Vietnam Journal of Agricultural Sciences

ISSN 2588-1299

VJAS 2019; 2(2): 370-375
https://doi.org/10.31817/vjas.2019.2.2.02

Identification, Structural Analysis, and Expression Profile of Genes Related to Starch Metabolism in Cassava (*Manihot esculenta* Crantz)

Chu Duc Ha[1], Nguyen Van Loc[2], Lai Thi Uyen[1,2], Pham Phuong Thu[3] & Pham Thi Ly Thu[1]

[1]Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences, Hanoi 129810, Vietnam
[2]Faculty of Agronomy, Vietnam National University of Agriculture, Hanoi 131000, Vietnam
[3]Faculty of Biology-Agricultural Technology, Hanoi Pedagogical University 2, Vinh Phuc 283460, Vietnam

Abstract

Starch metabolism is known to be an important pathway in the growth and development of plants. This study was conducted to investigate the genome-wide identification and structural analysis of genes encoding uridine diphosphate glucose pyrophosphorylase (UGPase), a key enzyme in starch synthesis in cassava, and to analyze the expression profiles of these genes based on publicly available RNA-seq data. A total of 11 members were found in the UGPase gene family (*MeUGP*) in cassava. Ten of the *MeUGP* genes were successfully mapped onto the chromosomes of the current cassava genome assembly. Based on their nucleotide sequences, the lengths of the genomic DNA sequences of the *MeUGP* genes ranged from 3,200 to 11,601bp, while the size of the coding sequence (CDS) varied from 831 to 3,654bp. According to the recent RNA-seq data, we found that a majority of the *MeUGP* genes were expressed in at least 1 tissue under normal conditions. Interestingly, *MeUGP4* was greatly expressed in the shoot apical meristem, while *MeUGP10* was more specific in the root apical meristem. The expression profiles of these *MeUGP* genes should be carried out in various conditions in further studies.

Keyword

Starch metabolism, identification, gene family, UGPase, bioinformatics, cassava

Received: May 25, 2018
Accepted: June 6, 2019

Correspondence to
hachu_amser@yahoo.com/
nvloc@vnua.edu.vn

ORCID
Nguyen Van Loc
https://orcid.org/0000-0003-3528-0551

Introduction

Cassava (*Manihot esculenta* Crantz) is considered to be a major multifunctional crop in Vietnam. Many parts of the cassava plant can be used as a staple food for humans, animal feed, and raw materials for industrial production (Ceballos *et al.,* 2004; Cutting, 1978). Among them, starch, a major storage form of glucose, is considered to serve as an important resource in providing energy for various biological processes during the growth and development of cassava plants (Li *et al.,* 2016).

The basic pathway of starch metabolism begins with CO_2 fixation, followed by transitory starch degradations, sucrose synthesis, and starch synthesis in the storage organs in the plant (Saithong *et al.,* 2013). A number of enzymes involved in starch metabolism in tuber crops have been found (Van Harsselaar *et al.,* 2017). Several recent studies focused on the proteomic profiling and functional characterization of several genes associated with starch metabolism in cassava (Chen *et al.,* 2015; Wang *et al.,* 2016). Among them, UDP-glucose pyrophosphorylase (UGPase) is a core enzyme that was clearly determined to have an important role in starch regulation in the tuber (Van Harsselaar *et al.,* 2017). Unfortunately, the role of UGPase in cassava is still poorly understood. Therefore, this study was conducted to investigate the genome-wide identification and structural analysis of genes encoding UGPase, a key enzyme in starch synthesis in cassava, and to analyze the expression of these genes based on the publicly available RNA-Seq data.

Materials and Methods

Materials

The newest genome and proteome data of the 'AM560-2' cassava cultivar (BioProject: PRJNA234389), an S3 line bred at CIAT from 'MCOL1505' (Bredeson *et al.,* 2016), were downloaded and used. These materials are available on the Phytozome website (https://phytozome.jgi.doe.gov/) (Goodstein *et al.,* 2012).

Methods

Identification of UGPase in cassava

At3G03250, UGPase1 in *Arabidopsis thaliana*, collected from a previous study (Van Harsselaar *et al.,* 2017), was used as the seed sequence to conduct a BlastP search in the current proteome of cassava (Bredeson *et al.,* 2016) in Phytozome 12 (Goodstein *et al.,* 2012). All identified proteins, with E-values $<1\times10^{-6}$, were then confirmed by the presence of the UGPase domain in the Pfam database (Finn *et al.,* 2016). The protein sequences were collected for further analysis.

Annotation of the MeUGP genes in cassava

The general annotation information of the of *MeUGP* genes, including GeneID, locus name, and TranscriptID, were collected from the cassava genome assembly in NCBI (Bredeson *et al.,* 2016). The genomic DNA sequence and the coding DNA sequence of each *MeUGP* gene were downloaded and used in further analyses.

Chromosomal distribution of the MeUGP genes in the cassava genome

The location of each *MeUGP* gene was retrieved from the cassava genome (Bredeson *et al.,* 2016) in Phytozome (Goodstein *et al.,* 2012). The physical size of each cassava chromosome was determined based on the current cassava genome assembly (BioProject: PRJNA234389) in NCBI (Bredeson *et al.,* 2016). The distributions of the *MeUGP* genes were drawn using Adobe Illustrator.

Structural analysis of the MeUGP genes in cassava

The genomic DNA sequence, CDS, and GC content of the *MeUGP* gene family were analyzed using BioEDIT software (Hall, 1999). The exon/intron organizations of the *MeUGP* genes were found in GSDS 2.0 (http://gsds.cbi.pku.edu.cn/) (Hu *et al.,* 2015).

Phylogenetic analysis of UGPase in cassava

Full-length protein sequences of the UGPases were used to construct an unrooted phylogenetic tree using the neighbor-joining method in MEGA 7.0 (Kumar *et al.,* 2016). The resulting tree was then drawn in Adobe Illustrator.

Expression profiles of the MeUGP genes in cassava under normal conditions

The expression profiles of the *MeUGP* gene family in various organs/tissues under normal conditions were analyzed based on previous RNA-seq data (Wilson *et al.*, 2017). In this study, five tissues, namely fibrous root, root apical meristem (RAM), shoot apical meristem (SAM), friable embryogenic callus (FEC), and organized embryogenic structure (OES) (Wilson *et al.*, 2017), were studied. The criteria of detection followed Wilson *et al.* (2017) in that FPKM values of 1 were indicated to represent "below the limit of detection", whereas FPKM values of 10 corresponded to "expressed". An expression value of ≥100 FPKM corresponded to "highly expressed".

Results and Discussion

Identification, annotation, and chromosomal distribution of genes encoding UGPase in cassava

To provide initial information about the genes encoding UGPase in cassava, At3G0325 (AtUGPase1) was used for a BlastP search against the current proteome of cassava (Bredeson *et al.*, 2016) in Phytozome (Goodstein *et al.*, 2012) and annotated in the genome assembly of cassava in NCBI (Bredeson *et al.*, 2016). As a result, a total of 11 genes encoding UGPase (*MeUGP*) were found in the cassava genome (**Table 1**).

Next, the distribution of the *MeUGP* genes was identified in the current cassava assembly. As a result, out of the 11 members of the *MeUGP* family, 10 genes were mapped onto 6 chromosomes of the cassava genome with different rates of distribution. Among them, chromosomes 1 and 2 each had three *MeUGP* genes while chromosomes 13, 15, 16, and 18 each had one *MeUGP* gene. Interestingly, 2 genes, *MeUGP4* and *MeUGP9*, were located on the subtelomeric regions of cassava chromosomes 2 and 16, respectively (**Figure 1**). Previously, the regions near centromeres (pericentromere) and near telomeres (subtelomere) were suggested to be more permissive to the expansion of segmental duplications (Emanuel & Shaikh, 2001). Thus, we also predicted that these genes may have played important roles in various biological processes during the evolution of the cassava plant.

Only one gene, *MeUGP11* (Manes.S044400.1), was not found in the cassava genome (**Figure 1**). This result could be explained by the fact that this newest cassava assembly is ~582.25Mb set on 18 chromosomes, while approximately 2001 scaffolds have not yet been mapped onto the chromosomes (Bredeson *et al.*, 2016). Previously, the expected cassava genome size was estimated to be approximately 772Mb (Awoleye *et al.*, 1994). Thus, we believe that *MeUGP11* could be mapped on the cassava genome assembly in the future.

Structural analysis of the MeUGP family in cassava

In this study, we also analyzed the structure of the *MeUGP* genes in cassava using various web-based tools. Firstly, the genomic DNA sequence of the genes encoding UGPase in cassava ranged from 3,200 (*MeUGP10*, *Manes.18G046300.1*) to 11,601bp (*MeUGP11*, *Manes.S044400.1*) in length, while the GC content varied from 32.15 (*MeUGP2*, *Manes.01G091700.1*) to 40.97% (*MeUGP4*, *Manes.02G001000.1*) (**Table 2**). Additionally, the coding sequence (CDS) length of the *MeUGP* family was found to be from 831 (*MeUGP10*) to 3,654bp (*MeUGP4*) (**Table 2**).

For further structural analysis, the exon/intron organization of the *MeUGP* gene family was also retrieved based on the Gene Structure Display Server (GSDS) tool (Hu *et al.*, 2015). As shown in **Figure 2**, *MeUGP* genes classified in the same clade often shared the same structure. For example, *MeUGP1* and *MeUGP4* contained 13 exons/12 introns, while '*MeUGP2* and *MeUGP5*', '*MeUGP8* and *MeUGP6*', and '*MeUGP11* and *MeUGP9*' seemed to share the same gene organization, although their genomic DNA sequences were different. These results showed that the structure of the genes encoding *UGP* in cassava was quite complicated, and the separation of exons in the gene family during the pressure of natural selection as previously described (Gorlova *et al.*, 2014).

Table 1. Annotation of putative genes encoding UGPase in cassava

No.	Gene name	Locus name	Gene ID	Protein ID	Transcript ID
1	*MeUGP1*	LOC110627580	*Manes.01G055700.1*	XP_021629622.1	XM_021773930.1
2	*MeUGP2*	LOC110620717	*Manes.01G091700.1*	XP_021620237.1	XM_021764545.1
3	*MeUGP3*	LOC110601941	*Manes.01G184000.1*	XP_021595082.1	XM_021739390.1
4	*MeUGP4*	LOC110609003	*Manes.02G001000.1*	XP_021603997.1	XM_021748305.1
5	*MeUGP5*	LOC110608695	*Manes.02G046900.1*	XP_021603663.1	XM_021747971.1
6	*MeUGP6*	LOC110609353	*Manes.02G082800.1*	XP_021604565.1	XM_021748873.1
7	*MeUGP7*	LOC110629968	*Manes.13G105300.1*	XP_021632906.1	XM_021777214.1
8	*MeUGP8*	LOC110601811	*Manes.15G118600.1*	XP_021594829.1	XM_021739137.1
9	*MeUGP9*	LOC110603572	*Manes.16G006900.1*	XP_021597024.1	XM_021741332.1
10	*MeUGP10*	LOC110606022	*Manes.18G046300.1*	XP_021600429.1	XM_021744737.1
11	*MeUGP11*	LOC110607682	*Manes.S044400.1*	XP_021602521.1	XM_021746829.1

Note: MeUGP: Genes encoding UGPase in cassava; ID: Identifier; LOC: Locus.

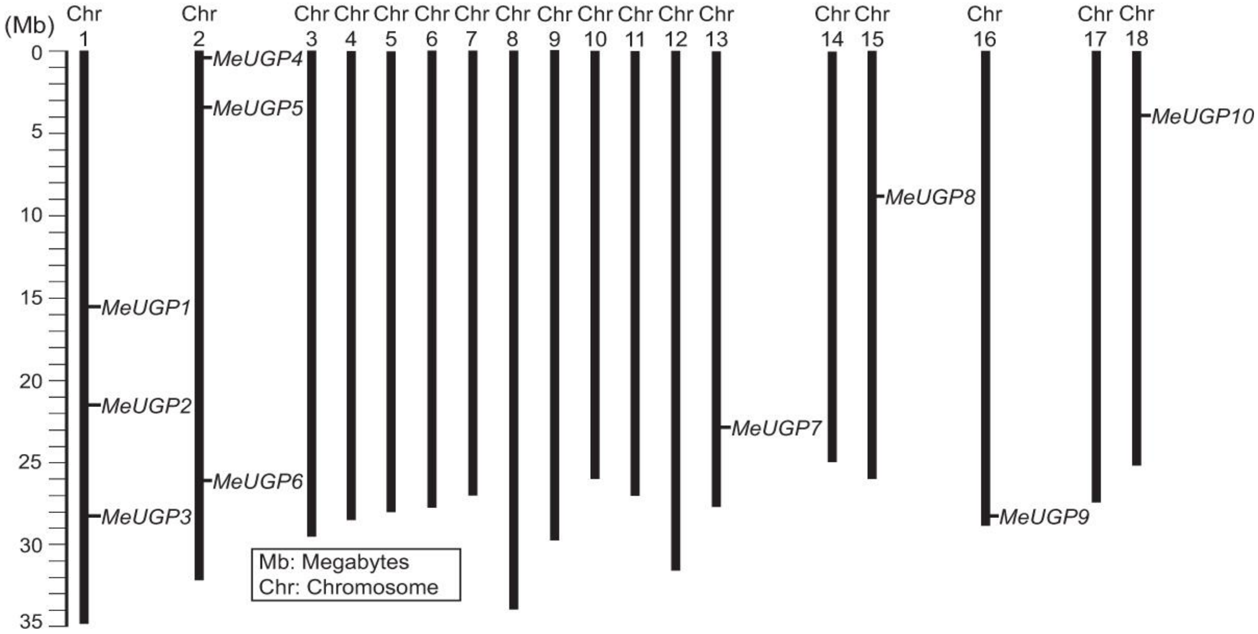

Figure 1. The chromosomal distribution of genes encoding UGPase in the current cassava genome

Table 2. Structure of the *MeUGP* genes in cassava

#	Gene name	Genomic DNA length	GC content	CDS length
1	*MeUGP1*	3568	39.46	1824
2	*MeUGP2*	5692	32.15	2034
3	*MeUGP3*	8000	32.95	1896
4	*MeUGP4*	4271	40.97	3654
5	*MeUGP5*	7306	37.76	2256
6	*MeUGP6*	10018	34.83	1827
7	*MeUGP7*	3951	32.37	906
8	*MeUGP8*	9176	36.16	3186
9	*MeUGP9*	11588	35.27	3441
10	*MeUGP10*	3200	36.63	831
11	*MeUGP11*	11601	35.26	2901

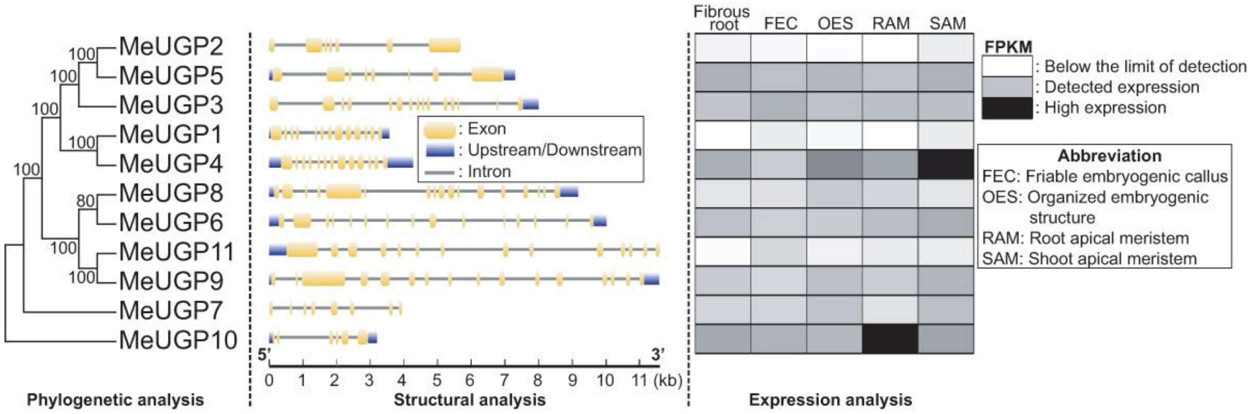

Figure 2. The exon/intron organization and the expression profiles of the *MeUGP* genes arranged based on the phylogenetic analysis

As shown in **Figure 2**, the majority of the *MeUGP* genes were expressed in at least 1 tissue. Among them, *MeUGP4* was strongly expressed in the SAM, while *MeUGP10* seemed to be specific in the RAM. Two genes, *MeUGP1* and *MeUGP2*, were not expressed in any tissues. Previously, the expression profiles of several genes encoding sucrose transporters (SWEET) in 11 tissues of the cassava plant under normal conditions were analyzed. *MeSWEET7* was found to be expressed in the FEC and OES, while *MeSWEET18* was specific in the RAM. Additionally, *MeSWEET26* and *MeSWEET27* were also expressed in the RAM and SAM (Ha *et al.*, 2017). Taken together, these results indicated that *MeUGP4* and *MeUGP10* may play a critical role in starch metabolism in the apical meristem, and thus, be involved in the growth and development of cassava plants.

Conclusions

Eleven members were identified in the UGPase gene family in the cassava genome. Ten of the genes were found to be located on six of cassava's eighteen chromosomes. *MeUGP4* and *MeUGP9* were mapped on the subtelomeric regions of chromosomes 2 and 16, respectively.

The size of the genomic DNA sequences of the *MeUGP* genes varied from 3,200 to 11,601bp. The CDS length of the *MeUGP* genes ranged from 831 to 3,654bp. Additionally, the *MeUGP* genes contained complicated exon/intron organizations.

Based on previous RNA-seq data, most of the *MeUGP* genes were found to be expressed in at least 1 tissue. *MeUGP4* was highly expressed in the SAM, while *MeUGP10* was more specific in the RAM. Two genes, *MeUGP1* and *MeUGP2*, were not expressed in any tissues.

References

Awoleye F., van Duren M., Dolezel J. & Novak F. J. (1994). Nuclear DNA content and in vitro induced somatic polyploidization cassava (*Manihot esculenta* Crantz) breeding. Euphytica. 76(3): 195-202.

Bredeson J. V., Lyons J. B., Prochnik S. E., Wu G. A., Ha C. M., Edsinger-Gonzales E., Grimwood J., Schmutz J., Rabbi I. Y., Egesi C., Nauluvula P., Lebot V., Ndunguru J., Mkamilo G., Bart R. S., Setter T. L., Gleadow R. M., Kulakow P., Ferguson M. E., Rounsley S. & Rokhsar D. S. (2016). Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nature Biotechnology. 34(5): 562-570.

Ceballos H., Iglesias C. A., Pérez J. C. & Dixon A. G. O. (2004). Cassava breeding: Opportunities and challenges. Plant Molecular Biology. 56(4): 503-516.

Chen X., Xia J., Xia Z., Zhang H., Zeng C., Lu C., Zhang W. & Wang W. (2015). Potential functions of microRNAs in starch metabolism and development revealed by miRNA transcriptome profiling of cassava cultivars and their wild progenitor. BMC Plant Biology. 15(1): 33.

Cutting W. A. (1978). Cassava - A valuable food but a possible poison. Tropical Doctor. 8(3): 102-103.

Ha C. D., Dung T. L., Huyen T. T. & Thu L. P. (2017). Evolutionary analysis and expression profiling of the sweet sugar transporter gene family in cassava (*Manihot esculenta* Crantz). The Journal of Science of Hanoi National University of Education. 62(10): 91-99 (in Vietnamese).

Emanuel B. S. & Shaikh T. H. (2001). Segmental duplications: an 'expanding' role in genomic instability and disease. Nature Reviews Genetics. 2(10): 791-800.

Finn R. D., Coggill P., Eberhardt R. Y., Eddy S. R., Mistry J., Mitchell A. L., Potter S. C., Punta M., Qureshi M., Sangrador-Vegas A., Salazar G. A., Tate J. & Bateman A. (2016). The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Research. 44(D1): D279-D285.

Goodstein D. M., Shu S., Howson R., Neupane R., Hayes R. D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N. & Rokhsar D. S. (2012). Phytozome: A comparative platform for green plant genomics. Nucleic Acids Research. 40(Database issue): D1178-D1186.

Gorlova O., Fedorov A., Logothetis C., Amos C. & Gorlov I. (2014). Genes with a large intronic burden show greater evolutionary conservation on the protein level. BMC Evolutionary Biology. 14(1): 50.

Hall T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98.

Hu B., Jin J., Guo A. Y., Zhang H., Luo J. & Gao G. (2015). GSDS 2.0: An upgraded gene feature visualization server. Bioinformatics. 31(8): 1296-1297.

Kumar S., Stecher G. & Tamura K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution. 33(7): 1870-1874.

Li Y. Z., Zhao J. Y., Wu S. M., Fan X. W., Luo X. L. & Chen B. S. (2016). Characters related to higher starch accumulation in cassava storage roots. Scientific Reports. 6: 19823.

Saithong T., Rongsirikul O., Kalapanulak S., Chiewchankaset P., Siriwat W., Netrphan S., Suksangpanomrung M., Meechai A. & Cheevadhanarak S. (2013). Starch biosynthesis in cassava: a genome-based pathway reconstruction and its exploitation in data integration. BMC System Biology. 7: 75.

Van Harsselaar J. K., Lorenz J., Senning M., Sonnewald U. & Sonnewald S. (2017). Genome-wide analysis of starch metabolism genes in potato (*Solanum tuberosum* L.). BMC Genomics. 18(1): 37.

Wang X., Chang L., Tong Z., Wang D., Yin Q., Wang D., Jin X., Yang Q., Wang L., Sun Y., Huang Q., Guo A. & Peng M. (2016). Proteomics profiling reveals carbohydrate metabolic enzymes and 14-3-3 proteins play important roles for starch accumulation during cassava root tuberization. Scientific Reports. 6: 19643.

Wilson M. C., Mutka A. M., Hummel A. W., Berry J., Chauhan R. D., Vijayaraghavan A., Taylor N. J., Voytas D. F., Chitwood D. H. & Bart R. S. (2017). Gene expression atlas for the food security crop cassava. New Phytologist. 213(4): 1632-1641.