Cloning and Structural-Functional Analysis of the *MiMYB44L* Gene from Macadamia Nut Kernel (Postprint)

Authors: Xu Peng, Tan Qiujin, Song Haiyun, Zheng Shufang, Yang Xiaozhou, Huan Xiuju, Zhang Tao, Zhou Chunheng, Wei Yuanrong, Wang Wenlin

Date: 2024-05-15T00:00:00+00:00

Abstract

Macadamia integrifolia is a high economic value evergreen nut tree, with kernels rich in fatty acids, proteins, and other nutrients. To identify genes associated with nutrient formation in macadamia kernels, this study employed transcriptomics, gene cloning, fluorescence quantification, and bioinformatics techniques to mine regulatory genes from kernel transcriptomes of varieties 'Guire 1'and 'A4' , which exhibit significant differences in nutrient content. The results showed: (1) Transcriptome analysis revealed 1,667 up-regulated genes and 1,798 down-regulated genes in 'Guire 1' kernels compared to 'A4' , and KEGG enrichment analysis indicated that differential genes were primarily enriched in starch and sugar metabolism, amino acid biosynthesis, and carbon metabolism pathways. (2) A differentially expressed gene, gene-LOC122077931, encoding the R2R3-MYB transcription factor MYB44L, was identified, and MiMYB44L was cloned from 'Guire 1' kernels using RACE technology, with a full length of 1,165 bp, an ORF of 999 bp, encoding 332 amino acids. (3) Bioinformatics analysis demonstrated that MiMYB44L contains the SANT domain characteristic of the R2R3-MYB family, lacks signal peptides and transmembrane domains, and contains phosphorylation sites. (4) Protein content was measured in kernels of 10 macadamia varieties, revealing that MiMYB44L expression was significantly higher in high-protein varieties than in low-protein varieties, with an overall correlation coefficient of 0.54, reaching an extremely significant level. These findings provide theoretical guidance for elucidating the regulatory mechanism of MiMYB44L in macadamia nutrient formation.

Full Text

Cloning, Structural and Functional Analysis of the MiMYB44L Gene in Macadamia integrifolia Kernels

XU Peng, TAN Qiujin, SONG Haiyun, ZHENG Shufang, YANG Xiaozhou, HUAN Xiuju, ZHANG Tao, ZHOU Chunheng, WEI Yuanrong, WANG Wenlin*

Guangxi South Subtropical Agricultural Science Research Institute, Longzhou 532400, Guangxi, China

Abstract

Macadamia nut (*Macadamia integrifolia*) is an evergreen nut tree with high economic value, whose kernel is rich in nutrients such as fatty acids and proteins. To identify genes associated with nutrient formation in macadamia kernels, this study employed transcriptomics, gene cloning, quantitative PCR, and bioinformatics techniques to mine regulatory genes from the kernel transcriptomes of varieties 'Guire No. 1' and 'A4' , which exhibit significant differences in nutrient content. The results showed: (1) Transcriptome analysis revealed 1,667 up-regulated genes and 1,798 down-regulated genes in 'Guire No. 1' kernels compared to 'A4', with KEGG enrichment analysis indicating that differential genes were primarily involved in starch and sucrose metabolism, amino acid biosynthesis, and carbon metabolism pathways. (2) A differentially expressed gene, gene-LOC122077931, encoding the R2R3-MYB transcription factor MYB44L, was identified and cloned from 'Guire No. 1' kernels using RACE technology. The full-length gene was 1,165 bp with an ORF of 999 bp, encoding 332 amino acids. (3) Bioinformatics analysis confirmed that MiMYB44L contains the SANT domain characteristic of the R2R3-MYB family, lacks signal peptides and transmembrane domains, and contains phosphorylation sites. (4) Protein content determination in kernels of ten macadamia varieties revealed that MiMYB44L expression was significantly higher in high-protein varieties than in low-protein varieties, with an overall correlation coefficient of 0.54 (highly significant). These findings provide theoretical guidance for elucidating the regulatory mechanism of MiMYB44L in nutrient formation in macadamia kernels.

Keywords: *Macadamia integrifolia*, MiMYB44L, kernel, structure, expression analysis

---

Introduction

Macadamia integrifolia is a high-value evergreen nut tree native to the rainforests of Australia' s east coast, now commercially cultivated in frost-free tropical and subtropical regions worldwide due to its premium edible qualities (Yang

et al., 2023). Macadamia kernels primarily consist of fats, proteins, and various trace elements (Wojdylo et al., 2022). Previous studies have demonstrated substantial variation in kernel nutrient content among different macadamia germplasms, with coefficients of variation from highest to lowest being Ca > Zn > P > Mg > soluble sugars > Fe > protein > amino acids > K > fat (Tan et al., 2021a). Principal component analysis clustered major nutrients into four factors—amino acid composition, minerals, fats, and proteins—with a cumulative contribution rate of 87.91%, highlighting the significant role of protein components in macadamia nutritional composition. However, while numerous studies have investigated fatty acids (Richards et al., 2020) and trace element nutrition (De Silva et al., 2023) in macadamia, research on protein content and its regulation remains limited.

Proteins serve as crucial macromolecules in living organisms, not only providing nutritional value but also participating in various biological processes including plant growth, development, and stress responses as transcription factors (Zumajo-Cardona et al., 2023), enzymes (Gabrielli et al., 2022), and protein complexes (Gisriel et al., 2022). Among these, the R2R3-MYB transcription factor MYB44 is involved in stress resistance and multiple plant hormone signaling pathways, serving as a key regulator of phytohormone signal crosstalk (Wang et al., 2023). In the *Arabidopsis* genome, AtMYB44 belongs to subgroup 22 with conserved motifs 22.1 (TGLYMSPxSP) and 22.2 (GxFMxVVQEMIxxEVRSYM) (Stracke et al., 2001). It regulates stress responses and participates in ethylene and jasmonic acid signaling by interacting with proteins such as PYL9 and WRKY70, while being subject to transcriptional and post-translational regulation itself. For instance, AtMYB44 enhances *Arabidopsis* resistance to peach aphids by activating the ethylene signaling pathway through EIN2. MYB44 also competitively inhibits MYB340-bHLH2-NAC56 complex formation to regulate anthocyanin biosynthesis in purple-fleshed sweet potatoes. The low nucleosome density in the AtMYB44 promoter region facilitates regulation by various transcription factors. Studies on other proteins have revealed that tree peony PsMYB44 negatively regulates petal blotch distribution by inhibiting dihydroflavonol-4-reductase gene expression (Luan et al., 2023), while VaMYB44 from Chinese wild grape negatively regulates cold tolerance in transgenic *Arabidopsis* and grape plants (Zhang et al., 2022). These findings suggest that MYB44 transcription factors may participate in regulating nutrient content in macadamia fruits.

This study focuses on macadamia as an important economic fruit tree in Guangxi Zhuang Autonomous Region. Leveraging transcriptome data from varieties with significantly different kernel protein content and the 'Guire No. 1' genome data (Xia et al., 2022), we employed transcriptomics, gene cloning, and bioinformatics to identify genes with significant differential expression in kernels and analyze their structures and functions. Specifically, we aimed to address: (1) What types of major regulatory genes are associated with nutrient formation in macadamia kernels? (2) What specific structural and functional features do these genes possess? (3) How does kernel protein content accu-

mulation correlate with gene expression across different macadamia varieties? Through structural analysis, expression profiling, and correlation analysis between expression levels and protein content for the MYB transcription factor subgroup 22 member MiMYB44L, this study establishes a theoretical foundation for transcriptional regulation research on macadamia nutrient composition.

Materials and Methods

1.1 Plant Materials

Experiments were conducted from 2021 to 2023 at the Macadamia Germplasm Resource Nursery of Guangxi South Subtropical Agricultural Science Research Institute, Guangxi Academy of Agricultural Sciences. The region features a south subtropical monsoon climate with abundant heat, rainfall, and sunshine. Kernels from ten varieties with significantly different protein content ('JW' , 'SH' , 'A16' , 'Guire No. 1' , 'B7' , 'B4' , 'Royal Large Fruit' , 'B3' , 'A38' , and 'A4') were selected for protein content determination and RNA extraction. The MiMYB44L gene was screened and cloned based on the 'Guire No. 1' genome data and kernel transcriptome data from 'Guire No. 1' and 'A4' developed by our institute.

1.2 MiMYB44L Cloning and Expression Analysis

Gene cloning was performed using primers MiMYB44L-F1 (5 -TCCGTTTCTCTCATCTTCTC-3) and MiMYB44L-R1 (5 -GTCTGTCTTCCATCTTCAATC-3). For quantitative PCR, MiACTIN served as the internal reference gene with primers MiMYB44L-QF (5 -AATCGCTCGTCTCCTCTC-3) and MiMYB44L-QR (5 -GGCTTGAACCACCTGAAC-3), and MiACTIN-QF (5 -TCTTCATTGCCTGCACTCCAGA-3) and MiACTIN-QR (5 -TTCCACCTGAATGCCGTCTAGC-3). Reactions were conducted on a Bio-rad CFX96 real-time PCR system following the method of Song et al. (2023), with data analyzed using the 2^(-ΔΔCt) method (Song et al., 2023).

1.3 Protein Content Determination

Protein content was determined according to the Kjeldahl method specified in the National Food Safety Standard of the People' s Republic of China GB/T5009.5–2016 (Tan et al., 2021a).

1.4 Statistical Analysis

Data were processed using OriginPro 2021 and Office 365. All gene expression data represent means and standard errors from three biological replicates and three technical replicates. SPSS version 27 was used for one-way ANOVA, multiple comparison analysis, and correlation analysis (Song et al., 2023).

Results

2.1 Transcriptome Analysis of 'Guire No. 1' and 'A4' Macadamia Kernels

Transcriptome analysis of high-protein variety 'Guire No. 1' and low-protein variety 'A4' kernels was performed using the Illumina high-throughput sequencing platform at Biomarker Technologies (Beijing, China). A total of 39.34 Gb clean data were obtained, with each sample yielding at least 5.92 Gb and Q30 base percentages of 94.06% or higher. Compared to 'A4', 'Guire No. 1' kernels showed 1,667 up-regulated and 1,798 down-regulated genes. GO and KEGG enrichment analysis revealed that differential genes were primarily involved in starch and sucrose metabolism, amino acid biosynthesis, and carbon metabolism pathways (Figure 1 [Figure 1: see original paper]A). Notably, a differentially expressed gene, gene-LOC122077931 encoding the R2R3-MYB transcription factor MYB44L, showed significantly higher expression in 'Guire No. 1' than in 'A4' kernels (Figure 1 [Figure 1: see original paper]B).

2.2 Bioinformatics Analysis of MiMYB44L

Based on transcriptome analysis of high-protein 'Guire No. 1' and low-protein 'A4' kernels, a significantly differential MYB transcription factor was identified. Primers were designed by searching the 'Guire No. 1' genome, and the MiMYB44L gene was cloned from 'Guire No. 1' kernels. The gene had a full length of 1,165 bp, a cDNA coding ORF of 999 bp, and encoded 332 amino acids (Figure 2 [Figure 2: see original paper]). The encoded protein had a molecular weight of 36.3 kDa, isoelectric point of 8.19, molecular formula of $C_{1572}H_{2486}N_{464}O_{500}S_{13}$, total atom number of 5,035, instability coefficient of 62.96 (classifying it as an unstable protein), aliphatic index of 67.26, and hydrophobicity value of -0.636 (indicating a hydrophobic protein).

To further analyze MiMYB44L protein structure and specific motifs, NCBI online tools and the MEME analysis website were employed. Results showed that macadamia MiMYB44L had the highest sequence similarity with Telopea speciosissima TsMYB44 and Nelumbo nucifera NnMYB44, clustering into one subfamily (Figure 3 [Figure 3: see original paper]). The protein contained eight characteristic motifs, consistent with plant MYB44 family proteins.

Conserved domain analysis revealed that MiMYB44L shared high sequence similarity with other plant MYB proteins, all containing the SANT superfamily domain. This domain spans amino acids 8-173 and is characteristic of Myb superfamily proteins involved in transcription, RNA processing and modification, cell division, and chromosome partitioning. The SANT domain binds tandem repeats of telomeric DNA as part of the capping complex, with sequence-dependent binding to G/C-rich motifs [$C_{2\ 3}A(CA)_{1\ 6}$]. This extremely conserved domain is also found in regulatory transcriptional repressor complexes that bind DNA. Protein property analysis (Figure 5 [Figure 5: see original paper]) indicated MiMYB44L is a hydrophobic protein lacking transmembrane

domains and signal peptides, but containing phosphorylation sites primarily for serine, threonine, and tyrosine.

Secondary and tertiary structure analysis showed that MiMYB44L secondary structure comprises 86 amino acids (25.90%) in α-helices, 19 amino acids (5.72%) in extended strands, 10 amino acids (3.01%) in β-turns, and 217 amino acids (65.36%) in random coils (Figure 6 [Figure 6: see original paper]).

Analysis of protein content in kernels of ten varieties including 'Guire No. 1', 'A38', and 'JW' revealed significant differences. RNA extraction and analysis from kernels of these ten varieties showed MiMYB44L expression was highest in high-protein varieties such as 'JW' and 'SH', and lowest in varieties like 'A4' and 'B4' (Figure 7 [Figure 7: see original paper]). The correlation coefficient between protein content and expression level was 0.54 ($P = 0.002$), reaching a significant level.

Discussion

Research on the molecular mechanisms and cultivation techniques underlying transcription factor regulation of macadamia protein content remains limited, though related physiological and molecular biological studies have been conducted domestically and internationally. For instance, investigations of metabolite and endogenous hormone content changes during early macadamia inflorescence development revealed that most differentially accumulated metabolites—including lipids, sugars, alcohols, amino acids, organic acids, and nucleotides—decreased significantly in fruits as they developed, while increasing substantially in rachises (except for amino acids). Compared to weak inflorescences, strong inflorescences accumulated primarily lipids in fruits, while rachises accumulated lipids, nucleotides, and sugars (Yang et al., 2023). Based on macadamia germplasm collection, studies on different varieties have shown that kernel nutrients are predominantly fat, with 'GR1' showing the highest fat content at 78 $g \cdot 100g^{-1}$ and 'A4' the lowest at 69.90 $g \cdot 100g^{-1}$. Significant correlations exist among some fruit quality traits (Tan et al., 2021b), suggesting that factors regulating kernel protein content may also regulate other nutritional components.

MYB44 is a multifunctional transcription factor and a key downstream regulator of numerous plant hormone signals. Recent research demonstrates that in *Arabidopsis*, MYB44 regulates pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) by promoting EIN2 and MPK3/6 expression (Wang et al., 2023). Phloem-mobile MYB44 negatively regulates phosphate transporter 1 expression in *Arabidopsis* roots to modulate plant tolerance to low phosphate stress (Olukayode et al., 2023). The abscisic acid signaling pathway transcription factor ABI5 promotes heat stress-induced chlorophyll degradation in cucumber by modulating MYB44 stability (Liu et al., 2023). Beyond model plants, MYB44's diverse functions have been demonstrated in tropical trees, such as rubber tree HbMYB44 being induced by multiple hormone signals (Qin et al., 2022) and *Brassica campestris* BcMYB44 regulating both anthocyanin synthe-

sis and drought tolerance (Hao et al., 2022). These conserved structures confer similar functions across different plant species.

In this study, the cloned MiMYB44L protein possesses a conserved SANT domain and clusters with MYB44 from Telopea speciosissima and Nelumbo nucifera. Molecular structural analysis confirms it as a member of the plant MYB44 transcription factor family. The highly significant correlation between kernel protein content analysis and expression analysis indicates that MiMYB44L may be a regulatory factor for protein content in macadamia. With advances in genetics and molecular biology, additional and updated research methods can be introduced to identify its function in macadamia. Although no direct regulation of protein content by MYB44 transcription factors has been reported in plants, targeted silencing of the BjMYB28 transcription factor gene has been shown to directly regulate low-glucosinolate line development in *Brassica juncea* (Augustine et al., 2013). Recently, phylogenetic, collinearity, and biochemical analyses identified and characterized a Lauraceae-specific citral biosynthetic gene cluster containing MYB44 as a transcription factor and two alcohol dehydrogenases (ADHs) as modification enzymes, derived from species-specific tandem and proximal duplication events (Zhao et al., 2023). Tree peony PsMYB44 contains a complete C2 motif at the C-terminus of its amino acid sequence that affects anthocyanin biosynthesis, clustering in the MYB44L transcriptional repression branch. PsMYB44 is localized in the nucleus, and its spatiotemporal expression pattern is negatively correlated with blotch formation (Luan et al., 2023). These findings suggest that evolutionary and transgenic technologies can be employed to demonstrate MiMYB44L' s transcriptional regulatory role in macadamia kernel nutrient synthesis.

Future research should employ evolutionary analysis to examine MYB44 relationships across species and analyze duplication events. Genomic DNA from ten varieties with significantly different protein content should be extracted to amplify the full-length MiMYB44L gene from each variety, enabling analysis of exon and intron structures and correlation with nutritional content including protein and fatty acids. A pGREEN vector containing a 35S × 2 pro strong promoter and GFP tag has been constructed for tobacco transient expression assays to verify gene function. Building upon this study, establishing a macadamia genetic transformation system and using overexpression and knockout approaches to demonstrate MiMYB44L regulation of protein and other nutrient contents will be crucial for advancing macadamia quality breeding.

Conclusion

MiMYB44L protein contains the SANT domain characteristic of the R2R3-MYB family, representing a typical R2R3-MYB transcription factor. MiMYB44L expression is significantly higher in high-protein macadamia varieties than in low-protein varieties, with a correlation coefficient of 0.54 (significant level).

References

AUGUSTINE R, MUKHOPADHYAY A, BISHT NC, 2013. Targeted silencing of BjMYB28 transcription factor gene directs development of low glucosinolate lines in oilseed *Brassica juncea*[J]. Plant Biotechnol J, 11(7): 855-866.

DE SILVA AL, TRUEMAN SJ, KAMPER W, et al., 2023. Hyperspectral imaging of adaxial and abaxial leaf surfaces as a predictor of Macadamia crop nutrition[J]. Plants-Basel, 12(3): 558.

GABRIELLI F, ANTINUCCI M, TOFANELLI S, 2022. Gene structure evolution of the short-chain dehydrogenase/reductase (sdr) family[J]. Genes, 14(1): 110.

GISRIEL CJ, CARDONA T, BRYANT DA, BRUDVIG GW, 2022. Molecular evolution of far-red light-acclimated photosystem II[J]. Microorganisms, 10(7): 1270.

HAO YL, WANG JJ, HU CM, et al., 2022. Regulation of BcMYB44 on anthocyanin synthesis and drought tolerance in non-heading Chinese cabbage (*Brassica campestris* ssp. chinensis Makino)[J]. Horticulturae, 8(5): 351.

LIU WK, CHEN GL, HE MM, et al., 2023. ABI5 promotes heat stress-induced chlorophyll degradation by modulating the stability of MYB44 in cucumber[J]. Horticul Res, 10(6): uhad089.

LUAN YT, CHEN ZJ, TANG YH, et al., 2023. Tree peony PsMYB44 negatively regulates petal blotch distribution by inhibiting dihydroflavonol-4-reductase gene expression[J]. Ann Bot, 131(2): 323-334.

OLUKAYODE T, CHEN J, ZHAO Y, et al., 2023. Phloem-mobile MYB44 negatively regulates expression of phosphate transporter 1 in *Arabidopsis* roots[J]. Plants-Basel, 12(20): 3617.

QIN B, FAN SL, YU HY, et al., 2022. HbMYB44, a rubber tree MYB transcription factor with versatile functions in modulating multiple phytohormone signaling and abiotic stress responses[J]. Front Plant Sci, 13: 893896.

RICHARDS TE, KAMPER W, TRUEMAN SJ, et al., 2020. Relationships between nut size, kernel quality, nutritional composition and levels of outcrossing in three Macadamia cultivars[J]. Plants-Basel, 9(2): 228.

SONG HY, ZHANG T, HE P, et al., 2023. Cloning and expression analysis of MibZIP1 from *Macadamia integrifolia*[J]. Bull Bot Res, 43(1): 131-139.

STRACKE R, WERBER M, WEISSHAAR B, 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*[J]. Curr Opin Plant Biol, 4(5): 447-456.

TAN QJ, CHEN HS, WEI YR, et al., 2021a. Nutritional analysis and evaluation of kernel of *Macadamia integrifolia*[J]. J Chin Cereals Oils Assoc, 36(2): 150-154.

TAN QJ, WEI YR, HUANG XY, et al., 2021b. Analysis of fruit characteristics

and nutrients of 10 accessions of *Macadamia integrifolia*[J]. J Fruit Sci, 38(5): 672-680.

WANG Z, LI X, YAO X, et al., 2023. MYB44 regulates PTI by promoting the expression of EIN2 and MPK3/6 in *Arabidopsis*[J]. Plant Commun, 4(6): 100628.

WOJDYLO A, TURKIEWICZ IP, TKACZ K, et al., 2022. Nuts as functional foods: Variation of nutritional and phytochemical profiles and their in vitro bioactive properties[J]. Food Chem, 15: 100418.

XIA CX, JIANG SR, TAN QJ, et al., 2022. Chromosomal-level genome of macadamia (*Macadamia integrifolia*)[J]. Trop Plants, 1(1): 1-9.

YANG WH, LI J, XIANG FL, et al., 2023. Comparative profiling of primary metabolites and endogenous hormones in different Macadamia infructescences during its early development[J]. Sci Horticul, 319: 112185.

ZHANG HJ, HU YF, GU B, et al., 2022. VaMYB44 transcription factor from Chinese wild *Vitis amurensis* negatively regulates cold tolerance in transgenic *Arabidopsis thaliana* and *V. vinifera*[J]. Plant Cell Rep, 41(8): 1673-1691.

ZHAO Y, CHEN Y, GAO M, et al., 2023. Alcohol dehydrogenases regulated by a MYB44 transcription factor underlie Lauraceae citral biosynthesis[J]. Plant Physiol, 194(3).

ZUMAJO-CARDONA C, GABRIELI F, ANIRE J, et al., 2023. Evolutionary studies of the bHLH transcription factors belonging to MBW complex: their role in seed development[J]. Ann Bot, 132(3): 383-400.

Note: Figure translations are in progress. See original paper for figures.