Genome-wide identification and characterization of glutathione S-transferase gene family in Musa acuminata L. AAA group and gaining an insight to their role in banana fruit development

Mahesh Basantani (mkbasantani@gmail.com)
Shri Ramswaroop Memorial University
Swati Vaish
Shri Ramswaroop Memorial University
Reshma Parveen
Shri Ramswaroop Memorial University
Divya Gupta
Shri Ramswaroop Memorial University

Research Article

Keywords: Banana genome hub, Glutathione S-transferase, Genome wide, Bioinformatics, Multifunctional protein

Posted Date: February 15th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1342566/v1

Abstract

Glutathione S-transferases are a multifunctional protein superfamily that is involved in diverse plant functions such as defense mechanisms, signaling, stress response, secondary metabolism, plant growth and development. Although the banana whole genome sequence is available but the distribution of GST genes on banana chromosomes, their subcellular localization, gene structure, their evolutionary relation with each other, conserved motifs and their roles in banana are still unknown. A total of 50 full length GST genes with the canonical thioredoxin fold have been identified belonging to seven GST classes namely tau, phi, zeta, lambda, DHAR, EF1G and GHR. The 50 GST genes were distributed into 11 banana chromosomes. All the MaGSTs were majorly localized in the cytoplasm. Gene architecture showed the conservation of exon numbers in individual GST classes. MEME analyses revealed few class specific motifs and many motifs were found in all the GST classes. Multiple sequence alignment of banana GST amino acid sequences with rice, Arabidopsis and soybean sequences revealed the Ser and Cys as conserved catalytic residue. Gene duplication analyses showed the tandem duplication as a driving force for GST gene family expansion in banana. Cis-regulatory element analysis showed the dominance of light responsive element followed by stress and hormone responsive element. Expression profiling analyses was also done by RNA-seq data. It was observed that MaGSTs are involved in various stages of fruit development. MaGSTU1 was highly upregulated. The comprehensive and organized studies of MaGST genes family provides groundwork for further functional analysis of MaGST genes in banana at molecular level and further for plant breeding approaches.

1. Introduction

Musa acuminata (L.) popularly known as 'Banana' belongs to the Musaceae family distributed in the hot, tropical and sub tropical regions worldwide. The presence of several vitamins (vitamin A, vitamin B1, B2, B3, B6 and vitamin C), minerals (especially potassium), good quantity of starch and fiber enhances the nutritional value of banana. Due to its antioxidant property its consumption is helpful against infectious disease and is also used to treat hypertension, cancer, coronary disease, diarrhea etc. (Kumar et al. 2020). Banana is world's major cash crop. India ranks first in bananas production worldwide. In the year of 2020, bananas production in India was 31.5 million tonnes that accounts for 26.43% of the world's bananas production (Link). China, Indonesia, Brazil, and Ecuador are other leading countries in banana production after India. The world's total bananas production was estimated at 119 million tonnes in 2020 (World DATA ATLAS). In India, Andhra Pradesh, Gujarat, Tamil Nadu, Maharashtra, Kerala, Uttar Pradesh, Bihar and Madhya Pradesh contribute more than 70 per cent of the country's banana production.

Considering the economic and nutritional significance of *M. acuminata*, it could be a fascinating challenge for plant breeders to enhance its fruit quality, to develop more resistant crop against diverse biotic and abiotic stress through transgenic approaches. The availability of sequenced genome of *M. acuminata* (DH- Pahang, AAA group) (D'Hont et al. 2012) provides an opportunity for researchers to search and characterize diverse gene families which are functionally important. In the recent studies many gene families like DCL, AGO and RDR gene families (Ahmed et al. 2021); Calcineurin B-Like (CBL) genes (Xiong et al.2021); cellulose synthase-like (Csl) gene family (Yuan et al. 2021); TCP gene family (Wang et al. 2020) and Aquaporin gene family (Hu et al. 2015) have been identified and well characterized in *M. acuminata*.

Glutathione S-transferases (GSTs) are an inbuilt antioxidant enzymatic defense system of plants, that works downstream of Cyt P450. GST enzyme superfamily is primarily occupied in scavenging diverse chemical compounds present in the soil or in the environment by conjugating glutathione (GSH), a natural non enzymatic antioxidant defense system in plants (Song et al. 2021), to a hydrophobic substrate to make chemical compound more hydrophilic to be expelled from the cell through vacuole (Basantani et al. 2007). GSTs are the key enzymes in plant growth and development, secondary metabolism, anythocyanin accumulation (Shao et al. 2021), signal transduction pathways (Nianiou-Obeidat et al. 2017), tetrapyrole metabolism and retrograde signaling (Sylvestre-Gonon et al 2020), detoxification of reactive carbonyl species (RCS) (Mano et al. 2019) and against various biotic and abiotic stresses. GST s are characterized by the presence of canonical thioredoxin fold at the highly conserved N-terminal domain which is dominantly consists of α- helices and β-strands with a β1α1β2α2β3β4α3 topology. It possesses G-site for glutathione binding. The variable C-terminal domains is consists of all α- helices and possesses H-site for secondary hydrophobic substrate binding (Vaish et al. 2020).

Based on GSTs' subcellular localization they are classified into three distinct superfamilies namely cytoplasmic, mitochondrial and microsomal. Cytopalsmic and mitochondrial GSTs are soluble; and microsomal GSTs (MAPEGs) are membrane associated proteins involved in eicosanoid and glutathione metabolism (Lallement et al. 2014a). Soluble plant GSTs are further categorized into 14 different classes based on sequence similarity, genomic organization, immunological cross reactivity and functions viz. tau (U), phi (F), theta (T), zeta (Z), lambda (L), Dehydroascorbate reductase (DHAR), Tetrachloro-hydroquinone dehalogenase (TCHQD), Elongation factor 1Bγ (EF1Bγ), microsomal prostaglandin E synthase type 2 (mPGES2), Glutathionyl hydroquinone reductase (GHR), iota, hemerythrin, metaxin and Ure2p (Lallement et al. 2014b). On the basis of active site residue GSTs are classified as serinyl and cysteinyl GSTs (Lallement et al. 2014a). Tau, phi, theta, zeta and TCHQD classes come under serinyl GSTs containing Ser as active site residue (Sylvestre-Gonon et al. 2019). The serinyl GSTs are dimeric proteins (homo or hetero forms). DHAR, lambda, GHR, iota, hemerythrin, mPGES-2 and metaxin GSTs are cysteinyl GSTs possessing Cys as catalytic residue and are monomeric in nature. The cysteinyl and serinyl GSTs function differently: Tau, phi, theta, zeta, TCHQD and EF1G GST proteins perform glutathionylation reaction i.e. conjugation of glutathione to xenobiotic compounds to make them more soluble and their vacuolar excretion from the plant cell; whereas cysteinyl GSTs are involved in deglutathionylation reactions and also catalyze the dehydroascorbate reduction. Tau, phi, theta and zeta are plant- specific GSTs.

Genome-wide analyses have identified 39 GSTs in *Cucumis melo var. saccharinus* (Song et al. 2021), 92 GSTs in *Medicago truncatula* (Hasan et al. 2021), 32 GSTs in *Cucurbita maxima* (Kayum et al. 2018), 330 GSTs *Triticum aestivum* (Wang et al. 2019), 82 GSTs in *Raphanus sativus* (Gao et al. 2020), 51 GSTs in chickpea (Ghangal et al. 2020) and 31 GSTs in *Vigna radiata* (Vaish et al. 2018). Since GSTs are functionally pivotal in plant growth and development, responsive against diverse biotic and abiotic stresses and there is no information related to genome-wide identification and characterization of GST gene family in *M. acuminata*, hence it was selected for genome-wide analyses in *M. acuminata.* The publically available whole-genome sequence of banana enables us to perform genome-wide analysis of GST gene family in banana using the integrated bioinformatics tools that are a cost effective, times and labor

saving approach. The present proposed the details of banana GSTs' physicochemical characteristics, sub cellular localization, chromosomal localization, gene duplication events, gene structure, protein secondary structure prediction, phylogenetic relationship with other the member of other taxa, abundance of cis-regulatory element, 3D structure modeling and their expression level during fruit development. The information of banana GSTs could be of potential importance for banana breeding program in the future.

2. Materials And Methods

2.1 Mining of Banana GSTs from Banana Genome Hub

The well characterized GST protein sequences of *Arabidopsis thaliana*, *Glycine max* and *Oryza sativa* were retrieved from The Arabidopsis Information Resource (TAIR) (http://www. arabidopsis.org/) and Rice Genome Annotation Project (RGAP) by their accession number respectively. With the pBLAST search in the genome of *M. acuminata* DH Pahang v4 (updated in September 2021) was employed on Banana Genome Hub (https://banana-genome-hub.southgreen.fr/download) database (Droc et al. 2013), with an e-value of 0.001. The identified sequences were subjected to NCBI Batch- CD Search (conserved domain) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer et al. 2017), SMART (Simple Modular Architecture Research Tool) database (http://smart.embl-heidelberg.de/) (Letunic et al. 2020) and Pfam search (http://pfam.xfam.org/search) online tool to find the key feature of GSTs i.e. the presence of conserved N-terminal domain with the thioredoxin fold and a C-terminal domain. The full length sequences containing both the conserved domain were selected for further analyses and characterization.

2.2 Chromosomal localization, evolutionary and gene duplication analyses of *MaGST* genes

The genomic locations of the identified MaGST genes were retrieved from genomic data. The locations of these genes were diagrammatically depicted on their respective chromosomes using the TBtools software v0.667 (https://github.com/CJ-Chen/TBtools). The phylogenetic analyses was carried out using the amino acid sequences of *M. acuminata*, *A. thaliana*, *O. sativa*, *G. max*, *P. patens* (a bryophyte) and *Larix kaempferi* (a gymnosperm) GST protein. The amino acid sequences of *P. patens* and *L. kaempferi* were downloaded from NCBI database using their accession number. The sequences were aligned using Clustal Omega and the tree was constructed following the neighbour-joining (NJ) method using MEGA X software (Kumar et al. 2018). For the accuracy of a constructed tree the bootstrap value was set at 1000 replicates. Gene duplication events were analyzed by pBLAST search of MaGSTs against each other on NCBI pBLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins). The MaGSTs exhibiting sequence similarity >80% were assumed as duplicated genes (Kong et al. 2013). The pair of homologous genes within 100 Kb regions on the same chromosome was considered as tandem duplicated (TD), while those located beyond 100 kb region or different chromosomal localization were designated as segmental duplicated genes (SD) (Holub et al. 2001). The estimation of synonymous rate (dS), non-synonymous rate (dN), and evolutionary constraint (dN/dS) between the duplicated MaGST gene pairs were analyzed using the PAL2NAL online tool (https://bio.tools/pal2nal) (Suyama et al. 2006) using their protein sequence alignment performed on Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) and their respective mRNA sequences. The mode of selection between duplicated genes was identified through dN/dS ratio. The value >1, =1 and <1 was considered as positive, neutral and purifying selection respectively. The divergence time T (million year (Mya) of each duplicated gene pair was calculated using the formula: (T=dS/2λ), where T is divergence time, dS is the number of synonymous substitutions per site, and λ is the fixed rate of 6.5×10^{-9} synonymous substitutions per site per year for monocotyledonous plants (Koch et al. 2000).

2.3 *In silico* physicochemical characterization of MaGST protein sequences and their subcellular localization prediction

The physicochemical parameters of MaGSTs such as molecular weight, pI, total number of negative and positive charged residues, extinction coefficient, aliphatic index and Grand Average of Hydropathy (GRAVY) were analyzed through ProtParam tool in Expasy (http://web.expasy.org/protparam/) (Gasteiger et al. 2005) with default parameters. The subcellular localization were predicted through three independent tools namely CELLO online tool v.2.5 (http://cello.life.nctu.edu.tw/) (Yu et al. 2006), DeepLoc (Armenteros et al. 2017) (http://www.cbs.dtu.dk/services/DeepLoc/) and WoLF pSORT (www.genscript.com/wolf-psort.html) (Horton et al. 2007).

2.4 Conserved motif and gene structure analyses

The amino acid sequences of Arabidopsis, rice and *P. somniferum* GSTs were used for conserved motif analysis. The conserved motifs of Ma GSTs were identified using the Mutiple Em for Motif Elicitation (MEME) program (http://meme-suite.org/) (Bailey et al. 2009). The parameters used for the analysis were 15 as the motif number and 6–50 as the motif width. The results were visualized with TBtools. The exon/intron organization of MaGSTs were analyzed by online available tool Gene Structure Display Server 2.0 (GSDS, https://gsds.cbi.pku.edu.cn/) (Hu et al. 2015) using their corresponding CDS sequences and genomic sequences, retrieved from Banana-Genome-Hub (https://banana-genome-hub.southgreen.fr/download).

2.5 Protein sequence alignments of MaGST protein and prediction of catalytic residue position

The MaGST protein sequences were aligned with the protein sequences of *A. thaliana*, *O. sativa* and *G. max* using Clustal Omega (Sievers et al. 2011). The protein alignments were than visualized through ESPript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) (Robert and Gouet 2014). The signature sequences and the conserved catalytic residues of MaGSTs were highlighted on the alignments.

2.6 Protein secondary structure prediction of MaGSTs

The secondary structure components viz. alpha helix, beta-strand and random coil of MaGSTs were predicted through SOPMA (Self-Optimized Prediction Method with Alignment), (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) (Combet et al. 2000).

2.7 Post- translational modifications observed in MaGSTs

The potential phosphorylation and glycosylation sites in the identified MaGSTs were predicted through local prediction software GPS5.0 (http://gps.biocuckoo.cn/online.php) (Wang et al. 2019) with the threshold of GPS5.0 as high and NetNGlyc 1.0 server (http:// www.cbs.dtu.dk/services/NetNGlyc/) (Gupta et al. 2004) respectively with default parameters.

2.8 Promoter analysis for cis-acting regulatory elements in *MaGST genes*

2 kbp promoter regions upstream of transcription start site (ATG) on MaGST genomic DNA sequences were extracted from JBrowse of *M. acuminata* DH Pahang v.4 using their locus ID to analyze cis-acting regulatory elements. The extracted promoter sequences were analyzed online through PlantCARE software (http://bioinformatics.psb.ugent.be/webtools/ plantcare/html/) (Lescot et al. 2002) to identify diverse hormone, stress and cellular development responsive elements.

2.9 Expression profiling of MaGSTs using RNA-seq data

To explore the basal gene expression patterns of the MaGST genes under fruit development condition, RNA-seq data of 45 genes of *M. acuminata* DH Pahang (AAA group) was retrieved from Expression Atlas (https://www.ebi.ac.uk/gxa/experiments). The GST gene expression data from endocarp tissue at 0days, 20 days and 80 days after flowering was analyzed to elucidate the importance of MaGST genes in fruit development. The heatmap was drawn by the TBtools software based on the transcripts per kilobase million (TPM) values of 45 MaGST genes
(Chen et al. 2018).

3. Results

3.1 M. acuminata genome has 50 GST genes that belong to seven canonical GST classes

With the pBLAST search of *M. acuminata* DH- Pahang v.4 genome at Banana Genome Hub, a total of 61 GST genes were obtained. The identified GST protein sequences were validated for the presence of conserved N and C-terminal domain through NCBI CD search, Pfam search and SMART database search. Out of 61 GSTs, 11 GST proteins that did not possess the conserved N or C-terminal domain were omitted from the study. 50 full length banana GST genes containing both the domains were named as MaGSTs belonging to seven established classes namely tau, phi, zeta, lambda, DHAR, EF1G and GHR. Elongation factor 1-gammma possessed an additional EF1G domain with the canonical N and C-terminal domain (Fig. 1). Plant specific tau and phi GSTs were highest in the number 24 and 11 respectively and theta GSTs were missing in the banana genome (Table 1). The nomenclature of banana GST protein was done as MaGSTs by taking the prefix Ma from *M. acuminata*, as proposed by Dixon et al. 2010 for *A. thaliana*. All the classes of MaGSTs were named MaGSTU, MaGSTF, MaGSTZ, MaGSTL, MaDHAR, MaGHR and MaEF1G. The numbering for each members of the class was done based on their chromosomal localization in the ascending order.

Table.1

List of identified GST members in Musa acuminata along with their detailed genomic information, physicochemical features and subcellular localization

S. No.	Locus ID	Gene Name	Chr. No.	Start	End	Strand	Gene (bp)	Protein (aa)	pI	Mol Wt (kDa)	GRAVY	AI	Sub Loc
1	Macma4_01_g17560.1	MaGSTU1	chr01	12645908	12646720	-	812	226	5.67	25.37	-0.104	95	Cy[a]
2	Macma4_01_g17570.1	MaGSTU2	chr01	12662270	12663054	-	784	217	5.1	24.43	-0.061	94.52	Cy[a]
3	Macma4_01_g17580.1	MaGSTU3	chr01	12673915	12674696	-	781	226	5.52	25.57	-0.087	90.27	Cy[a]
4	Macma4_01_g17600.1	MaGSTU4	chr01	12679873	12686709	-	6836	233	8.85	25.77	-0.152	89.7	Cy[a]
5	Macma4_01_g17620.1	MaGSTU5	chr01	12699355	12700135	-	780	226	5.66	25.55	-0.037	92.43	Cy[a]
6	Macma4_01_g17610.1	MaGSTU6	chr01	12711730	12712528	-	798	226	5.53	25.71	-0.124	88.14	Cy[a]
7	Macma4_01_g17590.1	MaGSTU7	chr01	12718722	12723573	-	4851	203	5.27	22.53	-0.08	91.82	Cy[a]
8	Macma4_03_g27660.1	MaGSTU8	chr03	38741904	38742965	+	1061	254	4.96	28.03	-0.106	93.31	Cy[a]
9	Macma4_04_g25520.1	MaGSTU9	chr04	33447586	33448535	+	949	227	5.26	25.45	0.02	104.45	Cy[a]
10	Macma4_05_g07780.1	MaGSTU10	chr05	5643248	5644184	-	936	224	5.38	25.75	-0.189	92.77	Cy[a]
11	Macma4_05_g10530.1	MaGSTU11	chr05	7490268	7491372	+	1104	232	5.4	25.64	0.003	99.22	Cy[a]
12	Macma4_05_g10520.1	MaGSTU12	chr05	7492167	7493157	+	990	239	5.92	26.86	0.115	108.54	Cy[a]
13	Macma4_05_g10540.1	MaGSTU13	chr05	7496933	7497902	+	969	230	5.59	25.54	0.088	100.96	Cy[a]
14	Macma4_05_g31120.1	MaGSTU14	chr05	45082780	45084014	+	1234	244	5.11	27.38	-0.139	93.11	Cy[a]
15	Macma4_08_g05550.1	MaGSTU15	chr08	33881616	33882956	-	1340	259	4.84	28.52	-0.082	91.47	Cy[a]
16	Macma4_09_g28170.1	MaGSTU16	chr09	43993848	43997243	+	3395	241	5.38	27.23	-0.177	90.21	Cy[a]
17	Macma4_09_g28180.1	MaGSTU17	chr09	44018622	44019703	+	1081	256	5.11	29.25	-0.2	85.35	Cy[a]
18	Macma4_09_g28210.1	MaGSTU18	chr09	44010622	44011745	+	1123	224	5.4	25.08	-0.11	91.43	Cy[a]
19	Macma4_09_g28220.1	MaGSTU19	chr09	44029677	44032258	+	2581	235	5.02	26.69	-0.134	96.64	Cy[a]
20	Macma4_09_g28230.1	MaGSTU20	chr09	44043122	44047786	+	4664	302	5.54	34.82	-0.236	94.3	Cy[a]
21	Macma4_09_g28240.1	MaGSTU21	chr09	44048790	44049518	+	728	214	5.28	24.24	-0.071	93.41	Cy[a]
22	Macma4_09_g28250.1	MaGSTU22	chr09	44051483	44052263	-	780	221	5.14	25.24	-0.203	82.94	Cy[a]
23	Macma4_10_g01170.1	MaGSTU23	chr10	1342950	1344013	+	1063	234	5.21	26.41	-0.026	100.85	Cy[a]
24	Macma4_10_g13200.1	MaGSTU24	chr10	26157275	26158656	+	1381	223	5.91	25.71	-0.371	87.89	Cy[a]
25	Macma4_01_g08790.1	MaGSTF1	chr01	6268539	6277061	-	8522	223	9.08	24.80	-0.27	81.03	Cy[a]
26	Macma4_02_g05750.1	MaGSTF2	chr02	20914202	20915401	-	1199	216	5.71	23.84	-0.087	87.59	Cy[a]
27	Macma4_02_g14640.1	MaGSTF3	chr02	26910761	26911782	-	1021	258	9.32	28.88	-0.267	80.89	Mt[a]
28	Macma4_03_g29570.1	MaGSTF4	chr03	39940354	39943419	+	3065	314	6.16	36.02	-0.404	85.06	Nu[a]
29	Macma4_04_g05170.1	MaGSTF5	chr04	4043451	4047857	+	4406	457	6.53	51.16	-0.23	87.26	Cy[a]
30	Macma4_04_g12310.1	MaGSTF6	chr04	8782518	8783723	-	1205	221	6.32	25.19	-0.243	92.58	Cy[a]
31	Macma4_04_g23890.1	MaGSTF7	chr04	32148837	32149900	-	1063	221	7.64	25.24	-0.38	90	Cy[a]
32	Macma4_04_g38330.1	MaGSTF8	chr04	42452829	42453742	+	913	221	8.38	25.57	-0.414	90.45	Cy[a]
33	Macma4_05_g16910.1	MaGSTF9	chr05	14241272	14242352	-	1080	221	6.43	24.52	-0.278	78.24	Cy[a]
34	Macma4_07_g04070.1	MaGSTF10	chr07	3080875	3083111	+	2236	549	6.54	58.66	-0.685	58.71	Nu[a]
35	Macma4_11_g19810.1	MaGSTF11	chr11	30559356	30560881	+	1525	215	5.44	23.57	0.095	102.56	Cy[a]

S. No.	Locus ID	Gene Name	Chr. No.	Start	End	Strand	Gene (bp)	Protein (aa)	pI	Mol Wt (kDa)	GRAVY	AI	Sub Loc
36	Macma4_06_g31870.1	MaGSTZ1	chr06	37035146	37040197	-	5051	225	5.83	25.16	-0.132	95.91	Cp[a] Nu[c]
37	Macma4_08_g30570.1	MaGSTZ2	chr08	47426590	47431263	-	4673	218	5.82	24.49	-0.267	95.87	Cy[a]
38	Macma4_08_g30590.1	MaGSTZ3	chr08	47436224	47440775	-	4551	218	5.1	24.36	-0.204	98.12	Cy[a]
39	Macma4_10_g20220.1	MaGSTZ4	chr10	30806066	30810276	+	4210	261	5.85	29.08	-0.102	90.46	Cp[a]
40	Macma4_06_g36420.1	MaGSTL1	chr06	40281165	40283472	+	2307	287	6.21	32.12	-0.22	83.38	Cp[a]
41	Macma4_08_g22450.1	MaGSTL2	chr08	41703039	41705386	-	2347	248	5.24	28.27	-0.371	88.1	Cy[a]
42	Macma4_09_g04240.1	MaGSTL3	chr09	2799886	2802861	-	2975	266	5.15	30.15	-0.382	80.04	Cp[a]
43	Macma4_10_g01880.1	MaGSTL4	chr10	2621522	2624676	+	3154	242	5.41	27.88	-0.524	77.02	Cy[a]
44	Macma4_06_g38000.1	MaDHAR1	chr06	41146917	41149619	-	2702	220	4.93	24.26	-0.117	90.36	Cy[a]
45	Macma4_11_g23820.1	MaDHAR2	chr11	33111424	33114740	-	3316	218	5.52	24.29	-0.211	98.3	Cy[a]
46	Macma4_02_g17210.1	MaEF1G1	chr02	28568385	28569603	-	1218	263	5	29.64	-0.307	81.29	Cy[a]
47	Macma4_09_g11850.1	MaEF1G2	chr09	8034508	8039435	-	4927	419	5.7	47.86	-0.333	83.99	Cy[a]
48	Macma4_09_g27020.1	MaEF1G3	chr09	42919531	42930570	-	11039	619	8.48	69.69	-0.345	81.42	Cp[a]
49	Macma4_09_g18510.1	MaGHR1	chr09	13698126	13701543	+	3417	425	8.37	46.44	-0.204	80.59	PM
50	Macma4_10_g19250.1	MaGHR2	chr10	30181763	30186696	-	4933	424	6.04	47.64	-0.381	70.64	Ex[a] Cy[c]

3.2 50 MaGST genes are clustered on 11 banana chromosomes and evolutionarily conserved. Tandem duplication was the driving force for MaGST gene family expansion

On the basis of the *M. acuminata* DH-Pahang v.4 annotation, 50 MaGST genes were assigned to the eleven chromosomes ranging from 1 (Chr7) to 11 (Chr9). Chromosome 11 contained the highest nine MaGST genes followed by Chr1 holding eight MaGST genes. The pattern of chromosomal allocation of GSTs was also noticeable i.e. mainly on proximal or distal end of the banana chromosome as depicted in Fig. 2. A phylogenetic tree was constructed using *M. acuminata*, Arabidopsis, rice, soybean (angiosperm), *P. patens* (a bryophyte) and *L. kaempferi* (a gymnosperm) GST protein sequences. The different GST classes branched out into their individual clades, with the members of each class clustering together. Two major clades are of plant specific tau and phi GSTs that made two superclades under which small clades were noticed (Fig. 3). The individual class of GST from different plants that belong to separate sub-groups of the plant kingdom were clustered together and an indicative of their divergent evolution from a common ancestor. Phylogenetic analysis of MaGSTs with angiosperm (Arabidopsis and rice), gymnosperm (*L. kaempferi*) and bryophyte (*P. patens*) was carried out. The outcome revealed that the evolution of plant GSTs might be earlier than their division into individual groups such as bryophyte, pteridophyte, gymnosperm and angiosperm and also the each GST classes have diverged prior to the division of monocot and dicot. Additionally, the numbers of each class of GSTs expanded in a species specific manner independently and irrespective of their genome size. Additionally, the gene pairs under tandem and segmental duplication were close together in a phylogenetic tree showing close relatedness with each other.

To elucidate the gene family expansion in MaGSTs the duplication mechanism were analyzed. A total of 21 duplication events were noticed in banana GST gene family expansion and evolution. Tandem duplication was found to play a major role as 16 gene pairs were involved in tandem duplication event creating 16 gene clusters on Chr 1, 5, 8 and 9. Among them 15 gene pairs of tau class genes were majorly duplicated. MaGSTZ2/ MaGSTZ3 gene pair was also tandem duplicated gene. Five gene pairs composed of MaGSTU8/ MaGSTU14, MaGSTU8/ MaGSTU15, MaGSTU14/ MaGSTU15, MaGSTF1/ MaGSTF9 and MaGSTL2/ MaGSTL4 genes were segmental duplicated. The duplication event majorly occurred on banana chr1, 5, 8, 9 and 10. Moreover, the dN/dS values of duplicated genes were calculated and found to be less than 1, which is an indicative of purifying selection. Lastly, the divergence time was also calculated for these duplicated genes. The estimated divergence time of these gene pairs were approximately 2.51~145.41 million year ago (MYA) (Table. 2).

Table. 2

Estimated dN/dS ratios and divergence times of the duplicated MaGST genes

S. No.	Gene Name 1	Chr No.	Gene Name 2	Chr No.	Percent Identity	d_N	d_S	d_N/d_S	Duplication time (Mya)	Duplication Type	Selection type
1	MaGSTU1	chr01	MaGSTU2	chr01	93.09%	0.0302	0.1293	0.2334	9.95	Tandem	Purifying
2	MaGSTU1	chr01	MaGSTU3	chr01	91.59%	0.0341	0.1636	0.2082	12.58	Tandem	Purifying
3	MaGSTU1	chr01	MaGSTU5	chr01	90.71%	0.0396	0.1267	0.3126	9.75	Tandem	Purifying
4	MaGSTU1	chr01	MaGSTU6	chr01	91.15%	0.0376	0.1035	0.3635	7.96	Tandem	Purifying
5	MaGSTU2	chr01	MaGSTU3	chr01	95.39%	0.0188	0.0738	0.2545	5.68	Tandem	Purifying
6	MaGSTU2	chr01	MaGSTU4	chr01	81.02%	0.1055	0.6032	0.175	46.40	Tandem	Purifying
7	MaGSTU2	chr01	MaGSTU5	chr01	93.09%	0.0303	0.065	0.4658	5.00	Tandem	Purifying
8	MaGSTU2	chr01	MaGSTU6	chr01	94.93%	0.0207	0.0418	0.494	3.22	Tandem	Purifying
9	MaGSTU3	chr01	MaGSTU5	chr01	94.69%	0.0231	0.0326	0.7076	2.51	Tandem	Purifying
10	MaGSTU3	chr01	MaGSTU6	chr01	95.13%	0.0197	0.0717	0.2751	5.52	Tandem	Purifying
11	MaGSTU4	chr01	MaGSTU7	chr01	89.32%	0.0591	0.1798	0.3287	13.83	Tandem	Purifying
12	MaGSTU5	chr01	MaGSTU6	chr01	92.92%	0.0306	0.042	0.7292	3.23	Tandem	Purifying
13	MaGSTU11	chr05	MaGSTU13	chr05	84.72%	0.0889	0.1082	0.8214	8.32	Tandem	Purifying
14	MaGSTU12	chr05	MaGSTU13	chr05	86.90%	0.069	0.0909	0.7584	6.99	Tandem	Purifying
15	MaGSTU17	chr09	MaGSTU20	chr09	82.73%	0.1048	1.2164	0.0861	93.57	Tandem	Purifying
16	MaGSTZ2	chr08	MaGSTZ3	chr08	95.41%	0.021	0.0377	0.5572	2.90	Tandem	Purifying
17	MaGSTU8	chr03	MaGSTU14	chr05	88.00%	0.0844	0.9435	0.0895	72.58	Segmental	Purifying
18	MaGSTU8	chr03	MaGSTU15	chr08	83.08%	0.0796	0.6984	0.1139	53.72	Segmental	Purifying
19	MaGSTU14	chr05	MaGSTU15	chr08	85.65%	0.0994	0.9109	0.1091	70.07	Segmental	Purifying
20	MaGSTF1	chr01	MaGSTF9	chr05	81.00%	0.1521	1.8903	0.0805	145.41	Segmental	Purifying
21	MaGSTL2	chr08	MaGSTL4	chr10	81.40%	0.0972	0.4392	0.2214	33.78	Segmental	Purifying

3.3 MaGST proteins are highly stable, hydrophilic and found majorly in the cytoplasm

The protein length of MaGSTs ranged from 203 (MaGSTU7) to 619 (MaEF1G3) with their corresponding molecular weight of 22.53 kDa to 69.69 kDa respectively. The isoelectric point (pI) ranged from 4.84 (MaGSTU15) to 9.32 (MaGSTF3). Among 50 MaGSTs, seven MaGSTs were basic and 43 MaGSTs were acidic in nature. The grand average of hydropathy values of most of the MaGST proteins of all the classes were negative indicating that all the MaGST proteins were hydrophilic having good interaction with water. The aliphatic index of MaGSTs ranged from 58.71 (MaGSTF10) to 108.54 (MaGSTU12). Most of the MaGSTs were having the AI value below 100 and hence these are hydrophilic in nature. The subcelluar localization prediction results showed that MaGSTs were centrally localized in the cytoplasm followed by chloroplast, nucleus, mitochondria, plasma membrane and extracellular (Table. 1, Fig. 4).

3.4 MaGST proteins are characterized by the presence of many class specific motifs and gene architecture among tau, phi, zeta, lambda and DHAR class is highly conserved

The conserved motif analyses identified many class specific motifs and few motifs were found to be distributed among all the GST classes. The tau class MaGSTs possessed the highest number of eight motifs i.e. motif 1, 2, 3, 4, 5, 7, 8 and 13 whereas MaDHAR possessed least number of motifs i.e. 4 and 6. Motif 4 was found to be distributed across all the MaGSTs except MaEF1G members likewise motif 1 was also present in all the MaGSTs omitting MaDHAR and MaGHR members. Motif 11 was found only in MaGSTF members. Motif 14 and 15 was found in MaGSTLs. Motif 12 was found only in MaGSTZ and MaEF1G class. Motif 10 was found in MaGSTF, MaGSTZ and MaEF1G. Motif 6 was found in all the MaGST classes except tau and lambda. Motif 3 was found in MaGSTU and MaGSTZ members (Fig. 5).

MaGST genes possessed two to ten exons. All the MaGSTU possessed two-exons/ one-intron except MaGSTU5/12/16/19/20. Except MaGSTF1/4/5, all the phi members contained three-exons/ two-introns. All the genes in the zeta and DHAR members possessed nine and six exons respectively. In the lambda class MaGSTL1/2/3 possessed nine exons whereas MaGSTL4 possessed eight exons. Two MaGHR genes possessed two and seven exons respectively whereas three MaEF1G genes contained five, six and nine exons correspondingly. The numbers of exons were highly variable in MaEF1G and MaGHR class (Fig. 6). The conservation in the number of exons can be correlated with the expansion of MaGST gene family.

3.5 Ser and Cys catalytic residues are highly conserved in MaGST classes

Multiple sequence alignment was performed by taking the GST protein sequences of *M. acuminata*, *A. thaliana*, *G. max* and *O. sativa* to identify the conserved residues and catalytic residue among different GST classes. The position of catalytic residues and their signature sequences have been depicted in Fig.. The position of active site residue varied among the classes. Tau and phi GSTs possessed Ser active site residue at the position 17 and 12 respectively whereas zeta hold active site Ser residue at position 41. Lambda and DHAR contained active site Cys residue at position 20. The GHR also possessed Cys active site residue at position 46.

3.6 MaGSTs are predominantly composed of α-helices

The secondary structure of plant GSTs is characterized by the dominance of alpha helix followed by coil, beta strand and beta turns. The percent of alpha helix was highest in all the MaGSTs especially tau proteins, except MaGSTF4/ MaGSTF10/ MaGSTZ4/ MaGSTL1 and MaDHAR2, in which the percent of coil was highest. The MaGSTU11 contained the 61.92 alpha helix which was highest among all the MaGSTs. Both the MaGHRs also possessed highest percent of coil than alpha helix (Fig. Table.S1).

3.7 Phosphorylation, as a major post translational modification in MaGSTs

Post translational modification plays an important role in protein structural modification and its functioning. In the post translational alteration prediction in banana GSTs, only 5 out of 50 MaGSTs possessed the glycosylation sites (Table S2). The phosphorylation prediction study revealed that Ser and Thr residues are highly phosphorylated accounting for 46.76% and 46% followed by Tyr that is 29.28% (Fig. 9; Table S3).

3.8 Five types of functional Cis-regulatory elements are present in the promoter region of MaGST genes

Cis-acting regulatory elements (CAREs) are found in the promoter region of target genes. They are basically short motif of 5-20 bp length, non-coding DNA that binds the transcription factors and regulate the gene transcription. In addition to four core elements (AT~TATA-box, CAAT-box, TATA-box, TATA), the current study identified 36 cis-elements categorized on the basis of its importance in plants physiology i.e. light responsive element, hormone responsive element, stress responsive element, cellular development related elements and other elements. The light responsive elements were highest in the number followed by stress and hormone responsive element (Fig. 10). The MaGSTU24 possessed the highest number of 57 cis-elements whereas MaGSTU6 possessed only 6 cis-elements in its promoter region. MYB and MYC cis-elements were highest in number 140 and 178 respectively. Abscisic acid responsiveness elements were than highest in number accounting for 127. The high number of stress responsive elements (MYB and MYC) in the promoter region can be associated with their role in combating against diverse stress in banana plant by upregulating the MaGST transcripts and total enzyme activity.

3.10 RNA expression profiling of MaGSTs showed most of the MaGST genes get up-regulated during fruit development stage

To predict the role of MaGSTs in fruit development, the expression level of 45 MaGST genes expressed in endocarp tissue were analyzed during 0 days, 20 days and 80 days after flowering (DAF) based on its RNA-seq data. Among the banana GSTs tau members the MaGSTU1 was highly up-regulated initially than down-regulated at 20 DAF and then up regulated at 80 DAF and could be involved in fruit ripening and at the phase of ripening the activity of the ethylene signal transduction pathway is also significantly increased (Dhar et al. 2019). Likewise the expression of MaGSTU9, MaGSTU11, and MaGSTU21 was also up-regulated at 80 DAF and they might be involved in fruit ripening. MaGSTU12, MaGSTU19 and MaGSTU20 were found to be down-regulated at 80 DAF and might not be involved in fruit development. Additionally, MaGSTU3, MaGSTU4, MaGSTU5, MaGSTU8, MaGSTU9, MaGSTU10, MaGSTU11, MaGSTU12, MaGSTU13, MaGSTU15, MaGSTU16, MaGSTU18 and MaGSTU22 genes were found to be up-regulated at 20 DAF. MaGSTF2 and MaGSTF11 were found to be down-regulated from 0 DAF to 80 DAF whereas MaGSTF3, MaGSTF4, MaGSTF6 and MaGSTF7 up-regulated at 20 DAF. MaGSTF5 was up-regulated at 80 DAF during fruit development. Gene expression level of MaGSTZ were found to be variable i.e. MaGSTZ2 was up-regulated from 0 to 20 DAF and than down-regulated at 80 DAF whereas MaGSTZ3 showed low level of expression at 0 to 80 DAF. The MaGSTZ4 showed increased expression at 0 DAF and then down-regulated from 20 DAF to 80 DAF. The MaGSTL1 showed a low expression level whereas MaGSTL3 and MaGSTL4 were highly expressed. All the MaGSTL, MaGHR genes and MaEF1G3 showed high level of expression at 20 DAF. The expression level of MaDHAR2 was high at 0 DAF and then low at 80 DAF. MaGSTU24, MaGSTF1, MaGSTF9, MaGSTF10 and MaGSTZ1 were down-regulated during the process of fruit development (Fig. 11, 12).

Discussion

GST gene family is well characterized is involved in diverse plant activities. It is functionally a versatile protein family involved in plant growth and development, signal transduction pathways, retrograde signaling, biotic and abiotic stress management, tetrapyrole signaling, hormone signaling etc. To date genome wide identification and characterization of GST gene family have been performed in variety of plants like hami melon (Song et al. 2021), melon (Wang et al. 2020), radish (Gao et al. 2020), medicago (Hasan et al. 2021) etc. The comprehensive genome wide identification of a gene family is precise and more significant using the whole genome information. In the current study, genome wide search of GST gene family in banana led to the identification of 50 GST genes in *M. acuminata* genome based on currently released genome v.4 of banana on Banana Genome Hub database. Like rice, barley, sweet potato, tomato, mungbean, soybean and various other crops, the numbers of tau and phi GST genes were the highest accounting for 24 and 11 respectively. The high

number of tau and phi GST genes reflecting their functional importance in plant growth and development. Tau and phi GSTs are also coupled with the plant response to various abiotic and biotic stresses. The pI values of GSTs were in the range of 4.84 to 9.32. Among 50 MaGST proteins 43 were acidic having pI value less than 7 and seven MaGSTs were basic in nature having pI value above 7. In a report by Mohanta et al. 2019, that plant proteomes from 145 species revealed a pI range of 1.99 (epsin) to 13.96 (hypothetical protein) and molecular mass of the plant proteins varied from 0.54 to 2236.8 kDa. The molecular weight and isoelectric point of a plant protein play a significant role in protein biochemical functioning and hence it is important to know these physicochemical features of a protein in detail. The grand average of hydropathy (GRAVY) of MaGSTs showed them highly stable protein as the values were highly negative indicating that all the MaGST proteins were hydrophilic (González-Faune et al. 2021). The higher aliphatic index is also related to the thermal stability of proteins due to the occurrence of aliphatic amino acids. These results indicate the higher thermostability of above mentioned MaGSTs (Hasan et al. 2021). The subcellular localization prediction of 50 MaGSTs showed the dominance of these proteins in cytoplasm which indicates that these are soluble proteins followed by chloroplast, nucleus and mitochondria. The subcellular localization of a protein is directly related to its involvement in biological process in a cell. Hence it is important to spot a protein to better understand their role at the cellular level (Glory et al. 2007). In a report, with the combination of C-terminal GFP fusion technique and confocal microscopy for visualizing the fusions in *Nicotiana benthamiana*, the subcellular localization of 16 *P. patens* GSTs out of 21 PpGST proteins were confirmed to be cytosolic and nuclear (Liu et al. 2013). Likewise, GSTs were also reported to be found in other cellular compartments such as mitochondria, chloroplast, and endoplasmic reticulum and also in plasma membrane (Lallement et a. 2014a). The MaGSTs found in mitochondria might be involved in maintaining GSH: GSSG ratios in mitochondria. This is one of the important functions which GSTs play to maintain GSH: GSSG ratios in mitochondria because a high concentration of glutathione have been observed in this cellular compartment (Zechmann et al. 2008).

Gene family expansion majorly depends upon gene duplication events, transposition or splicing. Segmental and tandem duplication are two major duplication types. In the current study, tandem duplication played a driving force for MaGST gene family expansion and among them MaGST tau genes was majorly duplicated and this class contributed more for MaGST gene family expansion. This might be due to high number of tau genes and their major role in detoxifying xenobiotics and providing stress tolerance to plants against diverse stress (He et al. 2016). It can also be inferred that tau GSTs of banana were earliest than other MaGST gene members. 28 gene pairs in melon, 9 gene pairs in hami melon and 11 gene pairs in apple were also found to have tandem duplication pattern. In a report by Flagel et al. 2009, the polyploidy can be a leading provider of duplicate genes by means of diverse duplication mechanisms and a source of evolutionary uniqueness in plant genome. The ratio of non-synonymous and synonymous substitutions (Ka/Ks) for duplicated MaGST genes was found to be less than 1 signifying that the duplicated genes were under purifying selection pressure which indicates the removal of deleterious duplications and increases the fixation possibility of novel duplicated genes (Tanaka et al. 2009). Interestingly, entire duplicated MaGST genes were primarily under strong purifying selection like Gossypium species (Dong et al. 2016) and tomato GST genes (Islam et al. 2017). Phylogenetic analysis showed the numbers of each class of GSTs expanded in a species specific manner independently and irrespective of their genome size (Hasan et al. 2021).

The gene structure of GST gene family is highly conserved among different plant species. There was conservation in exon numbers among common phylogenetic classes and different exons numbers can be correlated with different evolutionary patterns (Wang et al. 2019). There are 2 exons in tau class, 3 in phi class, 9 in zeta class, 6 in DHAR and 8-9 in lambda class. All the MaGST genes of tau, phi, zeta, lambda and DHAR classes contained the prominent number of exons exhibiting their evolutionary conservation. The same numbers were also reported in wheat, raddish, mung bean, apple and melon GSTs. In a report it is mentioned that the less number of introns respond fast to stress stimuli hence tau and phi containing less number of introns respond quickly to any stress (Jeffares et al. 2008).

MEME analyses identified class specific motifs and motifs that were found in many MaGST classes. The diversity in the occurrence of motifs in individual classes were probably due to GSTs common functions like plant growth and development, stress management as well as few different role in tyrosine metabolism and ascorbate metabolism etc. (Vaish et al. 2020). A well conserved signature motifs were identified in MaGSTs i.e. W(A/V)S(P/M) in tau, (E/Q)SR(A/K/G)I in phi, SSCS/A in zeta, CPF/YA in lambda, CPFC/S in DHAR and CPWA in GHR (Vaish et al. 2020) were also reported in tomato (Islam et al. 2017), capsicum (Islam et al. 2019), chinese cabbage (Du et al. 2018) and wheat (Wang et al. 2019). The presence of these signature motifs in the individual classes of MaGSTs clearly validates them as GST proteins and their involvement in diverse plant functions.

The promoter analyses revealed different CAREs that were related to hormonal, cellular, stress and light response functions (Kaur et al. 2017). In addition to these elements core promoters (AT~TATA-box, CAAT-box, TATA-box, TATA) (Rahman et al. 2021) were also found. The light responsive elements were majorly found in the promoter region of all the 50 MaGST genes. Photosynthetic reaction takes place in the leaves in response to light hence this can also be correlated with association of GST protein in photosynthesis. In a report by Gallé et al. 2019, that function and expression of GSTs and the level of its substrate i.e. GSH also depends on the quality and intensity of light. Due to significant role of light in GSTs activity and its expression, probably the numbers of light responsive elements were higher. Plant growth regulators play an important role in plant growth and development, seed germination, fruit development etc. Many regulatory elements such as ERE motif, GARE motif, CGTCA element, TGACG element, ABRE, TGA element, Aux RR-core, O2-site, P box etc. were found in the promoter region of most of the MaGSTs that can be responsive to diverse hormones like ethylene, gibberellins, salicylic acid, methyl jasmonate, abscisic acid and auxin. These all plant hormones play an important role in a range of plant metabolisms. Presence of different defense and stress responsive elements like STRE, DRE, LTR, MBS, W- box, WRE3, WUN motif, as-1, TC-rich repeats confirmed the role of MaGSTs providing resistance against many biotic and abiotic stresses (Kaur et al. 2017). Additionally, for circadian control, circadian element was also found in many MaGST genes. Although, very less information is available regarding the role of this element in plant metabolism, the study by Alderete et al. (2018) on tobacco seedlings, the NtGST gene (phi) was analyzed for the putative circadian regulation. The results of the study exhibited diurnal regulation with increased expression at the end of the light phase, with transcript levels decreasing in the dark period. Collectively, cis- acting regulatory elements are essential for driving the functioning of a protein efficiently.

The expression profiling of GST genes have been done in many plant species under various developmental stages in different plant parts like leaves, root, pericarp, endocarp, seeds and flowers etc (Dixon et al. 2010; Jain et al. 2010; Islam et al. 2017). In the current study the MaGSTs were found to be involved in fruit development during all the stages.

Conclusively, *in silico* identification and characterization of GST gene family in banana led to the identification of 50 full length GST genes. The identified MaGST genes can potentially be used for molecular and functional characterization in this agriculturally important crop. The individual MaGST genes can be cloned and characterized, and their expression can be studied in different tissues under normal developmental and diverse stress conditions. The results of the current study can be utilized for opium plant breeding programs for developing high yielding and/or stress tolerant varieties.

Declarations

Funding:

No funding was received for conducting this study.

Competing Interests:

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdul Kayum M, Nath UK., Park JI, Biswas MK, Choi EK, Song JY, Kim HT, Nou IS (2018) Genome-Wide Identification, Characterization, and Expression Profiling of Glutathione *S*-Transferase (GST) Family in Pumpkin Reveals Likely Role in Cold-Stress Tolerance. *Genes 9*:84. https://doi.org/10.3390/genes9020084
2. Ahmad MZ, Nasir JA, Ahmed S, *et al.* (2020) Genome-wide analysis of glutathione S-transferase gene family in *G. max*. Biologia 75:1691–1705. https://doi.org/10.2478/s11756-020-00463-5
3. Ahmed FF, Hossen M.I, Sarkar M.AR, Konak JN, Zohra FT, Shoyeb M., et al. (2021) Genome-wide identification of DCL, AGO and RDR gene families and their associated functional regulatory elements analyses in banana (Musa acuminata). PLoS ONE 16(9): e0256873. https://doi.org/10.1371/journal.pone.0256873
4. Alderete LGS, Guido ME, Agostini E, and Mas P (2018) Identification and characterization of key circadian clock genes of tobacco hairy roots: putative regulatory role in xenobiotic metabolism. Environ. Sci. Pollut. Res. Int. 25:1597–1608. doi: 10.1007/s11356-017-0579-9
5. Armenteros JJA, Sønderby CK, Sønderby SK, Nielsen H, Winther O (2017) DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* 21:3387-339. https://doi.org/10.1093/bioinformatics/btx431
6. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L et al. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research 37:202-208. https://doi.org/10.1093/nar/gkp335
7. Basantani M, Srivastava A (2007) Plant glutathione transferases: a decade falls short. Can J Bot 85:443-456. https://doi.org/10.1139/B07-033
8. Chen CJ, Chen H, He YH, Xia R (2018) TBtools, a toolkit for biologists integrating various biological data handling tools with a user-friendly interface. BioRxiv. https://doi.org/10.1101/289660
9. Combet C., Blanchet C., Geourjon C. and Deléage G (2000) NPS@: Network Protein Sequence Analysis. TIBS 25:147-150. 10.1016/s0968-0004(99)01540-6
10. D'Hont A, Denoeud F, Aury JM., Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, et al. (2012) The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. Nature, 10.1038/nature11241.
11. Droc G, Lariviere D, Guignon V, Yahiaoui N, This D, Garsmeur O, Dereeper A, Hamelin C, Argout X, Dufayard JF, Lengelle J, Baurens FC, Cenci A, Pitollat B, D'Hont A, Ruiz M, Rouard M, Bocs S*(2013)* **The Banana Genome Hub**. *Database* doi:10.1093/database/bat035
12. Ding N, Wang A, Zhang X, Wu Y, Wang R, Cui H, Huang R, Luo Y (2017) Identification and analysis of glutathione S-transferase gene family in sweet potato reveal divergent GST-mediated networks in aboveground and underground tissues in response to abiotic stresses. BMC Plant Biol 17:225. https://doi.org/10.1186/s12870-017-1179-z
13. Du J, Ren J, Ye X, Hou A, Fu W, Mei F, Liu Z (2018) Genome-wide identification and expression analysis of the glutathione S-transferase (GST) family under different developmental tissues and abiotic stresses in Chinese cabbage (Brassica rapa ssp. pekinensis) PeerJ Preprints 6:e26629v1/. https://doi.org/10.7287/peerj.preprints.26629v1
14. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. (2013) STRING v9. 1: protein protein interaction networks, with increased coverage and integration. Nucleic acids research. 41:D808–D15. 10.1093/nar/gks1094
15. Gallé Á, Czékus Z, Bela K, Horváth E, Ördög A, Csiszár J, Poór P (2019) Plant glutathione transferases and light. Front Plant Sci 9:1944. https://doi.org/10.3389/fpls.2018.01944
16. **Gao J, Chen B, Lin H, Liu Y, Wei Y, Chen F, Li W (2020) Identification and Characterization of the Glutathione S-Transferase (GST) Family in Radish Reveals a Likely Role in Anthocyanin Biosynthesis and Heavy Metal Stress Tolerance. Gene** 743:144484. https://doi.org/10.1016/j.gene. 2020.144484.
17. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins M, Appel RD et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook 571-607. https://doi.org/10.1385/1-59259-890-0:571
18. Ghangal R, Rajkumar MS, Garg R. *et al.* (2020) Genome-wide analysis of glutathione S-transferase gene family in chickpea suggests its role during seed development and abiotic stress. Mol Biol Rep 47:2749–2761. https://doi.org/10.1007/s11033-020-05377-8

19. Glory E, Murphy RF (2007) Automated subcellular location determination and high-throughput microscopy. Dev Cell. 12(1): 7–16. https://doi.org/10.1016/j.devcel.2006.12.007 PMID: 17199037.
20. González-Faune P, Sánchez-Arévalo I, Sarkar S, Majhi K, Bandopadhyay R, Cabrera-Barjas G, Gómez A, Banerjee A (2021) Computational Study on Temperature Driven Structure–Function Relationship of Polysaccharide Producing Bacterial Glycosyl Transferase Enzyme. Polymers 13:1771. https://doi.org/10.3390/ polym13111771
21. Gupta R, Jung E, Brunak S (2004) Prediction of N-glycosylation sites in human proteins. 46. 203-206.
22. Hasan MS, Singh V, Islam S, Islam MS, Ahsan R, et al. (2021) Genome-wide identification and expression profiling of glutathione S-transferase family under multiple abiotic and biotic stresses in *Medicago truncatula* L.. PLOS ONE 16:e0247170. https://doi.org/10.1371/journal.pone.0247170
23. (https://economictimes.indiatimes.com/news/economy/foreign-trade/india-exports-bananas-worth-rs-619-cr-during-2020 2021/articleshow/83566756.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst).
24. Hayes JD, Flanagan JU and Jowsey IR (2005) Glutathione transferases. Annu. Rev. Pharmacol. Toxicol. 45:51-88. 10.1146/annurev.pharmtox.45.120403.095857
25. Holub EB (2001) The arms race is ancient history in Arabidopsis, the wildflower. Nat. Rev. Genet. 2:516–527. 10.1038/35080508
26. Horton P, Park K, Obayashi T, Fujita N, Harada H, Adams-Collier, CJ et al. (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Research, 35 (Web Server Issue): W585-W587. https://doi.org/10.1093/nar/gkm259
27. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31:1296-1297. https://doi.org/10.1093/bioinformatics/btu817
28. Hu W, Hou X, Huang C, Yan Y, Tie W, Ding Z, Wei Y, Liu J, Miao H, Lu Z, Li M, Xu B, Jin Z (2015) Genome-Wide Identification and Expression Analyses of Aquaporin Gene Family during Development and Abiotic Stress in Banana. Int J Mol Sci. 20;16(8):19728-51. doi: 10.3390/ijms160819728. PMID: 26307965; PMCID: PMC4581322.
29. Islam S, Rahman IA, Islam T, Ghosh A (2017) Genome-wide identification and expression analysis of glutathione S-transferase gene family in tomato: Gaining an insight to their physiological and stress-specific roles. PLOS ONE 12: e0187504. https://doi.org/10.1371/journal.pone.0187504
30. Islam S, Sajib SD, Jui ZS, Arabia S, Islam T (2019) Genome-wide identification of glutathione S-transferase gene family in pepper, its classification, and expression profiling under different anatomical and environmental conditions. Scientific Reports 9:9101. https://doi.org/10.1038/s41598-019-45320-x.
31. Jain M, Ghanashyam C, Bhattacharjee A (2010) Comprehensive expression analysis suggests overlapping and specific roles of glutathione S-transferases during development and stress responses in rice. BMC Genomics 11:73. https://doi.org/10.1186/1471-2164-11-73
32. Jeffares DC, Penkett CJ, Baehler J (2008) Rapidly regulated genes are intron poor. Trends Genet. 24, 375–378. doi: 10.1016/j.tig.2008.05.006
33. Kaur A, Pati PK, Pati AM, Nagpal AK (2017) *In-silico* analysis of cis-acting regulatory elements of pathogenesis-related proteins of *Arabidopsis thaliana* and *Oryza sativa*. PLoS ONE 12:e0184523. https://doi.org/10.1371/journal.pone.0184523
34. Kayum MA, Nath UK, Park JI, Biswas MK, Choi EK, Song JY, Kim HT, Nou IS (2018) Genome-wide identification, characterization, and expression profiling of glutathione S-transferase (GST) family in pumpkin reveals likely role in coldstress tolerance. Genes. 9(2):84. https://doi.org/10.3390/genes9020084
35. Klausen MS, Jespersen MC, Nielsen H, Jensen KK, Jurtz VI, Sønderby CK, Sommer MOA, Winther O, Nielsen M, Petersen B Marcatili P (2019) NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning. Proteins 87:520-527. https://doi.org/10.1002/prot.25674
36. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae). Mol. Biol. 17:1483–1498. 10.1093/oxfordjournals.molbev.a026248
37. Kong H, Landherr LL, Frohlich MW, Leebens-Mack J, Ma H, DePamphilis CW (2007) Patterns of gene duplication in the plant *SKP1* gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. The Plant Journal. 50: 873-885. https://doi.org/10.1111/j.1365-313X.2007.03097.x
38. Kong X, Lv W, Jiang S, Zhang D, Cai G, Pan J, Li D (2013) Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. BMC Genom.14:433. https://doi.org/10.1186/1471-2164-14-433
39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549. https://doi.org/10.1093/molbev/msy096
40. Kumar N, Ved A, Yadav RR, Prakash O (2021) A Comprehensive Review on Phytochemical, Nutritional, and Therapeutic Importance of *Musa acuminate*. Int. J. of Current Research and Review. http://dx.doi.org/10.31782/IJCRR.2021.13901
41. **Labrou NE, Papageorgiou AC, Pavli O, Flemetakis E (2015) Plant GSTome: structure and functional role in xenome network and plant stress response. Current Op in Biotechnol 32:186–194. https://doi.org/10.1016/j.copbio.2014.12.024**
42. Lallement PA, Brouwer B, Keech O, Hecker A, Rouhier N (2014a) The still mysterious roles of cysteine-containing glutathione transferases in plants. Front Pharmacol 5:192. https://doi.org/10.3389/fphar.2014.00192
43. Lallement PA, Meux E, Gualberto JM, Dumarcay S, Favier F, Didierjean C et al. (2014b). Glutathionyl-hydroquinone reductases from poplar are plastidial proteins that deglutathionylate both reduced and oxidized glutathionylated quinones. FEBS Lett 589:37–44. doi: 10.1016/j.febslet.2014.11.021
44. Lescot M, De´hais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic acids research 30:325–7. 10.1093/nar/30.1.325
45. Letunic I, Bork P (2018) 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46(D1):D493–6. 10.1093/nar/gkaa937
46. **Liu YJ, Han XM, Ren LL, Yang HL, Zeng QY (2013) Functional divergence of the glutathione S-transferase supergene family in Physcomitrella patens reveals complex patterns of large gene family evolution in land plants. Plant Physiol 161:773-786.** https://doi.org/10.1104/pp.112.205815

47. Marchler Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S et al. (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research, 45:D200-D203. https://doi.org/10.1093/nar/gkw1129
48. **Marques AC, Vinckenbosch N, Brawand D, Kaessmann H (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. Genome biol 9:R54. https://doi.org/10.1186/gb-2008-9-3-r54**
49. Mano J, Kanameda S, Kuramitsu R, Matsuura N, Yamauchi Y (2019) Detoxification of Reactive Carbonyl Species by Glutathione Transferase Tau Isozymes. Front. Plant Sci 10:487. 10.3389/fpls.2019.00487
50. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer E, Tosatto S, Paladin L, Raj S, Richardson LJ, Finn RD and Bateman A (2021) Pfam: The protein families database in 2021. Nucleic acids research 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
51. Mohanta, TK, Khan A, Hashem A, *et al.* (2019) The molecular mass and isoelectric point of plant proteomes. BMC Genomics 20:631. https://doi.org/10.1186/s12864-019-5983-8
52. Nianiou-Obeidat I, Madesis P, Kissoudis C, Voulgari G, Chronopoulou E, Tsaftaris A, Labrou NE (2017) Plant glutathione transferase-mediated stress tolerance: functions and biotechnological applications. Plant Cell Rep 36:791-805. https://doi.org/10.1007/s00299-017-2139-7
53. Rahman MM, Rahman MM, Eom JS *et al.* (2021) Genome-wide Identification, Expression Profiling and Promoter Analysis of Trehalose-6-Phosphate Phosphatase Gene Family in Rice. J. Plant Biol. 64:55–71. https://doi.org/10.1007/s12374-020-09279-x
54. Rezaei MK, Shobbar ZS, Shahbazi M, Abedini R, Zare S (2013) Glutathione S-transferase (GST) family in barley: Identification of members, enzyme activity, and gene expression pattern. J Plant Physiol 170:1277-1284. https://doi.org/10.1016/j.jplph.2013.04.005
55. Rice Genome Annotation Project. Available online: http://rice.plantbiology.msu.edu (accessed on 22 November 2017).
56. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Research 42:W320-W324. https://doi.org/10.1093/nar/gku316
57. Separovich RJ, Wong MWM, Chapman TR, Slavich E, Hamey JJ, Wilkins MR (2021) Post-translational modification analysis of Saccharomyces cerevisiae histone methylation enzymes reveals phosphorylation sites of regulatory potential. J Biol Chem. 296:100192. doi: 10.1074/jbc.RA120.015995.
58. Shao D, Li Y, Zhu Q, Zhang X, Liu F, Xue F, Sun J (2021) GhGSTF12, a glutathione S-transferase gene, is essential for anthocyanin accumulation in cotton (Gossypium hirsutum L.), Plant Science. 305:110827 https://doi.org/10.1016/j.plantsci.2021.110827
59. Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539. https://doi.org/10.1038/msb.2011.75.
60. Song W, Zhou F, Shan C, Zhang Q, Ning M, Liu X, Zhao X, Cai W, Yang X, Hao G and Tang F (2021) Identification of Glutathione S-Transferase Genes in Hami Melon (Cucumis melo var. saccharinus) and Their Expression Analysis Under Cold Stress. Front. Plant Sci. 12:672017. doi: 10.3389/fpls.2021.672017
61. Suyama M, Torrents D.and Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609-W612. 10.1093/nar/gkl315
62. Sylvestre-Gonon E, Law SR, Schwartz M, Robe K, Keech O, Didierjean C, Dubos C, Rouhier N, Hecker A (2019) Functional, Structural and Biochemical Features of Plant Serinyl-Glutathione Transferases. Front Plant Sci 10:608. https://doi.org/10.3389/fpls.2019.00608
63. Sylvestre-Gonon E, Schwartz M, Girardet JM, Hecker A, Rouhier N (2020) Is there a role for tau glutathione transferases in tetrapyrrole metabolism and retrograde signalling in plants? Phil. Trans R Soc B 375:20190404. https://doi.org/10.1098/rstb.2019.0404
64. Tanaka K, Takahasi K, Takano-Shimizu T (2009) Enhanced fixation and preservation of a newly arisen duplicate gene by masking deleterioud loss-of-function mutations. Genetics Research. 91(4):267-280. 10.1017/S0016672309000196
65. The Arabidopsis Information Resource (TAIR). Available online: https://www.arabidopsis.org (accessed on 18 November 2017).
66. Thom R, Dixon DP, Edwards R, Cole DJ and Lapthorn AJ (2001) The structure of a zeta class glutathione S-transferase from Arabidopsis thaliana: characterisation of a GST with novel active-site architecture and a putative role in tyrosine catabolism. J. Mol. Biol. 308:949–962. doi: 10.1006/jmbi.2001.4638
67. Vaish S, Awasthi P, Tiwari S, Tiwari SK, Basantani MK (2018) In silico genome-wide identification and characterization of glutathione S-transferase gene family in Vigna radiata (L.) Wilczek. Genome 61:311-322. https://doi.org/10.1139/gen-2017-0192
68. Vaish S, Gupta D, Basantani MK (2020) Glutathione S-transferase: a versatile protein family. 3Biotech 10:321 https://doi.org/10.1007/s13205-020-02312-3
69. Wagner U, Edwards R, Dixon DP, Mauch F (2002) Probing the diversity of the *Arabidopsis* glutathione S- transferase gene family. Plant Mol. Biol 49:515-532. https://doi.org/10.1023/A:1015557300450
70. Wang C, Xu H, Lin S, Deng W, Zhou J, Zhang Y, Shi Y, Peng Di, Xue Y (2019) **GPS 5.0: An update on the prediction of kinase-specific phosphorylation sites in proteins**. Genomics, Proteomics & Bioinformatics 18:72-80. https://doi.org/10.1016/j.gpb.2020.01.001
71. Wang R, Ma J, Zhang Q. et al. (2019) Genome-wide identification and expression profiling of glutathione transferase gene family under multiple stresses and hormone treatments in wheat (*Triticum aestivum* L.). BMC Genomics 20:986. https://doi.org/10.1186/s12864-019-6374-x.
72. Wang J, Wang Z, Jia C, *et al.* (2021) Genome-Wide Identification and Transcript Analysis of TCP Gene Family in Banana (*Musa acuminata* L.). *Biochem Genet*. https://doi.org/10.1007/s10528-021-10100-8
73. Wongsantichon J, Ketterman AJ (2005) Alternative splicing of glutathione S-transferases. Methods Enzymol 401:100-116. https://doi.org/10.1016/S0076-6879(05)01006-2
74. Xiong Y, Li R, Lin X, Zhou Y, Tang F, Yao Y, Liu J, Wang L, Yin X, Liu Y, et al. (2021)Genome-Wide Analysis and Expression Tendency of Banana (Musa acuminata L.) Calcineurin B-Like (MaCBL) Genes under Potassium Stress. Horticulturae. 7:70. https://doi.org/10.3390/ horticulturae7040070

75. **Yang Q, Liu YJ, Zeng QY (2014) Biochemical functions of the glutathione transferase supergene family of Larix kaempferi. Plant Physiol and Biochem 7:99-107.** https://doi.org/10.1016/j.plaphy.2014.02.003
76. Yuan W, Liu J, Taká ˇc T, Chen H, Li X, Meng Jm Tan Y, Ning T, He Z, Yi G, et al. (2021)Genome-Wide Identification of Banana Csl Gene Family and Their Different Responses to Low Temperature between Chilling-Sensitive and Tolerant Cultivars. Plants. 10:122. https://doi.org/10.3390/ plants10010122
77. Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins 64:643-651. http://dx.doi.org/10.1002/prot.21018
78. Yu H, Yan Y, Zhang C, Dalby PA (2017) Two strategies to engineer flexible loops for improved enzyme thermostability. Sci. Rep 7:1–15. https://doi.org/10.1038/srep41212
79. **Zechmann B, Mauch F, Sticher L, Müller M (2008) Subcellular immunocytochemical analysis detects the highest concentrations of glutathione in mitochondria and not in plastids. J. of Exp Botany 59: 4017-4027.** https://doi.org/10.1093/jxb/ern243

Figures

Figure 1

The conserved N and C-terminal domain of MaGSTs confirmed through NCBI CD- search

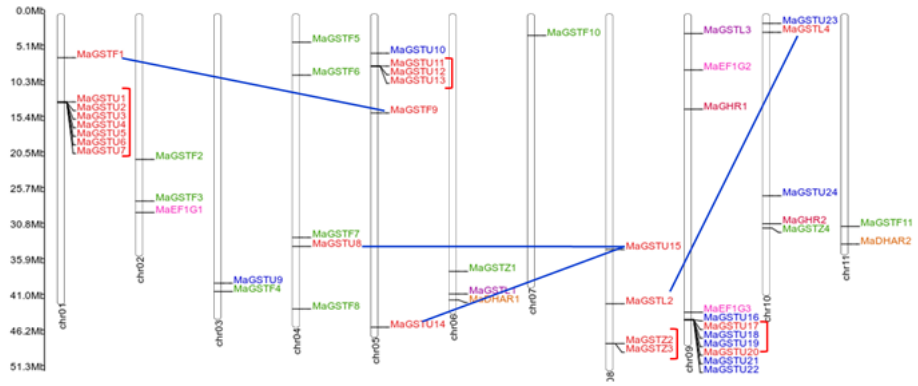

Figure 2

Chromosomal distribution of 50 MaGST genes on 11 *M. acuminata* chromosomes. Different colors are representative of different MaGST classes. The members in red colored font are duplicated gene pairs. Gene pairs showing tandem and segmental duplication are represented through red color bracket and blue lines respectively.

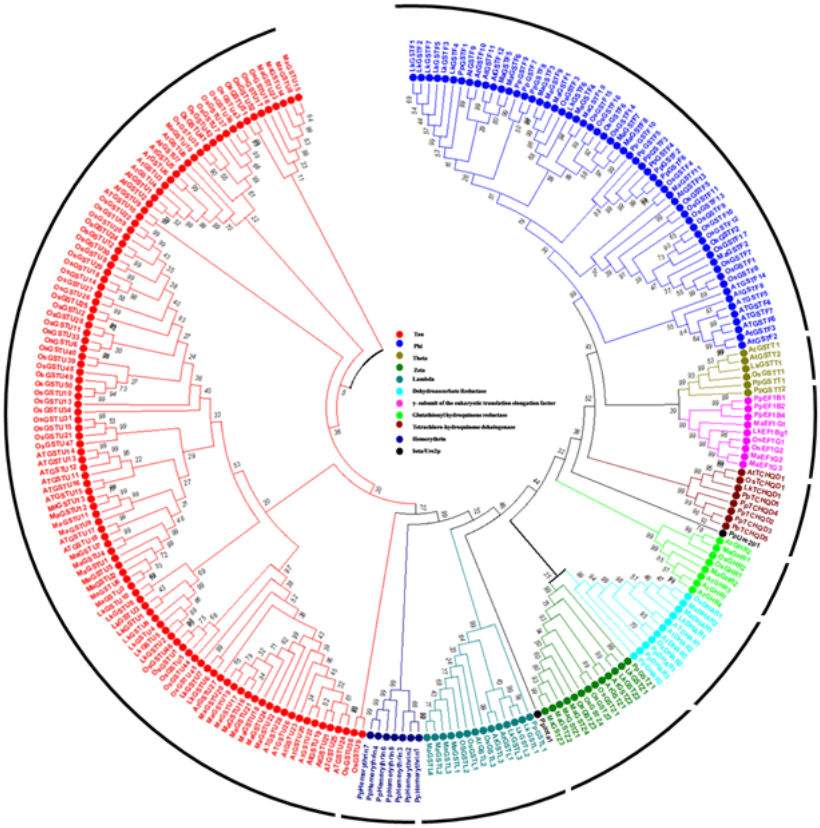

Figure 3

Phylogenetic tree of GST proteins among banana, rice, Arabidopsis, *P. patens* and *L. kaempferi*. A total of 10 different clades were depicted in different colors. Plant specific tau and phi classes made the largest clade.

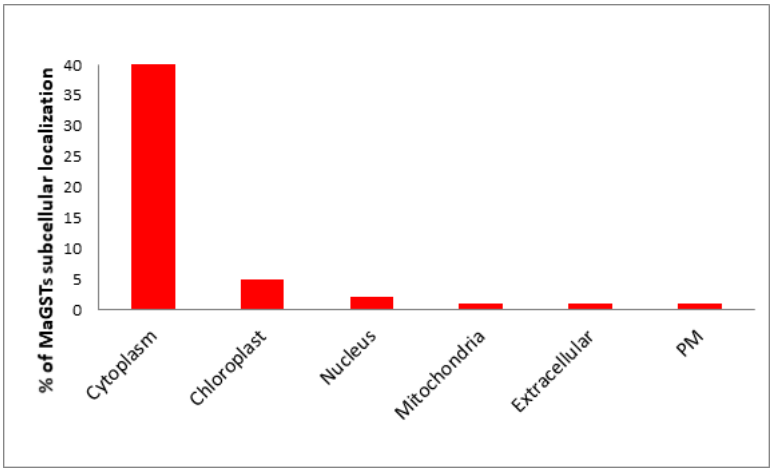

Figure 4

MaGSTs are majorly localized in the cytoplasm

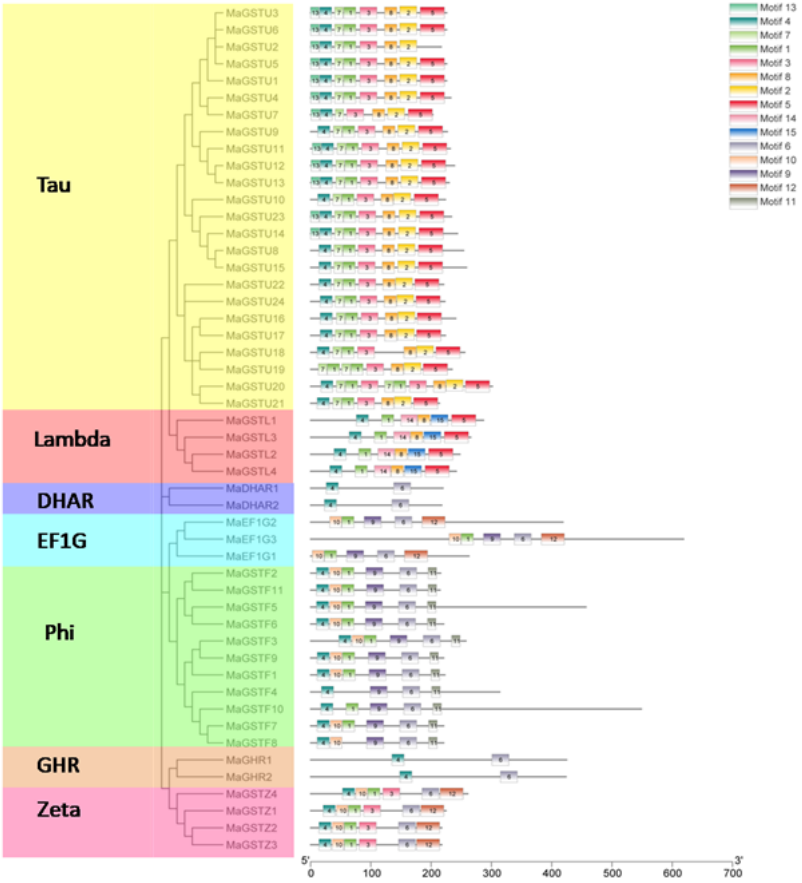

Figure 5

Conserved motif analyses of banana GSTs with 15 motifs. All the conserved motifs were identified by the MEME suite with complete MaGST protein sequences. Different motifs are represented by different colored boxes on their corresponding domain.

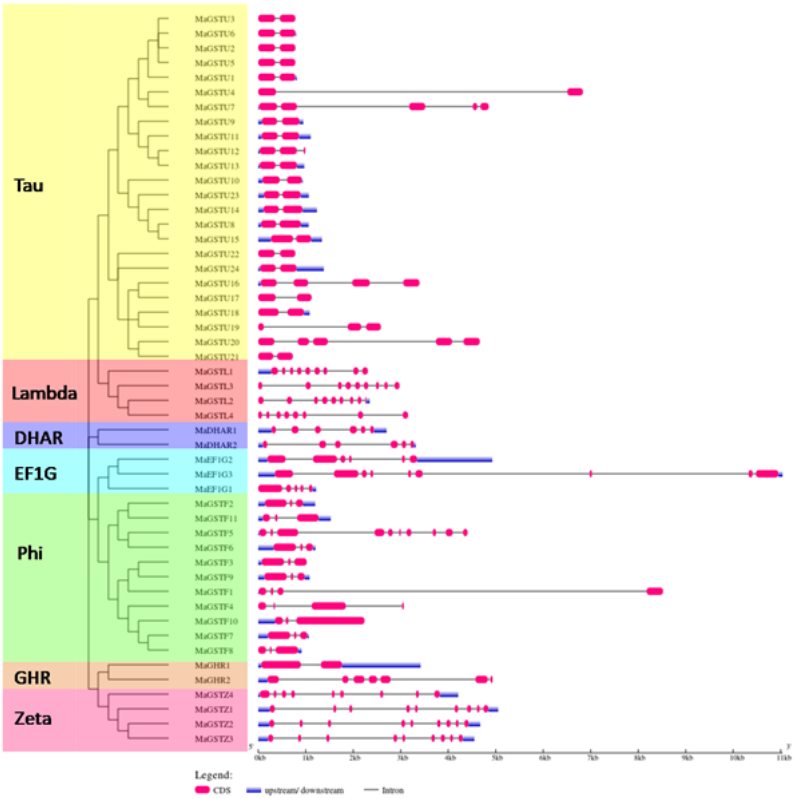

Figure 6

Gene structure analyses of banana MaGSTs. An exon-intron structure analysis was done using the GSDS tool.

Figure 7

Catalytic residue depiction in banana GST proteins. Multiple sequence alignments of MaGSTs were performed with *A. thaliana*, rice and soybean GST protein sequences. The asterisks (*) indicate the active site serine in tau, phi and zeta MaGSTs (a- c) and the active site cysteine in DHAR, lambda and GHR (d- f).

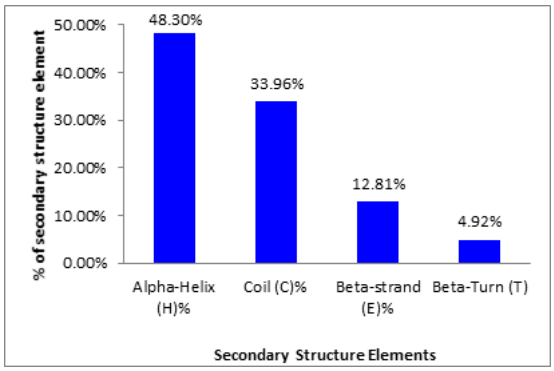

Figure 8

Secondary structure of MaGSTs were characterized by the dominance of alpha helices. The analyses was performed by SOPMA tool.

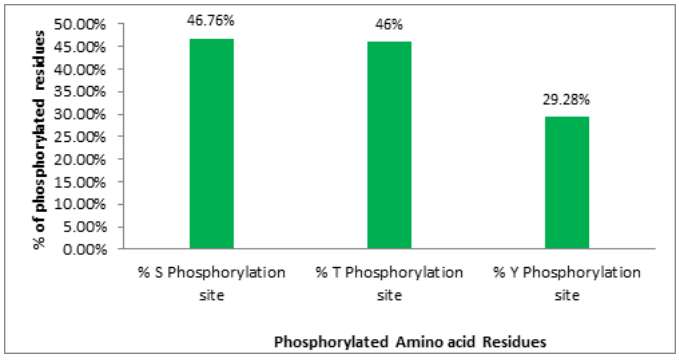

Figure 9

Potential phosphorylation prediction revealed serine as highly phosphorylated amino acid residue

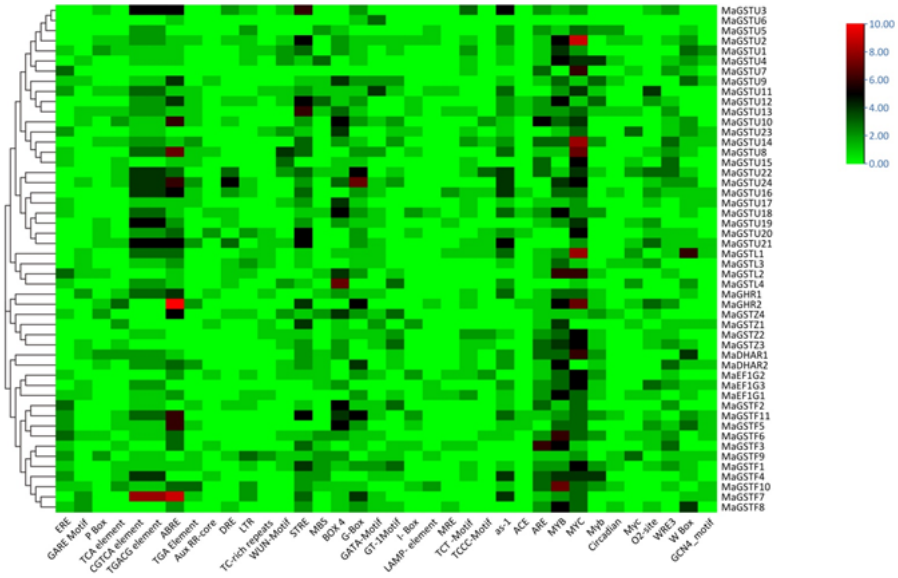

Figure 10

Cis-acting regulatory elements predicted in the upstream promoter region of predicted MaGSTs. The scale represents the number of particular elements in the corresponding genes. Grey color is indicative of absence of CAREs.

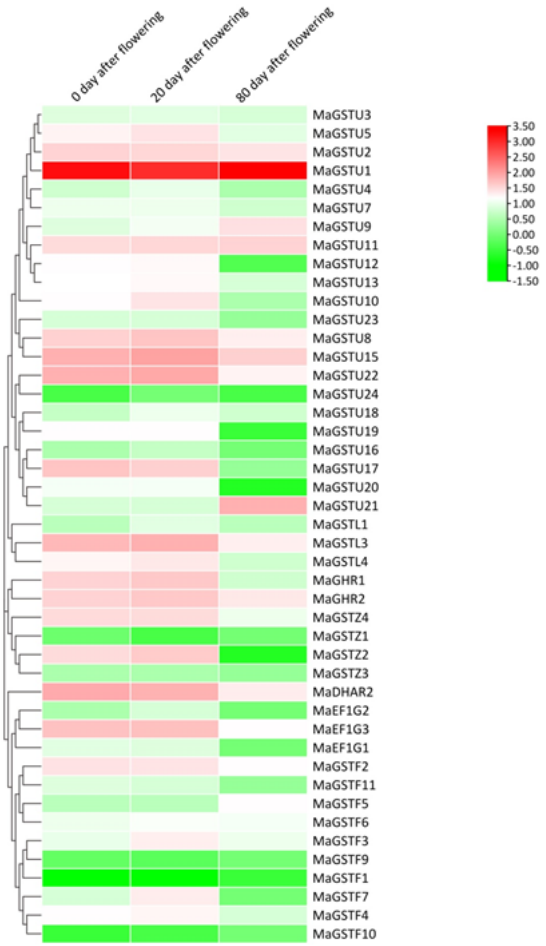

Figure 11

Expression profiling of MaGST genes based on RNA-seq data at fruit development stage. The heat-map was generated by TBtools software. The scale represents signal intensity of TPM values (converted in $\log_{10}$).

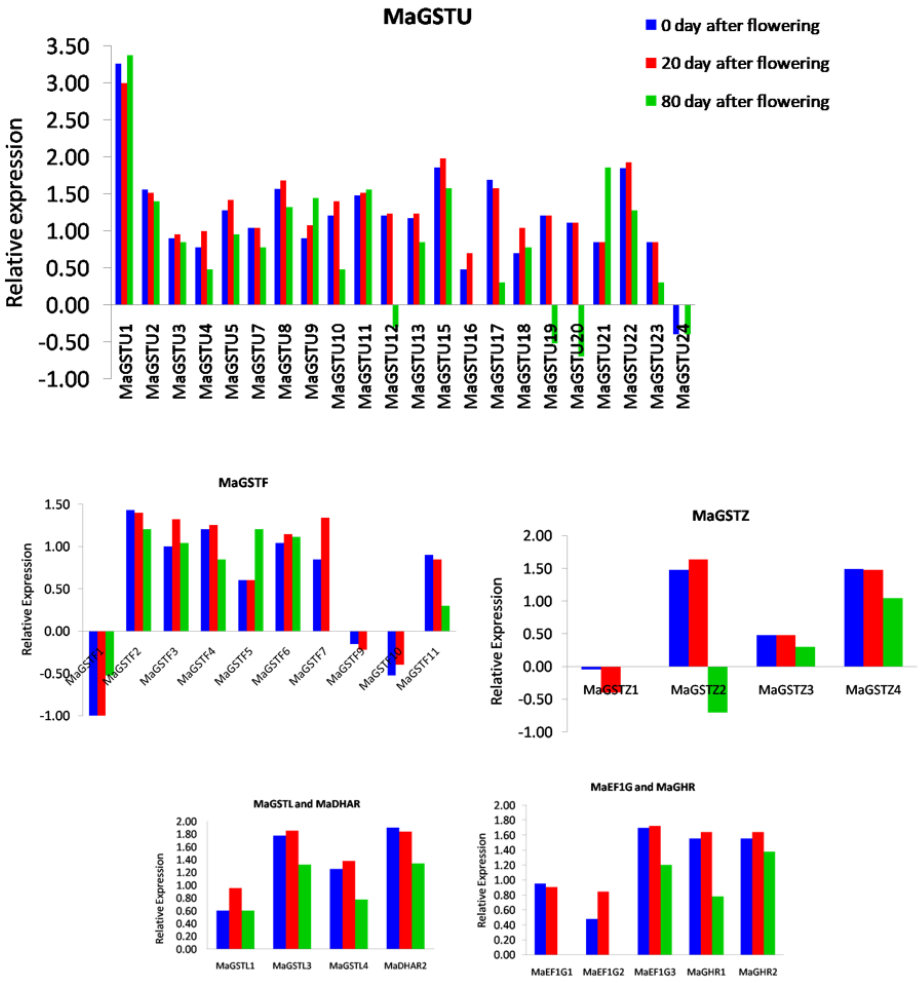

Figure 12

Expression profiles of MaGST genes during fruit development. Blue, red and green bar represents the expression level of 45 MaGST genes at 0, 20 and 80 days of flowering based on the $\log_{10}$ values of RNA seq data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- BananaSupplementaryfiles.docx