common enzyme classification mainly included oxidoreductase, transferase, hydrolase, lyase, isomerase and ligase. The three most frequent enzyme classes were transferase, hydrolase and oxidoreductase, which indicated that the biological processes of moso bamboo mainly included oxidation-reduction processes, amino acid metabolism processes and tricarboxylic acid cycle reactions. Interestingly, the number of enzyme classification for leaves was higher than that for seedlings. It may be that the physiological metabolism process in the leaves of moso bamboo is more active than the seedlings (Figure 2B).
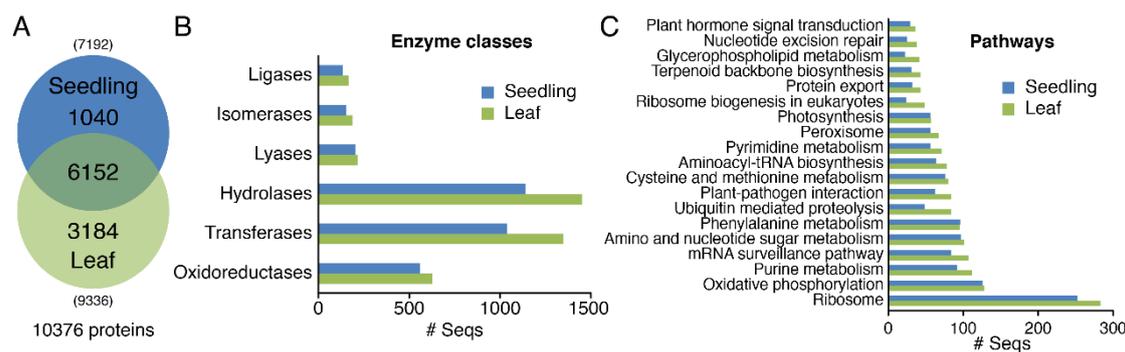
Figure 2. Profiles of proteins in seedlings and leaves of moso bamboo. (**A**) Comparison of protein numbers in seedlings and leaves of moso bamboo; (**B**) The classification of enzymes in seedlings and leaves of moso bamboo; (**C**) Pathway analysis, according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

The analysis of the metabolic pathways indicated that the proteins identified in the seedlings and leaves mainly included ribosome, oxidative phosphorylation and purine metabolism. Although the identified proteins in the leaves and seedlings of moso bamboo were part of the same metabolic pathways, the number of proteins in the same metabolic pathway was different. In all metabolic pathways, the number of proteins enriched in the leaves of moso bamboo was higher than that of the seedlings (Figure 2C). Other important pathways included photosynthesis and phenylalanine metabolic pathway. Oxidative phosphorylation was a kind of energy conversion of proteins, which occurred in mitochondria. The above results showed that part physiological metabolism was similar in the seedlings and leaves of moso bamboo, and there was also different physiological state. Comparing with seedling, physiological metabolism of the leaves was slow. However, the leaves activated and expressed more proteins related to plant stress response, disease defense and energy metabolism in order to ensure the normal physiological function and stress resistance of the organism.

3.4. GO Analysis of Proteomics in Seedlings and Leaves of Moso Bamboo

The correlation analysis of normalized spectral abundance factors (NASF) between 30-days-old seedlings and 60-days-old leaves had a correlation coefficient of 0.57 (Figure 3A), suggesting a moderate correlation between both tissues. These results indicated that moso bamboo had common as well as distinct physiological and biochemical processes at different growth stages. Biological processes of protein enrichment in seedlings and leaves of moso bamboo were similar with each other. There was no significant difference in the number of proteins involved in several biological processes, including glycolysis, translation and membrane transport. Moso bamboo presented as being fast-growing, and the fully developed leaves from 30-days and 60-days was significantly different in the size and color (Figure 1A). Thus, the correlation coefficient between 30-days-old seedlings and 60-days-old leaves was 0.57, suggesting a variation between two periods. Compared with the leaves from 30-days-old seedlings, there were 1011 differentially expressed proteins in the 60-days-old leaves, accounting for 22.1% of the total number of proteins quantified. Among 1011 differentially expressed proteins, 37 were down-regulation and 974 were up-regulation in 30-days-old seedlings compared to that in 60-days-old leaves.

In addition, we analyzed the differences in protein expression profiles and performed GO enrichment analysis on the differentially expressed proteins by using label-free quantitative proteomics with a *p* value cutoff < 0.05 in moso bamboo from 30-days-old and 60-days-old fully developed leaves, respectively (Figure 3B,C). Enrichment biological processes in down-regulated proteins in seedlings compared to leaves were mainly involved in GO terms related to hydrogen peroxide catabolism. Regulation of hydrogen peroxide was often associated with plant stress responses and disease defenses. Specific binding of kinase and substrate play a role in regulating plant growth and development,

and are also key regulators of signal transduction [61]. The number of proteins involved in redox process and oxidative phosphorylation in leaves was more than that in seedlings, which indicated that more energy might be needed to maintain the balance of the organism in the later growth stage of moso bamboo (Figure 3B). Enrichment biological processes in up-regulated proteins included amongst others translation and photosynthesis (Figure 3C).
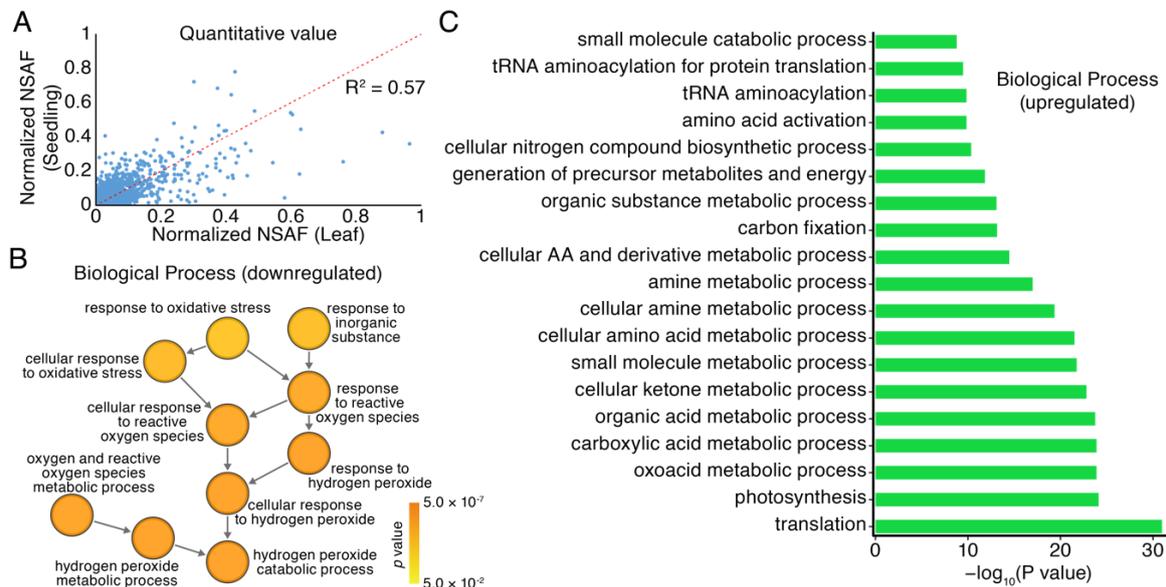
Figure 3. The gene ontology (GO) enrichment analysis of differential expression proteins in seedlings and leaves of moso bamboo based on mass spectrometry-based label-free quantitative proteomics. (**A**) The correlation analysis of protein abundance between seedlings and leaves of moso bamboo; (**B**) The biological process for down-regulated proteins; (**C**) The biological process for up-regulated proteins.

3.5. Discovery of New Protein Coding Genes

A comparison of all protein database produced by the mass spectrometer with the current annotated protein database created from genomic and transcriptome data revealed that many tandem mass spectra were not matched, which reminded us that the genome annotation information of moso bamboo is not completely accurate. To modify and supplement the coding information of moso bamboo genome, the remaining 21,611 unproven-gene annotation genes were translated using all potential open reading frame and were matched with tandem mass spectra to identify new genes and annotated coding genes. According to internationally recognized standards for protein mass spectrometry identification, each recognized protein had at least two or more peptides matching support. In total, we identified 1108 new translation frames, and this translation information complemented the previously predicted gene coding patterns (Figure 4).
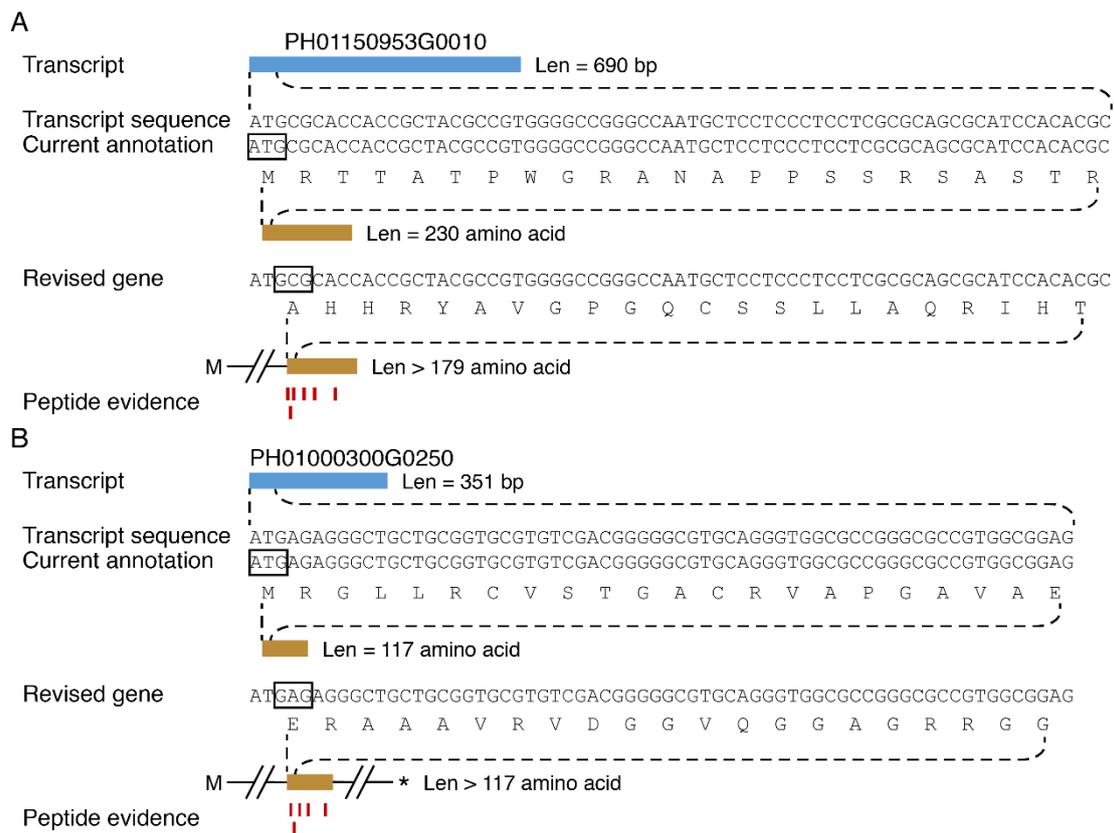
Figure 4. Revision of gene annotation in moso bamboo and discovery of new genes. Current annotation and translation frame (black frames) and revised gene models of PH01150953G0010 (**A**) and PH0100300G0250 (**B**) supported by matching peptide spectra. Stop codon was marked by an asterisk.

In addition, 1989 locus were predicted to be non-coding RNA in the transcriptome analysis of moso bamboo using PacBio long reads sequencing [9]. Since these lncRNAs were identified based on software prediction, parts of the RNA sequences might actually encode proteins. In order to test this hypothesis, all lncRNAs were translated and matched against tandem mass spectra (Figure 5A). Among the 1989 predicted lncRNA, 1015 original proteins were successfully identified by mass spectrometry. This shows that translational activity could be validated for about half (51.03%) of the in-silico predicted lncRNA (Figure 5B). We also performed the differentially expressed lncRNA encoded proteins and identified 291 down-regulated and 145 up-regulated protein, respectively (Figure 5F). The newly identified 1015 proteins-encoding genes greatly complement the moso bamboo gene annotation database, which laid the foundation for future studies of moso bamboo. The re-annotation according to mass spectrometry data had obvious advantages over that provided by prediction, which could accurately predict the correct translation framework.

Homology and functional analysis of these new protein-coding genes were performed by BLASTP against NCBI's non-redundant protein database (NR) using Blast2GO. InterProScan was used to annotate the functions of these new genes. Of the new genes coding translation frames, 971 genes were functionally annotated and classified and for 1211 genes, no annotation could be retrieved. Among the functional classification of annotated genes, 254 genes were involved in cellular processes, 164 genes were involved in nitrogen compound metabolic processes, and there were also many genes participated in carbon fixation, photosynthesis, response to stimulus and gene expression, etc. (Figure 5C–E).
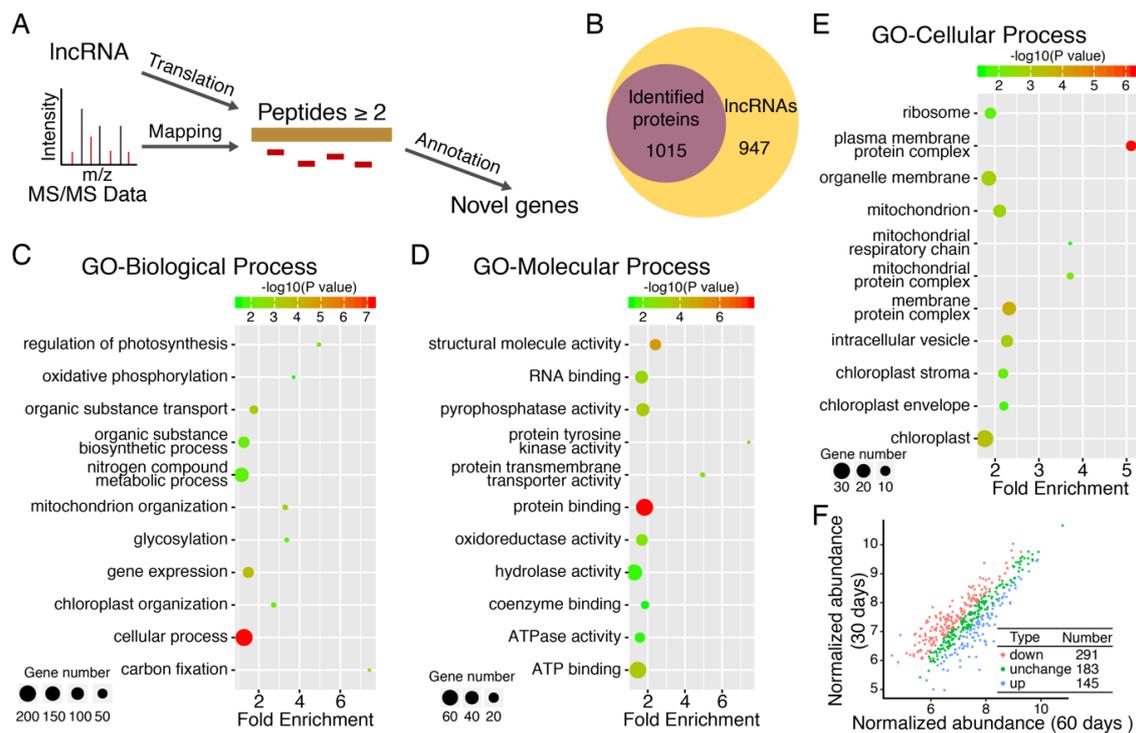
Figure 5. The re-annotation of lncRNA through mass spectrometry data. (**A**) Overview of pipeline for lncRNA re-annotation. All lncRNAs were translated into amino acid sequences. Selected Amino acid (aa) sequences (> 100 aa) were compared to mass spectrometry data and sequences with. more than 2 peptide-matches were considered to be authentic; (**B**) Venn diagram illustrating the number of lncRNA re-annotated by mass spectrometry; (**C**) The biological process for lncRNA re-annotated by mass spectrometry; (**D**) The molecular function for lncRNA re-annotated by mass spectrometry; (**E**) The cellular component for lncRNA re-annotated by mass spectrometry. Explanatory information on the functional enrichments and numbers of involved proteins in terms were all listed on the left. The cutoff of the *p* value is set at 0.05. Different colors (such as red, green and yellow) stand for –log10 (*p* value) in the GO analysis; (**F**) The plot of differentially expressed lncRNA encoded proteins.

3.6. Large Scale Identification of Protein Isoforms

We analyzed whether the 71,923 and 82,390 unique alternative splicing and alternative polyadenylation isoforms, respectively, from 12,091 genes obtained by PacBio Iso-sequencing encoded proteins, could be mapped with mass spectrometry data [9]. In total, 93,481 unique isoforms could be translated into 163,669 potential amino acid sequences by TransDecoder software (Figure 6A). Among all different transcript isoforms matched to mass spectrometry data, we verified that 3284 unique transcript isoforms from 2663 genes could encode proteins, accounting for 3.51% of all isoforms (Figure 6B). The low percentage might be caused by expression levels below the detection threshold for some of the isoforms. Mass spectrometry matching 3284 unique isoforms provided evidence of how each isoform was encoded. Domain analysis of 3284 protein isoforms exhibited a variety of domains (Figure 6C): (1) Long protein isoform contains all domains of the short protein isoform, a total of 298 genes, such as PH01000000G2440; (2) Different protein isoforms from the same gene contained both the same and specific domains, with a total of 187 genes, such as PH01000001G2110; (3) Different protein isoforms from the same gene contain completely different domains, a total of 134 genes, such as PH01000005G0510; (4) Different protein isoforms from the same gene contained the same domain, a total of 230 genes, such as PH01000529G0510. These results suggest that different protein isoforms from one locus have similar or distinct functions due to alternative splicing.
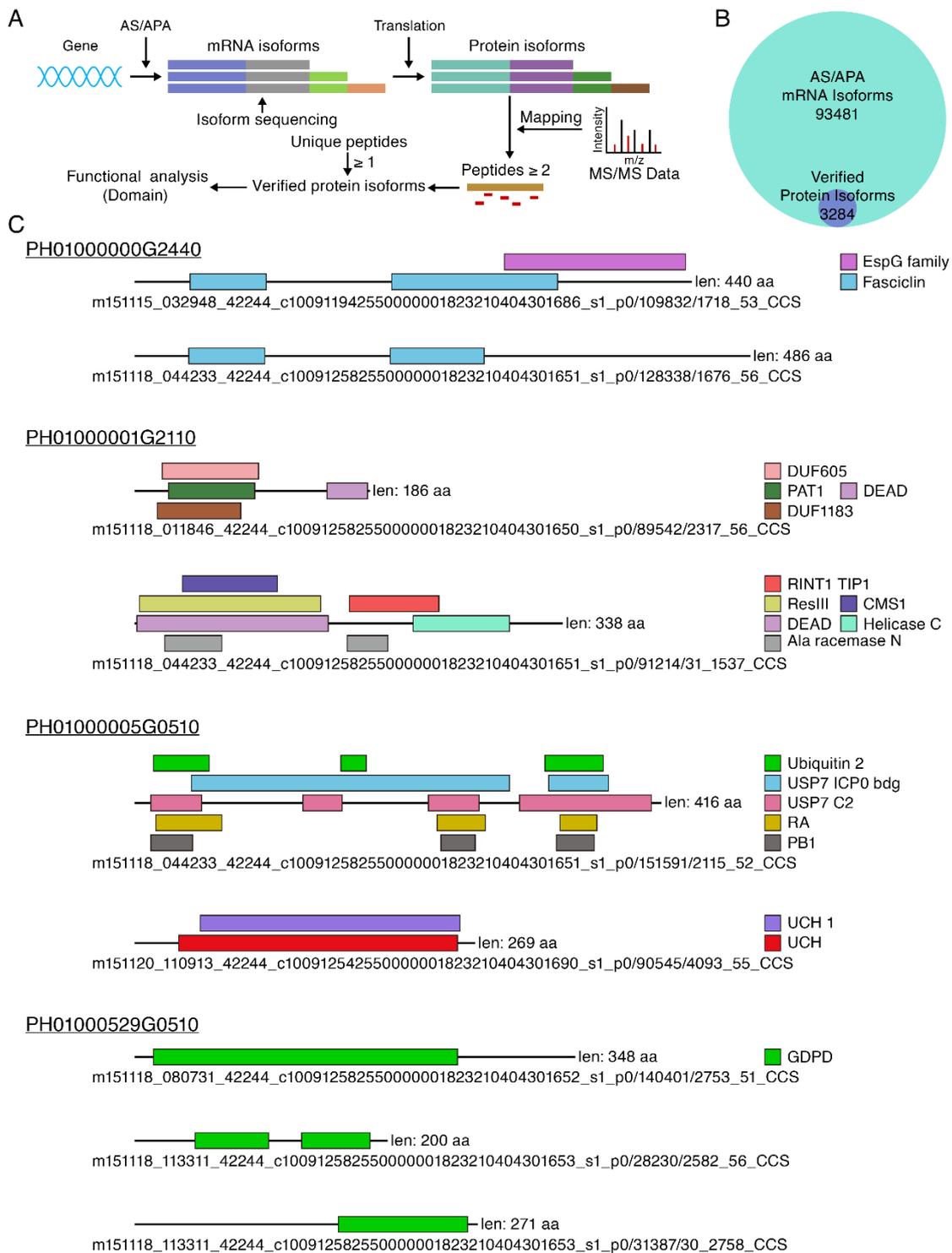
Figure 6. Different protein isoforms from the same locus were identified by mass spectrometry data. (**A**) The pipeline for the verification of different protein isoforms from the same locus for moso bamboo through protein profiling data; (**B**) The Venn diagram illustrating the number of different protein isoforms verified by mass spectrometry in all AS/APA isoform; (**C**) Domain analysis of different protein isoforms from the same locus.

The majority of alternative splicing-derived transcripts contain premature termination codons (PTCs), which are degraded by NMD pathways [62,63]. NMD is typically triggered only by stop codons at least ~50 nt upstream of the last exon–exon junction [64]. Among 93,481 unique transcripts

isoforms, 20,825 were found to be NMD-regulated, accounting for 22.2% of all transcript isoforms. Among the 3284 protein-encoding isoforms identified by mass spectrometry, 601 splicing isoforms received NMD regulation, which accounted for 18.3% of all protein-encoding isoforms and suggested that part transcripts targeted by NMD might escape from the mRNA surveillance mechanism.

4. Discussion

Moso bamboo has important ecological, economic and cultural values, and also has the characteristics of strong carbon sequestration, rapid growth and tall plant. The combination of moso bamboo genome data and transcriptional data contributes to annotating moso bamboo genome [4,9]. However, few studies have supported annotated data of these moso bamboo genes through proteomes to form a more reliable and better annotated reference database. Proteome analysis based on high-resolution mass spectrometry can accurately help to establish and modify genome coding gene annotation information of species, which has been reported in many model plants. In *Arabidopsis thaliana*, proteome analysis was successfully applied to correct and improve annotation information of the genome [65,66]. In monocotyledon maize, 165 new protein-coding genes were annotated and 741 gene-coding models were modified by proteomics [67]. Similarly, in rice, proteome data and transcriptome data have greatly improved the annotation of rice genome [68,69]. However, few studies have supported annotated data of these moso bamboo genes through proteomes to form a more reliable and better annotated reference database. In this study, the extraction method of moso bamboo protein was optimized, hence the differences of different extraction methods for the extraction of moso bamboo protein were compared. It was found that phenol extraction combined with acetonitrile precipitation method had the best extraction effect. The extraction of moso bamboo by phenol extraction combined with acetonitrile precipitation method, a total of 10,376 proteins in moso bamboo were successfully identified, and the number of proteins identified at one time accounted for more than 30% of the number of all annotated genes (Figure 1C). The number of proteins identified by phenol extraction combined with acetonitrile precipitation is 5 times higher than of that with TCA/acetone precipitation (Table 1). This result showed that the extraction of moso bamboo protein by phenol extraction combined with acetonitrile precipitation was highly feasible. There are many methods for protein extraction, among which protein precipitation with miscible organic solvents (usually acetonitrile, methanol, acetone and ACN precipitation method) is the most commonly used. Of these extraction methods, acetonitrile precipitation method is the most effective to precipitate proteins, in which the vast majority of proteins were precipitated [70]. In this study, we found that the acetonitrile method was more efficient than the TCA/acetone method. Since the polar of acetonitrile is higher than that of acetone, it is expected to compete more successfully for water bound to protein, leaving more charged functional group available on protein aggregates to attract free amino acids from solution [71]. Thus, acetonitrile has a higher compatibility with water than acetone. In this study, acetonitrile presented more dehydrated and made it more susceptible to precipitation. Another reason may be that acetonitrile can better dissolve fiber compared with acetone.

In this study, phenol extraction combined with acetonitrile precipitation method was more efficient than TCA/acetone method. The establishment of efficient method for extracting the total proteins from leaves, as well as the identification of proteins based on mass spectrometry provided an accurate and unbiased method for the identification of proteins and improving annotation information in moso bamboo. Importantly, effective annotation of moso bamboo genome will cause some more extensive impacts on subsequent genetic improvement, gene function analysis, germplasm identification and the protection of moso bamboo. In this study, we also extracted the root and stem proteins of moso bamboo after 60-days. However, the concentration and efficiency of protein extraction were low. Label-free quantitative proteomics could not be carried out. Thus, the method of extracting protein from root and stem still needs further optimization.

In this study, we used proteomic data to validate and modify the gene annotation information of moso bamboo. The study of genome annotation and proteome of moso bamboo is helpful to the

genetic breeding of moso bamboo and other species, with important scientific significance. In this study, we identified 55,200 peptides in the genome, which matched 10,376 predictive protein coding genes and validated the translation of predictive gene model. By matching peptides with all potential open reading frames of the corresponding gene sequences, we further improved 1108 gene models. Proteomic data could help us identify whether predicted lncRNA sequences were real non-coding sequences. Among other species, researchers corrected some predicted non-coding RNAs that actually encoded proteins by proteomic data [21,25]. In this study, the mass spectrometry peptides revealed that 1015 predicted lncRNAs might encode proteins.

Proteomics is also widely used to identify different protein isoforms from the same gene in various species [44,45,48–52]. In this study, 3284 transcript isoforms were verified to encode proteins through proteomic mass spectrometry data, accounting for 3.51% of all protein isoforms. In previous studies, only a small amount of alternative splicing proteins was verified by the proteomic method [48–52]. This may be mainly due to the sensitivity of current protein profiles and protein extraction methods. However, not all alternative splicing isoforms play important roles in organisms and are capable of encoding amino acids [50]. The majority of alternative splicing-derived transcripts contain premature termination codons (PTCs), which are degraded by nonsense-mediated mRNA decay (NMD) pathways [47,62,63]. NMD is a post-transcriptional quality control mechanism that selectively degrades mRNA containing PTCs [64,72,73]. In this study, mass spectrometry supported the translation of 18.3% transcript isoforms targeted by NMD. It will be interesting to investigate how these transcript isoforms escaped from mRNA surveillance mechanism. Alternative splicing can generate splicing isoforms targeted by NMD, which is a transcript quality control mechanism to regulate RNA turnover. This study suggested that isoforms with NMD regulation could also give rise to protein produces, which has been validated by label-free proteomics. For future studies that quantitative proteomics performed will be important for NMD factor mutants or for splicing factor mutants to investigate the interplay between NMD machinery and alternative splicing.

In vitro protein functional analysis experiments have demonstrated that different protein isoforms have diverse functions in organism [38,43]. Different protein isoforms play different roles in the organism, and the functional divergences are obvious [38,43]. For example, in humans the BCL-2 gene encodes for two protein isoforms with opposite functions: Bcl-x inhibits cell apoptosis, and Bak promotes cell apoptosis [74]. In this study, by analyzing domains from different isoforms of gene, the results suggest that different isoforms from the same gene have the same and specific functions in moso bamboo. In addition, in some different isoforms, the domains they contain are almost completely diverse, and the results strongly suggest that they play a completely diverse role in the growth and development of moso bamboo.

5. Conclusions

In this study, proteomics was used to improve and correct the protein coding information, which was previously annotated by genomic and transcriptomic data of moso bamboo. Here, we provided experimental evidence for predicted gene models and identified new protein-coding models, which further improved the gene annotation. Proteomic data can help us identify functions in many new transcripts of genome and transcriptome research. In addition, analysis of proteome can identify and correct some gene models that have been omitted by gene prediction algorithms. Interestingly, 18.3% transcript isoforms targeted by NMD could escape from the mRNA surveillance mechanism. The combination of proteome, transcriptome and genome analysis to improve the assembly and annotation of genomes of any species will become a model for future genome sequencing research. The proteome analysis in this study provided evidence for improving the annotation information of predictive genes in moso bamboo. Importantly, the proteomic evidence based on mass spectrometry provided an accurate and unbiased method for the identification of proteins in moso bamboo. The effective annotation of moso bamboo genome will cause some more extensive impacts on

subsequent genetic improvement, gene function analysis, germplasm identification, and the protection of moso bamboo.

Author Contributions: C.L., Z.Z., and L.G. conceived and designed the study; X.Y. performed the experiments and bioinformatics analysis with the help from Y.W., M.V.K., M.P., M.T. and Y.G.; X.Y., Z.Z., and L.G. wrote and revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lobovikov, M.; Paudel, S.; Piazza, M.; Ren, H.; Wu, J.Q. *World Bamboo Resources: A Thematic Study Prepared in the Framework of the Global Forest Resources Assessment 2005*; FAO: Rome, Italy, 2007.

2. Scurlock, J.; Dayton, D.; Hames, B. Bamboo: An overlooked biomass resource? *Biomass Bioenergy* **2000**, *19*, 229–244. [CrossRef]

3. Fu, J. Chinese moso bamboo: Its importance. *Bamboo* **2001**, *22*, 5–7.

4. Peng, Z.; Lu, Y.; Li, L.; Zhao, Q.; Feng, Q.; Gao, Z.; Lu, H.; Hu, T.; Yao, N.; Liu, K.; et al. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat. Genet.* **2013**, *45*, 456. [CrossRef] [PubMed]

5. Wang, B.; Wei, W.; Liu, C.; You, W.; Niu, X.; Man, R. Biomass and carbon stock in Moso bamboo forests in subtropical China: Characteristics and implications. *J. Trop. For. Sci.* **2013**, *25*, 137–148.

6. Zhou, R.; Moshgabadi, N.; Adams, K.L. Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 16122–16127. [CrossRef] [PubMed]

7. Li, P.; Zhou, G.; Du, H.; Lu, D.; Mo, L.; Xu, X.; Shi, Y.; Zhou, Y. Current and potential carbon stocks in Moso bamboo forests in China. *J. Environ. Manag.* **2015**, *156*, 89–96. [CrossRef] [PubMed]

8. Zhao, H.; Gao, Z.; Wang, L.; Wang, J.; Wang, S.; Fei, B.; Chen, C.; Shi, C.; Liu, X.; Zhang, H.; et al. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*). *Gigascience* **2018**, *7*. [CrossRef]

9. Wang, T.; Wang, H.; Cai, D.; Gao, Y.; Zhang, H.; Wang, Y.; Lin, C.; Ma, L.; Gu, L. Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*). *Plant J.* **2017**, *91*, 684–699. [CrossRef]

10. Peng, Z.; Zhang, C.; Zhang, Y.; Hu, T.; Mu, S.; Li, X.; Gao, J. Transcriptome sequencing and analysis of the fast growing shoots of moso bamboo (*Phyllostachys edulis*). *PLoS ONE* **2013**, *8*, e78944. [CrossRef]

11. Zhang, H.; Wang, H.; Zhu, Q.; Gao, Y.; Wang, H.; Zhao, L.; Wang, Y.; Xi, F.; Wang, W.; Yang, Y.; et al. Transcriptome characterization of moso bamboo (*Phyllostachys edulis*) seedlings in response to exogenous gibberellin applications. *BMC Plant Biol.* **2018**, *18*, 125. [CrossRef]

12. Liu, L.; Cao, X.L.; Bai, R.; Yao, N.; Li, L.B.; He, C.F. Isolation and characterization of the cold-induced *Phyllostachys edulis* AP2/ERF family transcription factor, *peDREB1*. *Plant Mol. Bio. Rep.* **2012**, *30*, 679–689. [CrossRef]

13. Cui, X.W.; Zhang, Y.; Qi, F.Y.; Gao, J.; Chen, Y.W.; Zhang, C.L. Overexpression of a moso bamboo (*Phyllostachys edulis*) transcription factor gene *PheWRKY1* enhances disease resistance in transgenic *Arabidopsis thaliana*. *Botany* **2013**, *91*, 486–494. [CrossRef]

14. Huang, Z.; Zhong, X.J.; He, J.; Jin, S.H.; Guo, H.D.; Yu, X.F.; Zhou, Y.J.; Li, X.; Ma, M.D.; Chen, Q.B.; et al. Genome-wide identification, characterization, and stress-responsive expression profiling of genes encoding LEA (late embryogenesis abundant) proteins in moso bamboo (*Phyllostachys edulis*). *PLoS ONE* **2016**, *11*, e0165953. [CrossRef] [PubMed]

15. Li, Y.; Dai, C.; Hu, C.; Liu, Z.; Kang, C. Global identification of alternative splicing via comparative analysis of SMRT-and Illumina-based RNA-seq in strawberry. *Plant J.* **2017**, *90*, 164–176. [CrossRef] [PubMed]

16. Li, L.; Cheng, Z.; Ma, Y.; Bai, Q.; Li, X.; Cao, Z.; Wu, Z.; Gao, J. The association of hormone signalling genes, transcription and changes in shoot anatomy during moso bamboo growth. *Plant Biotechnol. J.* **2018**, *16*, 72–85. [CrossRef] [PubMed]

17. Wilusz, J.E.; Sunwoo, H.; Spector, D.L. Long noncoding RNAs: Functional surprises from the RNA world. *Genes Dev.* **2009**, *23*, 1494–1504. [CrossRef] [PubMed]

18. Mattick, J.S. RNA regulation: A new genetics? *Nat. Rev. Genet.* **2004**, *5*, 316–323. [CrossRef] [PubMed]

19. Choudhary, J.; Grant, S.G. Proteomics in postgenomic neuroscience: The end of the beginning. *Nat. Neurosci.* **2004**, *7*, 440–445. [CrossRef]

20. Armengaud, J. A perfect genome annotation is within reach with the proteomics and genomics alliance. *Curr. Opin. Microbiol.* **2009**, *12*, 292–300. [CrossRef]

21. Prasad, T.K.; Mohanty, A.K.; Kumar, M.; Sreenivasamurthy, S.K.; Dey, G.; Nirujogi, R.S.; Pinto, S.M.; Madugundu, A.K.; Patil, A.H.; Advani, J.; et al. Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res.* **2017**, *27*, 133–144. [CrossRef]

22. Mahesh, H.B.; Subba, P.; Advani, J.; Shirke, M.D.; Loganathan, R.M.; Chandana, S.L.; Shilpa, S.; Chatterjee, O.; Pinto, S.M.; Prasad, T.S.K.; et al. Multi-omics driven assembly and annotation of the sandalwood (*Santalum album*) genome. *Plant Physiol.* **2018**, *176*, 2772–2788. [CrossRef] [PubMed]

23. Armengaud, J. Reannotation of genomes by means of proteomics data. In *Methods in Enzymology*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 585, pp. 201–216.

24. Bock, T.; Chen, W.-H.; Ori, A.; Malik, N.; Silva-Martin, N.; Huerta-Cepas, J.; Powell, S.T.; Kastritis, P.L.; Smyshlyaev, G.; Vonkova, I.; et al. An integrated approach for genome annotation of the eukaryotic thermophile *Chaetomium thermophilum*. *Nucleic Acids Res.* **2014**, *42*, 13525–13533. [CrossRef] [PubMed]

25. Kim, M.-S.; Pinto, S.M.; Getnet, D.; Nirujogi, R.S.; Manda, S.S.; Chaerkady, R.; Madugundu, A.K.; Kelkar, D.S.; Isserlin, R.; Jain, S.; et al. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581. [CrossRef]

26. Silmon de Monerri, N.C.; Weiss, L.M. Integration of RNA-seq and proteomics data with genomics for improved genome annotation in Apicomplexan parasites. *Proteomics* **2015**, *15*, 2557–2559. [CrossRef] [PubMed]

27. Armengaud, J. Next-generation proteomics faces new challenges in environmental biotechnology. *Curr. Opin. Biotechnol.* **2016**, *38*, 174–182. [CrossRef] [PubMed]

28. Wang, X.; Ren, H.; Zhang, B.; Fei, B.; Burgert, I. Cell wall structure and formation of maturing fibres of moso bamboo (*Phyllostachys pubescens*) increase buckling resistance. *J. R. Soc. Interface* **2011**, *9*, 988–996. [CrossRef] [PubMed]

29. Cui, K.; He, C.Y.; Zhang, J.G.; Duan, A.G.; Zeng, Y.F. Temporal and spatial profiling of internode elongation-associated protein expression in rapidly growing culms of bamboo. *J. Proteome Res.* **2012**, *11*, 2492–2507. [CrossRef]

30. Wang, E.T.; Sandberg, R.; Luo, S.; Khrebtukova, I.; Zhang, L.; Mayr, C.; Kingsmore, S.F.; Schroth, G.P.; Burge, C.B. Alternative isoform regulation in human tissue transcriptomes. *Nature* **2008**, *456*, 470–476. [CrossRef]

31. Pan, Q.; Shai, O.; Lee, L.J.; Frey, B.J.; Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **2008**, *40*, 1413–1415. [CrossRef]

32. Nilsen, T.W.; Graveley, B.R. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **2010**, *463*, 457–463. [CrossRef]

33. Zhao, L.; Zhang, H.; Kohnen, M.V.; Prasad, K.; Gu, L.; Reddy, A.S. Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and Nanopore-based direct RNA sequencing. *Front. Genet.* **2019**, *10*, 253. [CrossRef]

34. Xu, Z.; Peters, R.J.; Weirather, J.; Luo, H.; Liao, B.; Zhang, X.; Zhu, Y.; Ji, A.; Zhang, B.; Hu, S.; et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *Plant J.* **2015**, *82*, 951–961. [CrossRef]

35. Wang, B.; Tseng, E.; Regulski, M.; Clark, T.A.; Hon, T.; Jiao, Y.; Lu, Z.; Olson, A.; Stein, J.C.; Ware, D. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat. Commun.* **2016**, *7*, 11708. [CrossRef]

36. Abdel-Ghany, S.E.; Hamilton, M.; Jacobi, J.L.; Ngam, P.; Devitt, N.; Schilkey, F.; Ben-Hur, A.; Reddy, A.S. A survey of the sorghum transcriptome using single-molecule long reads. *Nat. Commun.* **2016**, *7*, 11706. [CrossRef]

37. Li, L.; Hu, T.; Li, X.; Mu, S.; Cheng, Z.; Ge, W.; Gao, J. Genome-wide analysis of shoot growth-associated alternative splicing in moso bamboo. *Mol. Genet. Genom.* **2016**, *291*, 1695–1714. [CrossRef]

38. Kelemen, O.; Convertini, P.; Zhang, Z.; Wen, Y.; Shen, M.; Falaleeva, M.; Stamm, S. Function of alternative splicing. *Gene* **2013**, *514*, 1–30. [CrossRef]

39. Kurosaki, T.; Popp, M.W.; Maquat, L.E. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 406–420. [CrossRef]

40. Drechsel, G.; Kahles, A.; Kesarwani, A.K.; Stauffer, E.; Behr, J.; Drewe, P.; Rätsch, G.; Wachter, A. Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. *Plant Cell* **2013**, *25*, 3726–3742. [CrossRef]

41. Kalyna, M.; Simpson, C.G.; Syed, N.H.; Lewandowska, D.; Marquez, Y.; Kusenda, B.; Marshall, J.; Fuller, J.; Cardle, L.; McNicol, J.; et al. Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res.* **2011**, *40*, 2454–2469. [CrossRef]

42. Belostotsky, D.A.; Sieburth, L.E. Kill the messenger: mRNA decay and plant development. *Curr. Opin. Plant Biol.* **2009**, *12*, 96–102. [CrossRef]

43. Yang, X.; Coulombe-Huntington, J.; Kang, S.; Sheynkman, G.M.; Hao, T.; Richardson, A.; Sun, S.; Yang, F.; Shen, Y.A.; Murray, R.R.; et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **2016**, *164*, 805–817. [CrossRef]

44. Ezkurdia, I.; Rodriguez, J.M.; Carrillo-de Santa Pau, E.; Vázquez, J.; Valencia, A.; Tress, M.L. Most highly expressed protein-coding genes have a single dominant isoform. *J. Proteome Res.* **2015**, *14*, 1880–1887. [CrossRef]

45. Liu, Y.; Gonzàlez-Porta, M.; Santos, S.; Brazma, A.; Marioni, J.C.; Aebersold, R.; Venkitaraman, A.R.; Wickramasinghe, V.O. Impact of alternative splicing on the human proteome. *Cell Rep.* **2017**, *20*, 1229–1241. [CrossRef]

46. Reixachs-Sole, M.; Ruiz-Orera, J.; Alba, M.; Eyras, E. Ribosome profiling at isoform level reveals an evolutionary conserved impact of differential splicing on the proteome. *BioRxiv* **2019**. [CrossRef]

47. Chaudhary, S.; Jabre, I.; Reddy, A.S.; Staiger, D.; Syed, N.H. Perspective on alternative splicing and proteome complexity in plants. *Trends Plant Sci.* **2019**, *24*, 496–506. [CrossRef]

48. Brosch, M.; Saunders, G.I.; Frankish, A.; Collins, M.O.; Yu, L.; Wright, J.; Verstraten, R.; Adams, D.J.; Harrow, J.; Choudhary, J.S.; et al. Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and "resurrected" pseudogenes in the mouse genome. *Genome Res.* **2011**, *21*, 756–767. [CrossRef]

49. Tress, M.L.; Martelli, P.L.; Frankish, A.; Reeves, G.A.; Wesselink, J.J.; Yeats, C.; Ísólfur Ólason, P.; Albrecht, M.; Hegyi, H.; Giorgetti, A.; et al. The implications of alternative splicing in the ENCODE protein complement. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 5495–5500. [CrossRef]

50. Tress, M.L.; Bodenmiller, B.; Aebersold, R.; Valencia, A. Proteomics studies confirm the presence of alternative protein isoforms on a large scale. *Genome Biol.* **2008**, *9*, R162. [CrossRef]

51. Abascal, F.; Ezkurdia, I.; Rodriguez-Rivas, J.; Rodriguez, J.M.; del Pozo, A.; Vázquez, J.; Valencia, A.; Tress, M.L. Alternatively spliced homologous exons have ancient origins and are highly expressed at the protein level. *PLoS Comput. Biol.* **2015**, *11*, e1004325. [CrossRef]

52. Tress, M.L.; Abascal, F.; Valencia, A. Alternative splicing may not be the key to proteome complexity. *Trends Biochem. Sci.* **2017**, *42*, 98–110. [CrossRef]

53. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [CrossRef]

54. Elias, J.E.; Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214. [CrossRef]

55. Searle, B.C. Scaffold: A bioinformatic tool for validating MS/MS-based proteomic studies. *Proteomics* **2010**, *10*, 1265–1269. [CrossRef]

56. Zybailov, B.; Mosley, A.L.; Sardiu, M.E.; Coleman, M.K.; Florens, L.; Washburn, M.P. Statistical analysis of membrane proteome expression changes in *Saccharomyces c erevisiae*. *J. Proteome Res.* **2006**, *5*, 2339–2347. [CrossRef]

57. Zybailov, B.L.; Florens, L.; Washburn, M.P. Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors. *Mol. Biosyst.* **2007**, *3*, 354–360. [CrossRef]

58. Florens, L.; Carozza, M.J.; Swanson, S.K.; Fournier, M.; Coleman, M.K.; Workman, J.L.; Washburn, M.P. Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **2006**, *40*, 303–311. [CrossRef]

59. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [CrossRef]

60. Kanehisa, M.; Sato, Y.; Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **2016**, *428*, 726–731. [CrossRef]

61. Stone, J.M.; Walker, J.C. Plant protein kinase families and signal transduction. *Plant Physiol.* **1995**, *108*, 451–457. [CrossRef]

62. Filichkin, S.A.; Mockler, T.C. Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes. *Biol. Direct* **2012**, *7*, 20. [CrossRef]

63. Marquez, Y.; Brown, J.W.; Simpson, C.; Barta, A.; Kalyna, M. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* **2012**, *22*, 1184–1195. [CrossRef]

64. Jaffrey, S.R.; Wilkinson, M.F. Nonsense-mediated RNA decay in the brain: Emerging modulator of neural development and disease. *Nat. Rev. Neurosci.* **2018**, *19*, 715–728. [CrossRef]

65. Rödiger, A.; Agne, B.; Baerenfaller, K.; Baginsky, S. *Arabidopsis* proteomics: A simple and standardizable workflow for quantitative proteome characterization. *Methods Mol. Biol.* **2014**, *1072*, 275–288.

66. Castellana, N.E.; Payne, S.H.; Shen, Z.; Stanke, M.; Bafna, V.; Briggs, S.P. Discovery and revision of *Arabidopsis* genes by proteogenomics. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 21034–21038. [CrossRef]

67. Castellana, N.E.; Shen, Z.X.; He, Y.P.; Walley, J.W.; Cassidy, C.J.; Briggs, S.P.; Bafna, V. An automated proteogenomic method uses mass spectrometry to reveal novel genes in *Zea mays*. *Mol. Cell. Proteomics* **2014**, *13*, 157–167. [CrossRef]

68. Helmy, M.; Tomita, M.; Ishihama, Y. OryzaPG-DB: Rice proteome database based on shotgun proteogenomics. *BMC Plant Biol* **2011**, *11*, 63. [CrossRef]

69. Watanabe, K.A.; Homayouni, A.; Tufano, T.; Lopez, J.; Ringler, P.; Rushton, P.; Shen, Q.J. Tiling Assembly: A new tool for reference annotation-independent transcript assembly and novel gene identification by RNA-sequencing. *DNA Res.* **2015**, *22*, 319–329. [CrossRef]

70. Kay, R.; Barton, C.; Ratcliffe, L.; Matharoo-Ball, B.; Brown, P.; Roberts, J.; Teale, P.; Creaser, C. Enrichment of low molecular weight serum proteins using acetonitrile precipitation for mass spectrometry based proteomic analysis. *Rapid Commun. Mass Spectrom.* **2008**, *22*, 3255–3260. [CrossRef]

71. Sedgwick, G.W.; Fenton, T.F.; Thompson, J.R. Effect of protein precipitating agents on the recovery of plasma free aminoacids. *Can. J. Anim. Sci.* **1991**, *71*, 953–957. [CrossRef]

72. Chang, Y.-F.; Imam, J.S.; Wilkinson, M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* **2007**, *76*, 51–74. [CrossRef]

73. Lykke-Andersen, S.; Jensen, T.H. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 665–677. [CrossRef]

74. Schwerk, C.; Schulze-Osthoff, K. Regulation of apoptosis by alternative pre-mRNA splicing. *Mol. Cell* **2005**, *19*, 1–13. [CrossRef]