Functional & Integrative Genomics (2026) 26:161
https://doi.org/10.1007/s10142-026-01911-2

RESEARCH

Identification and characterization of C-glycosyltransferase (CGT) gene family associated with flavonoid-mediated shelf-life extension in pearl millet flour

Kommineni Jagadeesh[1,2] · Eswari B. Kancharla[2] · Mani Vetriventhan[1] · Kuldeep Singh[1] · Roopa Banerjee[1] · Kalyani Prasad[1] · Sudhakar Reddy Palakolanu[1]

Received: 26 March 2026 / Revised: 8 May 2026 / Accepted: 25 May 2026

Abstract

Pearl millet is a vital dryland, nutrient-rich C_4 crop with significant potential to reduce hidden hunger among vulnerable groups. However, its wider commercialization is limited by the short shelf life of its flour. Several endogenous compounds, particularly antioxidant flavonoids, are associated with reduced rancidity. Among these, C-glycosylflavonoids are considered important contributors to the enhanced shelf life of flour. C-glycosyl transferases (*CGTs*) encode proteins containing the UDP-glycosyl transferase domain and are predicted to play a regulatory role in flavonoid glycosylation. However, a detailed study on the quality of pearl millet flour is lacking. In this study, 32 *PgCGT* genes were identified and mapped to all the chromosomes except chromosome 3. Most of the *CGT* genes (18) are predicted to be located in chloroplasts. Gene structure analysis showed that most (25) genes had only one exon. Motif analysis identified conserved motifs within the *CGT* gene family in pearl millet. Comparative synteny analysis across cereal genomes suggested that purifying selection likely shaped the evolution of this gene family, with 15, 18, 20, and 15 conserved orthologs identified in rice, sorghum, foxtail millet, and maize, respectively. Promoter analysis identified *cis*-regulatory elements predicted to be associated with seed-specific regulation and endosperm-specific expression. The low-rancid genotype DS-SP-15 showed significantly higher flavonoid content at day 0 (131.44 mg QE/100 g) and day 10 (141.88 mg QE/100 g) than the high-rancid genotype DS-SP-9, which recorded 107.04 mg QE/100 g and 117.17 mg QE/100 g at the respective time points. The observed gene expression differences in contrasting inbred lines at various times suggest a potential role in flour shelf life in pearl millet. This study provides a comprehensive characterization of the *PgCGT* gene family and highlights possible functional divergence among its members. These results provide a solid foundation for future functional validation and targeted genome editing to improve grain quality and extend the shelf life of pearl millet flour.

Keywords C-glycosyltransferases · Pearl millet · Grain quality · PSPG-box motif · Gene expression · Rancidity

Introduction

Pearl millet [*Pennisetum glaucum* (L.) R. Br.] is one of the primary millets grown in arid and semi-arid regions of Asia and Africa. Globally, it is cultivated on approximately 30 million ha, predominantly in these regions, with about 60% of the area in Africa and 35% in Asia (Satyavathi et al. 2024). India is the largest producer of pearl millet, with cultivation on 7.21 million ha, producing 10.86 million metric tons at a productivity of 1507 Kg/ha (https://www.indiastat.com/). Owing to its nutritional richness and micronutrients, such as iron (Fe) and zinc (Zn), as well as

✉ Mani Vetriventhan
vetriventhan.mani@icrisat.org

✉ Sudhakar Reddy Palakolanu
Sudhakarreddy.Palakolanu@icrisat.org

[1] International Crops Research Institute for Semi-Arid Tropics (ICRISAT), Patancheru, Hyderabad 502324, India

[2] Prof. Jayashankar Telangana Agricultural University (PJTAU), Rajendranagar, Hyderabad 500030, India

Published online: 24 June 2026

several other essential nutrients, it is considered a nutrient-dense food crop (Kanatti et al. 2014; Govindaraj et al. 2020; Satyavathi et al. 2021; Singh et al. 2024). In addition to its high nutritional value, pearl millet contains a relatively high fat content (5.24–9.99%) (Tomar et al. 2021). Despite its superior agronomic resilience, nutritional richness, and widespread cultivation, its commercial utilization is limited by the short shelf life of its flour. This is primarily due to the activity of several oxidative enzymes that cause rancidity. This remains a major bottleneck and underscores the need to elucidate the biochemical and genetic mechanisms underlying rancidity development.

Rancidity in pearl millet is a complex phenomenon resulting from the interaction of several enzymes, including lipases (Aher et al. 2022), lipoxygenases (LOXs) (Palakolanu et al. 2026), and peroxidases, as well as environmental factors, such as oxygen, temperature, and moisture. Rancidity development is broadly classified into two types: hydrolytic and oxidative rancidity (Goswami et al. 2020). Hydrolytic rancidity involves the breakdown of fatty acids into free fatty acids, thereby increasing the acid value of pearl millet flour (Bhargavi et al. 2024). These free fatty acids serve as substrates for enzymes that initiate oxidative rancidity, which then oxidize them to yield aldehydes and ketones (Sharma et al. 2022). This results in off-odors and off-flavors in the flour, which ultimately leads to its deterioration. Owing to the greater role of these enzymes in rancidity development, considerable research efforts have focused on the identification and characterization of gene families encoding phospholipases (Moin et al. 2024), TAG-lipases (Aher et al. 2022), and lipoxygenases (LOXs) (Palakolanu et al. 2026). In addition to the degradation compounds, hydrolytic and oxidative activities also generate free radicals, which further deteriorate flour quality during storage. To mitigate such oxidative stress, several metabolic pathways have been activated, resulting in the synthesis of antioxidant compounds, among which flavonoids play a pivotal role. This enhanced the importance of flavonoids and their role in mitigating rancidity. Among the various flavonoid compounds with antioxidative and free radical-scavenging potential, notably C-glycosylflavonoid (C-Glf) compounds have greater potential to mitigate the negative impact of free radicals. In the case of pearl millet, C-glycosyl-vitexin, vitexin, glycosyl-orientin (Akingbala 1991), and orientin have been quantified (Ali et al. 2023). This underscored the importance of characterizing the C-glycosyltransferase gene family in combating rancidity.

C-glycosyltransferases (CGTs) are part of the glycosyltransferases (GTs; EC 2.4.x. y) belonging to a large family commonly found in viruses, archaea, bacteria, and eukaryotes. Recent studies have classified the glycosyl transferases into 138 families (https://www.cazy.org/GlycosylTransferase-family). These enzymes catalyze the transfer of UDP-activated sugars to a wide range of acceptors. Glycosylation is primarily classified based on the type of glycosidic linkage formed. The transfer of sugar to the hydroxyl group and to the carbon atom of the flavonoid skeleton is called O-glycosylation and C-glycosylation, respectively. Functional characterization of CGTs has been reported in maize, where these genes are associated with the biosynthesis of the insecticidal C-glycosyl flavone maysin (Ferreyra et al. 2013). Similarly, CGT genes have been identified in horticultural crops such as citrus, where they are involved in the production of value-added di-C-glucosyl flavonoids derived from phloretin (Ito et al. 2017) and in mango, where they participate in the biosynthesis of mangiferin, a compound of pharmaceutical importance (Chen et al. 2015). In buckwheat, the *CGT* gene has been reported to play a physiological role during seed germination, with expression specifically observed in cotyledons (Nagatomo et al. 2014). Flavonoids are widely distributed across many cereal and plant species and contribute significantly to their antioxidant potential. Studies have reported considerable variation in flavonoid content in crops such as sorghum (Ghimire et al. 2021), rice (Shen et al. 2009), maize (Kaur and Kaur 2016), and wheat (Punia et al. 2019). In the case of pearl millet, significant variability in total flavonoid content has been reported among the 13 pearl millet cultivars, which ranged from 54.49 to 116.51 mg/100 g catechin equivalents (Krishnan and Meera 2017). Although several genes linked to flavonoids have been identified, a comprehensive genome-wide analysis of *CGT* genes in pearl millet remains lacking. A preliminary study on an orthologous gene involved in C-GF biosynthesis has been documented (Boncompagni et al. 2018). Therefore, this research aims to systematically identify and analyze *CGT* genes across the entire genome to understand their variation, evolutionary connections, and possible functions in affecting flour quality in pearl millet.

Materials and methods

Sequence retrieval from databases

Genes involved in the C-glycosylflavonoid biosynthetic pathway reported in rice (8), along with genes reported in pearl millet (5), soybean (1), maize (1), citrus (3), and mango (1) (Boncompagni et al. 2018; Dorjjugder et al. 2021), were used as reference sequences for genome-wide identification of candidate homologs in pearl millet. The corresponding amino acid sequences were retrieved from UniProt (https://www.uniprot.org/) and the NCBI protein database. These reference protein sequences were used as queries in BLASTP searches against the de novo assembled genome of the Indian pearl millet

genotype PI583800 available on the Milletdb platform (http://milletdb.novogene.com/tools/blast). An HMMER search was performed using the command-line interface to identify proteins containing the conserved UDP-glycosyltransferase domains (PF00201), and candidates harboring this domain were retained for further analysis. Redundant hits were manually curated to obtain a non-redundant set of unique candidate homologs for downstream analyses. The gene, coding DNA sequence (CDS), and protein sequences of the selected candidates were subsequently extracted from the Milletdb database for downstream analyses. (http://milletdb.novogene.com/tools/sequenceextraction).

Phylogenetic analysis of CGT genes

The protein sequences of the unique candidates identified were extracted from Milletdb database (http://milletdb.novogene.com/tools/blast). In addition to the unique candidates, genes reported as UDP-glycosyltransferases and C-glycosyltransferases were also considered. Multiple sequence alignment was performed using the ClustalW method in MEGA11. The phylogenetic tree was constructed using MEGA11 software with the neighbour-joining (NJ) method, with the Poisson substitution model, and 1000 bootstrap replicates under standard settings. The phylogenetic tree was further formatted using iTOL (https://itol.embl.de).

Chromosomal mapping, physicochemical properties, and subcellular localization

Information regarding chromosomal positions, strand orientation, gene lengths, and open reading frames (ORFs) length, and the number of exons and introns for identified candidate genes was retrieved from the Milletdb Database (http://milletdb.novogene.com/home). Based on the chromosomal positions of the genes and the exon–intron organization, chromosomal mapping and gene structure were performed using TBtools-II (Chen et al. 2020). Additionally, physicochemical properties, including the number of amino acids, molecular weight (kDa), isoelectric point (pI), aliphatic index, Grand Average of Hydropathy (GRAVY), and stability index, were computed using the ProtParam database (https://web.expasy.org/protparam/). The subcellular localization of the predicted proteins was determined using the WoLF-PSORT database (https://wolfpsort.hgc.jp). The subcellular localization of each protein was assigned based on the highest score.

Motifs and cis-regulatory elements identification

A 1500 bp upstream promoter sequence from the transcription start site of each identified gene was extracted, and *cis*-regulatory elements were identified using the PlantCare Database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/). Conserved motifs within the coding DNA sequences (CDS) were identified using (http://meme-suite.org/index.html). The distribution of identified cis-acting regulatory elements across genes was visualized using a bubble plot generated in Python.

Secondary and tertiary structure analysis of proteins

The secondary structures of the predicted protein sequences were analyzed using SOPMA (Self-Optimized Prediction Method with Alignment) (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html), considering four conformational states: α-helices, β-strands, turns, and random coils. A similarity threshold of 8 and a window width of 17 were used for the prediction. Three-dimensional (3D) protein structures were predicted using AlphaFold (https://alphafold.ebi.ac.uk).

Evolutionary analysis

Dual synteny analysis was performed to investigate the evolutionary conservation and divergence of *CGT* genes between pearl millet and other cereal crops, including rice, maize, sorghum, and foxtail millet, using TBtools-II (Chen et al. 2020). Homologous gene pairs identified through synteny analysis were subjected to evolutionary analysis by estimating Ka and Ks values and substitution rates using TBtools (Chen et al. 2020). Divergence time was estimated using the formula $T = Ks/2\lambda$, where λ represents the neutral substitution rate, assumed to be 6.5×10^{-9} substitutions per site per year (Li et al. 2022).

Plant material and treatment conditions

Two inbred lines (DS-SP-15 and DS-SP-9), contrasting in rapid peroxide value (RPV; unpublished data), were used as genetic material for gene expression studies. These inbred lines were derived using the single-seed descent (SSD) method through continuous selfing up to the S5 generation from the biofortified cultivar Dhanashakti. Deglumed grains of the inbred lines were milled into flour using a Foss Cyclotech-1093 mill equipped with a 0.5 mm sieve. Flour samples collected immediately after milling of these contrasting inbred lines were considered day 0 samples. Furthermore, the samples stored for 10 days under accelerated conditions (37°C and 75% RH) were designated day 10 samples and used to estimate total flavonoid content and analyze gene expression.

Estimation of rapid peroxide value (RPV)

Two hundred milligrams of milled pearl millet flour from each inbred line was weighed into conical flasks, with three technical replicates per inbred line. Each sample was mixed with 1.5 mL of a chloroform–methanol mixture (7:3, v/v). The mixture was vortexed for 1 min to extract lipids, then centrifuged at 4000 rpm for 5 min to obtain a clear supernatant. From the obtained extract, an aliquot of 0.8 mL was transferred to a fresh tube containing 1 mL of 25 mM H_2SO_4 and 0.2 mL of 25 mM ferrous ammonium sulfate. Subsequently, 10 µL of xylenol orange reagent was added, followed by vortexing for 1 min to ensure proper mixing. The reaction mixture was incubated at room temperature for 5 min. After incubation, absorbance was recorded at 589 nm using a Thermo Scientific Multiskan Sky microplate spectrophotometer. The RPV was calculated using a standard calibration curve prepared from ferric chloride standards ranging from 0 to 20 µg Fe^{3+}. The RPV of each sample was expressed as µg Fe^{3+} g^{-1} or mEq peroxides kg^{-1} on a dry weight basis.

Estimation of total flavonoid content

The total flavonoid content was estimated in triplicate using the aluminium chloride colorimetric method (Isla et al. 2014). The samples were extracted in 80% acidified methanol, then vortexed, diluted, and filtered prior to analysis. Briefly, 1 mL of the extract was mixed with 4 mL of distilled water and 0.3 mL of 5% $NaNO_2$, then incubated for 5 min. Subsequently, 0.3 mL of 10% $AlCl_3$ was added, and the mixture was incubated for 6 min, followed by the addition of 2 mL of 1 M NaOH. The total reaction volume was adjusted to 10 mL with distilled water. Absorbance was measured at 510 nm using a microplate reader against a reagent blank. A standard calibration curve was generated using Quercetin, ranging from (10–100 µg/mL) ($R^2 \geq 0.99$). The total flavonoid content was expressed as mg quercetin equivalents (mg QE/100 g sample on a dry weight basis). Precision was maintained with % RSD$\leq$2%, and the blank absorbance was $\leq$0.1.

RNA isolation and quantitative real-time PCR (qRT-PCR) analysis

Total RNA was extracted independently from flour samples of each genotype collected at two sampling time points (Day 0 and Day 10), using two biological replicates per treatment, following a slightly modified protocol provided by the kit manufacturer (https://www.mn-net.com/bioanalysis/kits/rna/). RNA quality was assessed by agarose gel electrophoresis, and RNA concentration was quantified using a Qubit 4 fluorometer (Thermo Fisher, USA). Gene-specific primers were designed using Primer3 Plus (https://primer3.ut.ee) with amplicon sizes of 90–180 base pairs. qRT-PCR was performed using a Bio-Rad CFX96 Real-Time PCR System with a one-step RT-PCR kit in a 10 µL reaction volume with reaction components, including 2 × One Step RT-PCR Buffer (5 µL), RTase Enzyme Mix (0.4 µL), primer (0.8 µl), RNase-free water (2.6 µl), and RNA (1.2 µl). The standard thermal cycling program included reverse transcription at 42°C for 5 min, followed by an initial denaturation step at 95°C for 2 min. PCR amplification was carried out for 40 cycles, consisting of denaturation at 95°C for 5s, and combined annealing/extension at 62°C for 25s, with fluorescence data collected during the annealing/extension step. Following amplification, melt curve (amplicon dissociation) analysis was conducted to confirm the specificity of the PCR products for each gene. For each biological replicate, qRT-PCR reactions were performed in three technical replicates. The relative expression of C-glycosyltransferase genes was calculated using $2^{(-\Delta\Delta CT)}$ (Livak and Schmittgen 2001) with normalization against the reference gene *PgEIF4a* (Reddy et al. 2015). Expression data were summarized as mean values across biological replicates, and clustering-based heat maps of CGT gene expression patterns were generated using R.

Protein–protein interaction

The protein–protein interaction (PPI) network was constructed using the STRING database (https://string-db.org). The amino acid sequences of the identified proteins were used as input, and when required, orthologous gene identifiers were mapped to a model plant reference (*Oryza sativa*) to ensure compatibility with the STRING database. A medium-to-high confidence interaction score threshold ($\geq$0.4) was applied to filter for biologically meaningful associations. The network expansion was limited to a maximum of 10 first-shell interactors to capture direct functional associations while minimizing network complexity.

Results

Identification of CGT genes in pearl millet

Based on previously reported *CGT* genes across various crops, including rice, maize, and pearl millet, a genome-wide BLASTp search was conducted against the pearl millet genotype PI583800. After removing redundant proteins, the search yielded 151 unique proteins that were analyzed using HMMER. This scan identified 65 genes containing the conserved UDP-glycosyltransferase domain (PF00201), which

underpins the identification of glycosyl transferase genes. Among these, one gene was located on an unanchored scaffold and was excluded from further classification analysis, while the remaining 64 genes were retained for phylogenetic analysis.

In addition to the *glycosyl transferase* genes identified through BLAST, several previously reported genes from different crops were included in the phylogenetic analysis. The phylogenetic analysis classified the 64 glycosyl transferase genes into two well-defined clades, corresponding to *UDP-glycosyl transferases* and *C-glycosyltransferases*. This classification was further supported by the clustering of previously characterized *UGTs* and *CGTs* within their respective groups. Based on the phylogenetic classification, among the 64 genes, 32 were identified as *CGTs* and 32 as *UGTs* (Fig. 1a). The 32 *CGT* genes were subsequently analyzed.

Chromosomal mapping of PgCGT genes

The identified genes were distributed unevenly across the seven chromosomes of the pearl millet genome (Fig. 1b). Chromosome 6 harbored the highest number of *PgCGT* genes (12), followed by chromosome 7 (7 genes), chromosome 2 (6 genes), and chromosome 1 (5 genes). Only one gene each was located on chromosomes 4 and 5, whereas no *PgCGT* genes were identified on chromosome 3.

Physiochemical properties of pearl millet CGT genes

The physicochemical properties and subcellular localizations of the *CGT* genes are listed in Table 1. Among the 32 identified *CGT* genes, protein lengths ranged from 415 to 614 amino acids. The molecular weights of the proteins ranged from 45,296.14 to 66,641.35 Da. The isoelectric points (pI) ranged from 4.94 to 7.02, with most proteins exhibiting pI values of 5 and 6, indicating an acidic nature and a negative charge at physiological pH. The gene *PgCGT1.1* had the lowest isoelectric point (pI; 4.94), whereas *PgCGT5.1* had a higher pI value (7.02). Furthermore, the aliphatic index of the identified proteins ranged from 83.42 to 99.23, indicating relatively high thermostability. The grand average of hydropathy (GRAVY) values of the proteins ranged from −0.187 to 0.222, indicating that the proteins were moderately hydrophilic to moderately hydrophobic.

Regarding subcellular localization, approximately 50% of the proteins were predicted to localize to the chloroplast (16). Similarly, five have been localized to the cytoplasm, two to the endoplasmic reticulum, and three to the extracellular space. Two genes showed dual localization, that is, chloroplast/cytoplasm by *PgCGT1.1* and chloroplast/endoplasmic reticulum (ER) by *PgCGT6.4*.

Identification and Sequence Conservation of the PSPG Motif

The PSPG box is a defining feature of glycosyltransferase proteins. Based on motif analysis using ScanProsite, the presence of the PSPG box was confirmed in only 25 of the 32 genes. However, manual searches using conserved amino acid sequences identified the PSPG box in an additional six genes: *PgCGT6.2*, *PgCGT6.5*, *PgCGT6.10*, *PgCGT7.1*, *PgCGT7.4*, and *PgCGT7.7*. The PSPG box sequences of the CGT proteins were extracted and aligned using multiple sequence alignment (Fig. S1). It presents the consensus logo, highlighting the most common amino acid

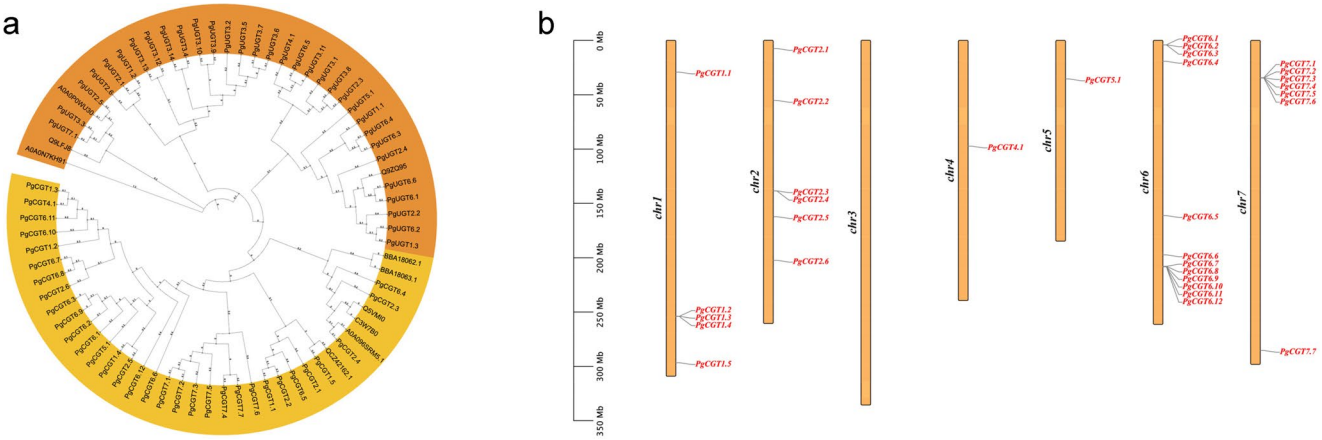

Fig. 1a Circular phylogenetic tree showing the evolutionary relationship of glycosyltransferase proteins in pearl millet. Phylogenetic analysis of the amino acid sequences was performed using MEGA11 with the maximum likelihood method with 1000 bootstrap replications. An unrooted phylogenetic tree was generated using iTOL. The tree displays two main clusters: an outer clade of O-glycosyltransferases (*PgUGT*) and an inner clade of C-glycosyltransferases (*PgCGT*). Fig. **1b** Chromosomal distribution of *PgCGT* genes across the pearl millet genome. The physical positions of *CGT* genes were mapped to individual chromosomes using genome annotation, illustrating their genomic locations and distribution patterns across pearl millet chromosomes. The scale bar on the left indicates the size of the chromosomes

Table 1 Predicted genomic features and physicochemical properties of *PgCGT* genes identified in pearl millet, including chromosomal location, gene structure, protein length, molecular weight, isoelectric point, and subcellular localization

Gene name	Gene ID	Chr	Genomic sequence (bp)	ORF (bp)	No. of exons	Protein properties							
						Number of Amino Acids	Molecular Weight (Da)	UDP domain		pI	Aliphatic index	GRAVY	Subcellular localization
								Start	End				
PgCGT1.1	PMD1G00816.1	1	2,841	1539	1	512	54256.46	374	417	4.94	91.13	0.046	Chloroplast/Cytoplasm
PgCGT1.2	PMD1G05852.1	1	2,078	1443	1	480	51761.84	354	397	5.72	94.12	0.123	Plastid/ER
PgCGT1.3	PMD1G05853.1	1	2,038	1404	1	467	51299.37	337	380	5.32	95.67	0.01	Chloroplast
PgCGT1.4	PMD1G05858.1	1	1,665	1,440	2	479	52235.9	342	385	5.17	95.07	0.002	Chloroplast
PgCGT1.5	PMD1G07473.1	1	1,506	1,491	1	496	52500.52	366	409	6.14	99.23	0.183	Chloroplast
PgCGT2.1	PMD2G00370.1	2	1,762	1,482	1	493	51615.9	362	405	5.73	90.73	0.068	Cytoplasm
PgCGT2.2	PMD2G01678.1	2	1,710	1,485	1	494	52988.32	361	404	5.86	92.37	-0.028	Chloroplast
PgCGT2.3	PMD2G03100.1	2	1,902	1,470	1	489	51809.97	362	405	5.55	92.82	0.14	Chloroplast
PgCGT2.4	PMD2G03101.1	2	1,740	1,404	1	467	49134.01	344	387	5.41	94.07	0.222	Plastid/ER
PgCGT2.5	PMD2G03793.1	2	1,742	1,491	1	496	54063.12	365	409	5.38	95.32	0.045	Chloroplast
PgCGT2.6	PMD2G05284.1	2	3,744	1,434	2	477	50873.56	343	386	4.99	97.8	0.144	Chloroplast
PgCGT4.1	PMD4G03235.1	4	1,527	1,476	1	491	53093.61	366	409	6.69	88.68	0.055	Chloroplast
PgCGT5.1	PMD5G01988.1	5	3,787	1,845	2	614	66641.35	358	401	7.02	83.42	-0.112	Chloroplast
PgCGT6.1	PMD6G00199.1	6	1,628	1,383	2	460	50210.18	335	378	5.51	90.5	0.004	cytoplasm
*PgCGT6.2**	PMD6G00200.1	6	1,901	1,248	1	415	45296.14	288	332	5.37	91.59	-0.092	cytoplasm
PgCGT6.3	PMD6G00201.1	6	1,464	1386	3	461	50605.08	334	377	6.08	84.38	-0.187	cytoplasm
*PgCGT6.4***	PMD6G00855.1	6	1,686	1,395	1	464	49757.88	288	332	5.27	95.24	0.086	Chloroplast/ER
PgCGT6.5	PMD6G03816.1	6	1,821	1,440	1	479	51076.67	350	393	5.93	97.64	0.125	ER
PgCGT6.6	PMD6G04667.1	6	1,727	1467	1	488	52370.4	360	403	5.8	93.57	0.118	Chloroplast
PgCGT6.7	PMD6G05003.1	6	2,241	1404	1	467	50404.93	344	387	5.09	88.16	0.001	Chloroplast
PgCGT6.8	PMD6G05004.1	6	1,967	1,413	1	470	51536.42	343	386	5.25	93.17	0.005	Chloroplast
PgCGT6.9	PMD6G05007.1	6	1,646	1,452	1	483	52858.81	356	399	5.1	90.6	-0.095	Chloroplast
*PgCGT6.10**	PMD6G05012.1	6	1,434	1,434	1	477	52318.48	345	388	5.29	95.68	0.087	Extracellular
PgCGT6.11	PMD6G05016.1	6	1,846	1,704	1	567	61180.58	441	484	6.14	91.36	-0.079	Cytoplasm
PgCGT6.12	PMD6G05017.1	6	1,566	1,419	1	472	50647.17	345	388	5.5	93.86	0.103	Chloroplast
*PgCGT7.1**	PMD7G01538.1	7	1,742	1590	1	529	57407.34	343	386	5.46	93.46	0.152	Chloroplast
PgCGT7.2	PMD7G01540.1	7	1,431	1,431	1	476	51258.69	344	387	6.07	96.11	0.179	Chloroplast
PgCGT7.3	PMD7G01542.1	7	2,034	1,464	1	487	52213.04	348	391	5.48	87.39	0.12	Extracellular
*PgCGT7.4**	PMD7G01543.1	7	1,637	1,455	1	484	52014	350	393	5.91	89.34	0.04	Extracellular
PgCGT7.5	PMD7G01546.1	7	1,452	1,452	1	483	51203.99	350	393	5.35	90.7	0.095	ER
PgCGT7.6	PMD7G01551.1	7	1,462	1374	2	457	49171.14	322	365	5.64	90.2	-0.014	Chloroplast
*PgCGT7.7**	PMD7G07001.1	7	3227	1476	2	491	52430.2	360	403	5.65	87.52	0.099	Peroxisome

* Indicates the genes where the PSPG box sequence was manually identified, and ** marks the gene in which the PSPG box sequence was not completely conserved

at each position, while a bar graph illustrates conservation level across the genes. The PSPG domain ranged from 44 to 45 amino acids; *PgCGT2.5* possessed a 45-amino acid domain, as identified by ScanProsite. Similarly, *PgCGT6.2*, which contained the PSPG domain identified manually, also harbored a 45-amino-acid domain. In the case of *PgCGT6.4*, the PSPG box sequence was absent, whereas certain amino acid motifs, such as HCGWNS, were highly conserved. Notably conserved sequences across the genes include WAPQ, marking the beginning of the N-terminal PSPG box; HCGWNS, located in the central region; and YAEQ, present at the end of the C-terminal region.

Gene structure analysis

Gene structure predictions indicated that approximately 78% of the genes (25 genes) contained a single exon without introns (Fig. 2a). This simple gene structure is characterized by the absence of introns, resulting in a continuous coding sequence without interruptions of introns. Among the remaining genes, six contained two exons, and one contained three exons.

Motif analysis

Motif analysis identified 10 conserved motifs, designated as Motifs 1–10 (Fig. 2b). The sequences and length of the motifs identified are presented in Table S1. The functional similarity among the proteins was also reflected in the shared motif composition. Among the 32 CGT genes, all 10 motifs were consistently present in 30 genes, whereas only nine motifs were detected in two genes, *PgCGT6.1* and *PgCGT6.2*. Motif 1 was absent in *PgCGT6.1*, while motif 2 was absent in *PgCGT6.2*.

Cis-regulatory elements prediction

Among the 32 *PgCGT* genes, seven genes were in proximity to neighboring genes, resulting in overlaps between their upstream promoter regions and the coding sequences of adjacent genes. Therefore, these genes were excluded, and 25 genes were used to identify *cis*-regulatory elements. Analysis of the promoter sequences, 1,500 bp upstream, identified approximately 89 cis-regulatory elements, of which 48 had known functions, and 41 had unknown functions. A bubble plot was generated based on the number of *cis*-regulatory elements identified in each gene (Fig. 3). The identified cis-elements were further classified into functional groups, including core functional elements, light-response regulatory elements, growth- and development-regulatory elements, and elements involved in abiotic and biotic stress regulation (Table S2).

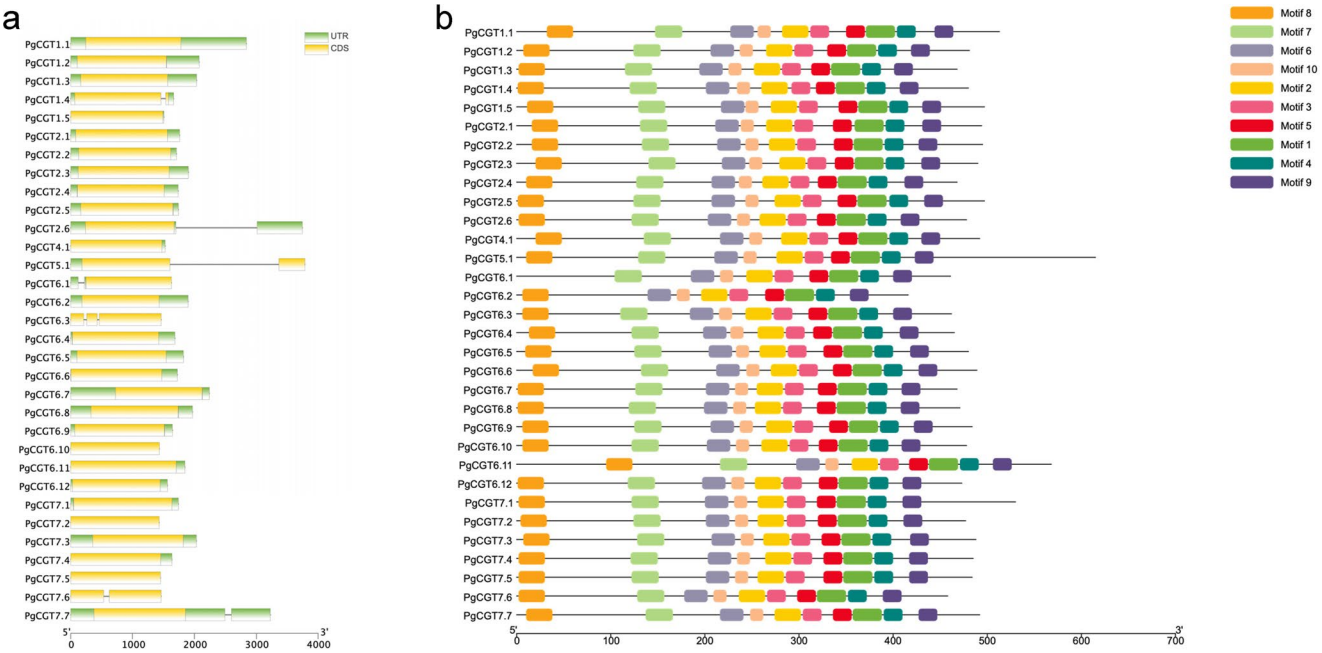

Fig. 2a Gene structure of *PgCGT* genes in pearl millet. The gene structure illustrates the arrangement of exons and introns for each identified *PgCGT* gene based on genomic sequence annotation. Yellow boxes represent coding sequences (CDS/exons), green boxes indicate untranslated regions (UTRs), and thin lines connecting the boxes denote introns. The scale at the bottom indicates gene length in base pairs from the 5′ to 3′ direction. **2b** Conserved motif analysis of the *PgCGT* protein family in pearl millet. Distribution and organization of these motifs across the 32 *PgCGT* proteins. Each colored box represents a distinct conserved motif, and the length of the protein sequences is indicated by the horizontal lines from 5′ to 3′

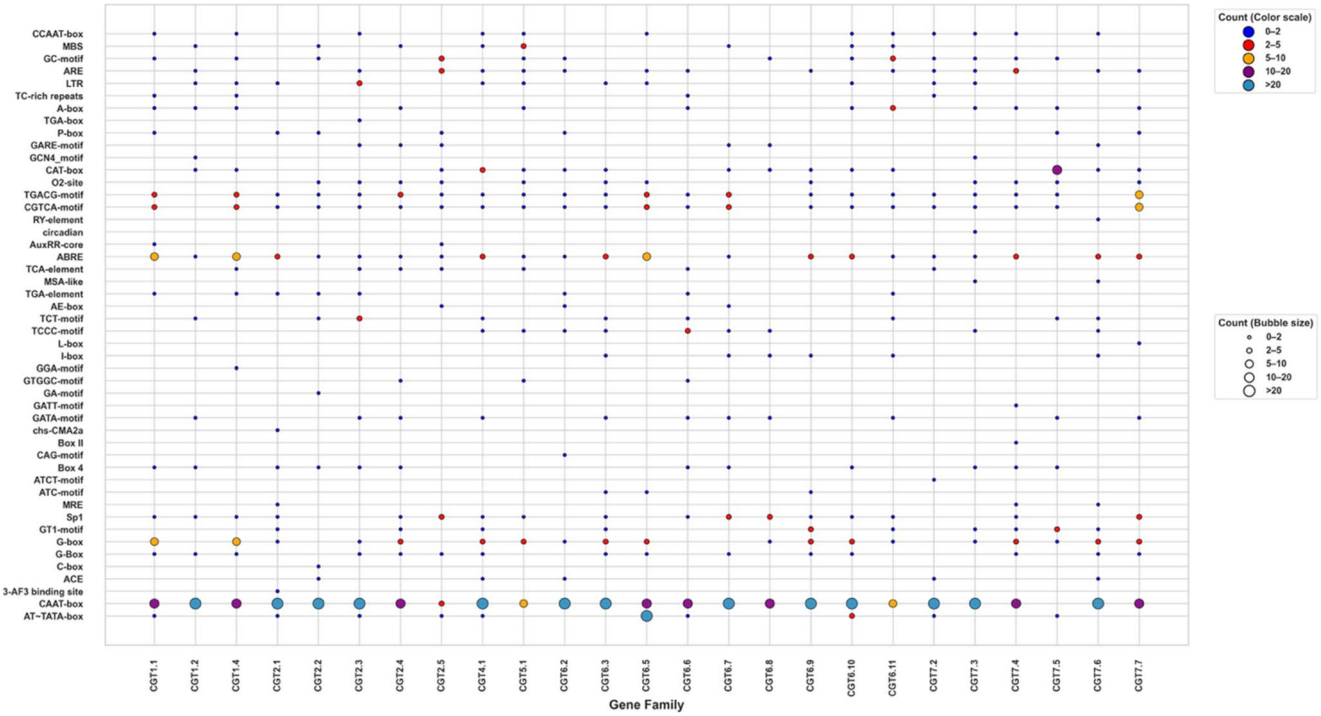

Fig. 3 Bubble plot showing the distribution of cis-regulatory elements identified in the promoter regions of *PgCGT* genes in pearl millet. The x-axis represents *PgCGT* genes, and the y-axis represents different cis-acting regulatory elements. The size and color of the bubbles indicate the number of occurrences of each element within the promoter regions

Structural properties of CGT proteins

Secondary-structure analysis of the *PgCGT* gene family revealed the proportions of α-helices, β-strands (extended strands), β-turns, and random coils in the protein structures (Table 2). The protein structures mainly consist of α-helices (37.21%–44.35%) and random coils (35.76%–47.07%). The percentage of extended strands ranged from 11.73% to 15.25%, whereas β-turns were fewer, ranging from 3.91% to 6.64%. The three-dimensional (3D) structures predicted with AlphaFold have predicted Local Distance Difference test (pLDDT) scores above 90%, indicating high reliability and accuracy (Fig. S2).

Collinearity analysis of PgCGT genes

Collinearity analysis provides insights into the evolution of the *PgCGT* gene family by comparing pearl millet with rice, maize, sorghum, and foxtail millet (Fig. 4). Notably, strong collinearity was observed between pearl millet *CGT* genes and those in other species. Synteny analysis identified 15 orthologous gene pairs between rice and pearl millet, 18 between sorghum and pearl millet, 20 between foxtail millet and pearl millet, and 15 between pearl millet and maize. The Ka/Ks ratio was also calculated, yielding values ranging from 0.267 to 0.572 for rice–pearl millet pairs, 0.248 to 0.654 for pearl millet–maize pairs, 0.196 to 0.601 for pearl millet–sorghum pairs, and 0.210 to 0.621 for pearl millet–foxtail millet pairs. The estimated divergence times for these orthologs across the four species ranged from approximately 9.242 to 77.796 million years ago.

Selection of contrasting genotypes for rancidity

Rapid peroxide value (RPV), an indicator of lipid oxidation and rancidity in pearl millet flour, was used to identify contrasting inbred lines. Among the evaluated lines, DS-SP-9 showed higher peroxide values during storage, whereas DS-SP-15 exhibited comparatively lower values (Table 3). On day 10 of storage, DS-SP-9 recorded an RPV of 13.59 mg ferric/kg, compared with 8.68 mg ferric/kg in DS-SP-15. Based on these contrasting rancidity patterns on day 10, which was identified as the key stage for peroxide accumulation, these two inbreds, DS-SP-15 (low rancidity) and DS-SP-9 (high rancidity), were selected for flavonoid estimation and gene expression analysis.

Total flavonoid content in high and low rancid inbreds

The total flavonoid content, which contributes to antioxidant capacity by scavenging free radicals, was estimated in contrasting lines at days 0, 10, and 20. Differential trends were observed in the contrasting lines (Fig. 5). In fresh flour

Table 2 Predicted secondary structure composition of *PgCGT* proteins in pearl millet, showing the percentage distribution of α-helix, extended strand (β-sheet), β-turn, and random coil

Gene	Gene ID	α-Helix (%)	Extended Strand (%)	β-Turn (%)	Random Coil (%)
PgCGT1.1	PMD1G00816.1	41.80%	13.28%	4.49%	40.43%
PgCGT1.2	PMD1G05852.1	42.92%	14.58%	4.58%	37.92%
PgCGT1.3	PMD1G05853.1	43.47%	15.20%	4.07%	37.26%
PgCGT1.4	PMD1G05858.1	43.22%	13.78%	3.97%	39.04%
PgCGT1.5	PMD1G07473.1	40.32%	14.92%	4.84%	39.92%
PgCGT2.1	PMD2G00370.1	40.16%	13.39%	6.09%	40.37%
PgCGT2.2	PMD2G01678.1	40.69%	13.97%	4.66%	40.69%
PgCGT2.3	PMD2G03100.1	41.92%	13.91%	4.29%	39.88%
PgCGT2.4	PMD2G03101.1	44.33%	14.78%	5.14%	35.76%
PgCGT2.5	PMD2G03793.1	41.13%	14.31%	5.44%	39.11%
PgCGT2.6	PMD2G05284.1	43.40%	14.26%	4.19%	38.16%
PgCGT4.1	PMD4G03235.1	42.57%	14.05%	5.70%	37.68%
PgCGT5.1	PMD5G01988.1	37.30%	11.73%	3.91%	47.07%
PgCGT6.1	PMD6G00199.1	44.35%	13.48%	4.78%	37.39%
PgCGT6.2	PMD6G00200.1	40.96%	14.22%	5.06%	39.76%
PgCGT6.3	PMD6G00201.1	43.60%	14.75%	4.99%	36.66%
PgCGT6.4	PMD6G00855.1	42.46%	14.22%	4.31%	39.01%
PgCGT6.5	PMD6G03816.1	42.38%	13.78%	4.59%	39.25%
PgCGT6.6	PMD6G04667.1	42.21%	14.34%	5.33%	38.11%
PgCGT6.7	PMD6G05003.1	42.40%	14.78%	6.64%	36.19%
PgCGT6.8	PMD6G05004.1	43.40%	14.47%	4.68%	37.45%
PgCGT6.9	PMD6G05007.1	41.61%	14.29%	4.55%	39.54%
PgCGT6.10	PMD6G05012.1	41.93%	14.47%	4.61%	38.99%
PgCGT6.11	PMD6G05016.1	37.21%	14.29%	4.06%	44.44%
PgCGT6.12	PMD6G05017.1	41.74%	15.25%	5.51%	37.50%
PgCGT7.1	PMD7G01538.1	40.45%	14.56%	4.73%	40.26%
PgCGT7.2	PMD7G01540.1	43.91%	14.08%	5.25%	36.76%
PgCGT7.3	PMD7G01542.1	43.12%	14.58%	5.54%	36.76%
PgCGT7.4	PMD7G01543.1	42.77%	14.67%	5.99%	36.57%
PgCGT7.5	PMD7G01546.1	40.99%	13.46%	4.76%	40.79%
PgCGT7.6	PMD7G01551.1	42.23%	15.10%	6.13%	36.54%
PgCGT7.7	PMD7G07001.1	42.97%	13.85%	4.89%	38.29%

(day 0), higher base flavonoid content was observed in the low-rancid inbred line DS-SP-15 (131.44 mg QE/100 g) compared with the high-rancid line DS-SP-9 (107.04 mg QE/100 g). A similar trend was observed on day 10, where DS-SP-15 (141.88 mg QE/100 g) exhibited higher values than DS-SP-9 (117.17 mg QE/100 g). However, on day 20, the high-rancid line DS-SP-9 showed higher flavonoid content (162.80 mg QE/100 g) than the low-rancid line DS-SP-15 (140.98 mg QE/100 g). The increased flavonoid accumulation in DS-SP-9 at later storage stages may reflect a stress-induced response to oxidative conditions rather than a preventive mechanism against rancidity. The declining trends in the low-rancid line (DS-SP-15) after day 10 may be associated with flavonoid utilization in free-radical scavenging, leading to thereby mitigating hydroperoxide formation. Overall, these flavonoid trends were consistent with the changes in the peroxide value, suggesting a possible association between oxidative stress and secondary metabolite accumulation during storage.

Gene expression profiling of PgCGT genes

The expression patterns of the identified *PgCGT* genes were analyzed using qRT-PCR in grain flour samples collected on day 0 and day 10. Heatmap visualization, combined with hierarchical clustering based on Z-score-normalized mean expression values, revealed clear temporal and genotype-specific differences in gene expression between the contrasting inbred lines (Fig. 6). The analysis identified a distinct cluster of genes, namely *PgCGT2.1*, *PgCGT6.7*, *PgCGT6.9*, *PgCGT6.5*, *PgCGT6.6*, *PgCGT1.1*, and *PgCGT7.5* which showed high expression in the low-rancid inbred DS-SP-15 on day 10, while exhibiting minimal expression in the high-rancid inbred DS-SP-9 at the same time point. Genotype-wise fold-change analysis further confirmed significantly higher expression of these genes in DS-SP-15 compared with DS-SP-9, with fold changes of 1.33 (*PgCGT2.1*), 7.45 (*PgCGT6.7*), 7.61 (*PgCGT6.9*), 50.45 (*PgCGT6.5*), 85.71 (*PgCGT6.6*), 5.29 (*PgCGT1.1*), and 16.91 (*PgCGT7.5*). Additionally, the genes *PgCGT1.5*, *PgCGT2.2*, *PgCGT2.4*, *PgCGT6.3*, *PgCGT6.8,*

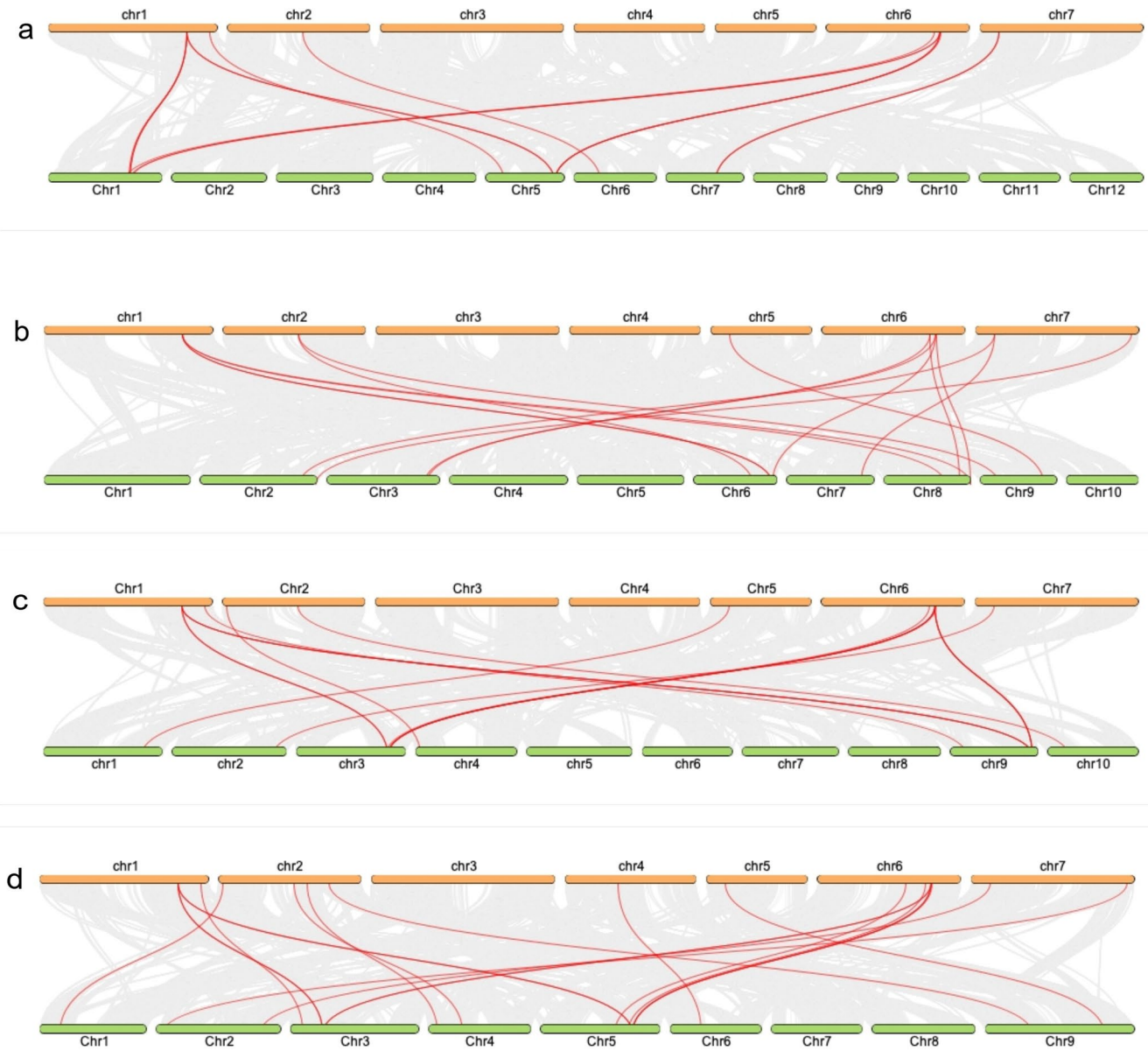

Fig. 4 Interspecific collinearity analysis of *PgCGT* genes between pearl millet and cereal genomes: (**a**) rice, (**b**) maize, (**c**) sorghum, and (**d**) foxtail millet. Chromosomal regions of pearl millet (top) and the corresponding chromosomes of rice, maize, sorghum and foxtail millet are shown as horizontal bars. Red lines indicate syntenic gene pairs, highlighting conserved genomic regions and evolutionary relationships of *CGT* genes across these species

PgCGT6.11 and *PgCGT6.12* were highly expressed on day 0 in the high-rancidity genotype, with minimal expression in the low-rancidity inbred. In contrast, *PgCGT7.6*, *PgCGT7.7*, *PgCGT1.4*, *PgCGT6.10*, and *PgCGT7.2* showed higher expression on day 0 in the low-rancidity genotype, suggesting their potential involvement during the early stages of grain storage. Furthermore, increased expression of *PgCGT6.1*, *PgCGT6.2*, *PgCGT6.4*, *PgCGT2.6*, *PgCGT7.1*, *PgCGT1.2*, and *PgCGT5.1* was observed in the high-rancidity genotype on day 10, indicating their possible association in lipid degradation and rancidity development during storage. No consistent expression patterns were observed among the remaining genes. These results highlight the potential role of *PgCGT* genes in determining grain quality and flour shelf life. Overall, this study provides evolutionary insights and unveils the

Table 3 Rapid peroxide value (RPV; mg ferric/kg) of contrasting pearl millet inbreds selected for gene expression analysis at different storage periods (day 0 and day 10 under accelerated storage conditions)

Inbred	RPV (mg Ferric/kg)		
	Day0	Day10	Day20
DS-SP-9	3.71	13.59	6.13
DS-SP-15	1.49	8.68	4.84

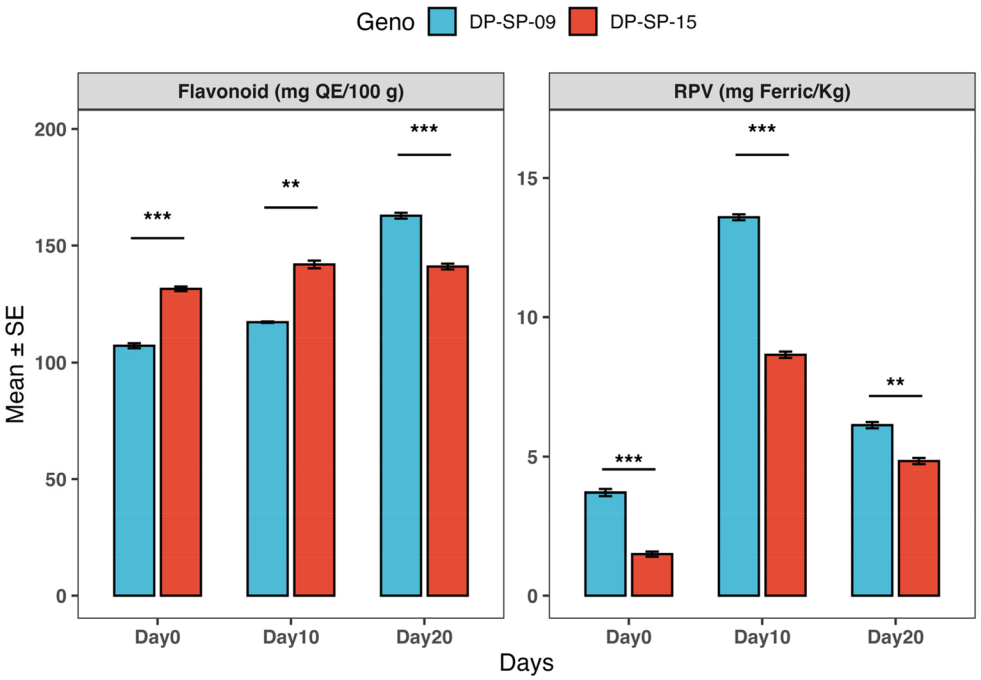

Fig. 5 Changes in total flavonoid content and peroxide value (RPV) in contrasting pearl millet inbreds during storage. Total flavonoid content (mg QE/100 g) and peroxide value (mg ferric/kg) were measured at Day 0, Day 10, and Day 20 in DS-SP-15 (low-rancid) and DS-SP-9 (high-rancid) genotypes stored under accelerated storage conditions. The left y-axis represents peroxide value, and the right y-axis represents total flavonoid content. Bars indicate mean values, and error bars represent standard error. The asterisks above the bars indicate statistically significant differences, where ** = $P < 0.01$ (more significant) and *** = $P < 0.001$ (highly significant)

functional diversity of the *CGT* gene family. The identified genes represent potential candidates for functional validation using transgenic approaches or gene editing.

Protein-protein interaction analysis

The protein–protein interaction (PPI) network consisted of 28 nodes and 38 edges, with an average node degree of 2.71 and a local clustering coefficient of 0.151. The observed number of interactions was significantly higher than expected for a random set of proteins (expected edges = 10), indicating strong functional connectivity among the proteins. The PPI enrichment analysis was highly significant ($p = 1.47 \times 10^{-11}$), suggesting that the proteins are biologically connected and likely involved in shared functional pathways. A gene interaction network analysis revealed that *PgCGT7.3*, *PgCGT7.6*, and *PgCGT7.7* functioned as major hub nodes with extensive associations to several rice homologs, including *OsI_24306*, *OsI_15049*, *OsI_15048*, *OsI_15050*, *OsI_07085*, *OsI_15051*, *OsI_24307*, *OsI_24308*, *OsI_24309*, and *OsI_17739*. Among these, *PgCGT7.3* exhibited the highest connectivity, suggesting a potentially important regulatory role within the network.

The functional enrichment analysis of the network revealed that the gene set was strongly associated with oxidative stress and reactive oxygen species (ROS) metabolism. Significant enrichment was observed for responses to oxidative stress, the hydrogen peroxide catabolic process, and peroxidase activity, indicating a key role in cellular redox homeostasis and in detoxifying reactive oxygen species. In parallel, a clear enrichment of flavonoid biosynthesis (KEGG: map00941) and related phenylpropanoid-derived pathways was observed, suggesting coordinated regulation of antioxidant secondary metabolite production (Fig. S3).

At the molecular function level, the network was predominantly enriched for UDP-glycosyltransferase and O-methyltransferase activities, supporting their roles in flavonoid modification, stabilization, and diversification. Peroxidase-related functions were also significantly enriched, supporting their role in hydrogen peroxide scavenging and oxidative stress regulation. STRING cluster analysis further revealed distinct functional modules associated with flavonoid metabolism, ROS detoxification, and secondary metabolite biosynthesis (Fig. S4). Collectively, these results indicate that the identified gene set forms a tightly connected functional network involved in flavonoid biosynthesis and oxidative stress regulation, particularly hydrogen peroxide detoxification in plants.

Discussion

Pearl millet is one of the most important crops, after major cereals such as rice, wheat, and maize, owing to its high micronutrient content (Govindaraj et al. 2020, 2022; Pujar et al. 2020). The International Year of Millets has renewed global interest in many millet crops, thereby increasing demand and consumption, particularly in the form of value-added products and processed forms. However, the short shelf life of pearl millet flour remains a major constraint on its wider utilization. Advances in genome sequencing technology have enabled the sequencing of complete genomes for several crops, including pearl millet, thereby providing new

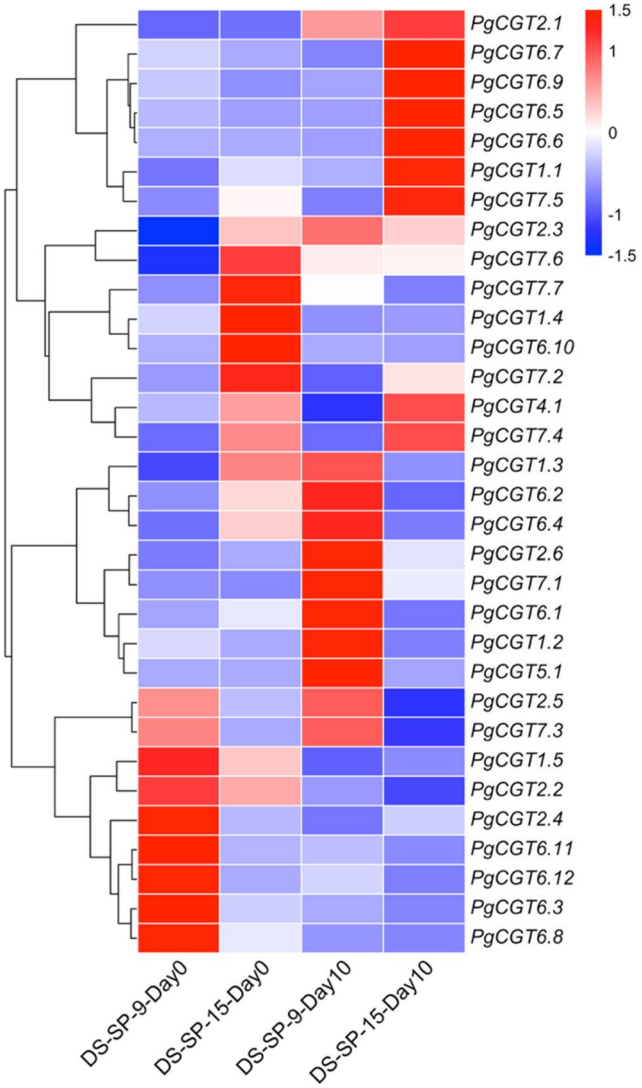

Fig. 6 Heat map showing the expression patterns of *PgCGT* genes in contrasting pearl millet inbreds during storage. Gene expression levels were analyzed in DS-SP-9 (high-rancid) and DS-SP-15 (low-rancid) at Day 0 and Day 10, which were stored under accelerated storage conditions. The color scale represents relative expression levels, where red indicates higher expression and blue indicates lower expression. Hierarchical clustering groups genes with similar expression patterns

opportunities for in silico characterization of gene families. C-glycosylflavonoids play a key role in regulating grain quality traits, particularly grain color and antioxidant activity in the embryo and endosperm, thereby improving flour shelf life and reducing rancidity (Kim et al. 2018). This is supported by evidence that flavonoids enhance antioxidant capacity and are positively associated with oxidative stability (Pushparaj and Urooj 2014). In addition to their role in grain quality, flavonoids have been implicated in plant defense responses, including downy mildew resistance in cucumbers (McNally et al. 2003). Furthermore, flavonoids contribute to drought tolerance in maize, and C-glycosylflavonoids may have specific roles in stress adaptation that warrant further investigation (Li et al. 2021). Therefore, characterization of *CGT* genes may help elucidate their additional functional roles in pearl millet. This highlights the importance of genome characterization of Pg*CGT* genes in pearl millet.

A genome-wide identification and characterization of glycosyltransferase genes in pearl millet identified 32 putative C-glycosyltransferase (CGT) genes in the present study. Glycosyl transferase gene family has been characterized in several crops, including cotton (Wu et al. 2019), groundnut (Ouyang et al. 2023), chickpea (Sharma et al. 2014), and sorghum (Rehana et al. 2025). However, these studies did not distinguish between C- and other glycosyl transferases. In this study, we classified genes using a phylogenetic analysis of C- and other glycosyl transferase genes, treating those that grouped with C-glycosyltransferases as *CGT* genes in pearl millet. Approximately 80% of the identified C-glycosyltransferase genes lacked introns. Previous research reported that approximately 58% of glycosyl transferase genes in rice (Dauda et al. 2022) and 55% in soybean (Mamoon Rehman et al. 2016) are intron-less. The low intron content observed in this gene family may be associated with the requirement for rapid transcriptional responses, involved in diverse regulatory roles.

The PSPG domain is recognized as the signature motif of the glycosyl transferase gene family and has been reported in other species (Yin et al. 2017; An et al. 2025). The presence of the PSPG domain in these sequences further supports their classification as members of the glycosyl transferase gene family. Additionally, multiple sequence alignment revealed both variable and conserved regions within the PSPG domain, thereby improving our understanding of the identified genes. Typically, UDP-glucose is the most common sugar donor for U-glycosyl transferases, although other sugar donors, such as UDP-galactose, glucuronic acid, and rhamnose, were also utilized (Gharabli and Welner 2024). The terminal amino acid of the PSPG domain sequence, typically glutamine (Q), is associated with sugar donor specificity because of its direct interaction with the sugar moiety and its higher affinity for UDP-glucose. However, some wild-type galactosyl transferases are also known to contain glutamine (Q) at the C-terminal end of the PSPG domain. Some point mutation studies have demonstrated that glutamine (Q) is strongly associated with UDP-glucose specificity and activity (Kubo et al. 2004; Louveau et al. 2018). Among the identified *PgCGT* genes in pearl millet, *PgCGT7.1* contained lysine instead of glutamine at the terminal amino acid of the PSPG domain. Further investigation is required to elucidate the functional significance of the point mutation within the PSPG domain. The presence of a 45-amino-acid PSPG domain, comprising 45 amino acids along with conserved regions in the gene *PgCGT6.1*, further supports its classification as a CGT gene and warrants further study.

Among the identified cis-acting regulatory elements, approximately 6% were associated with basic regulatory

functions, 33% with growth and development, 12.5% to biotic/abiotic stress responses, and nearly 48% with light-responsive elements. Several cis-elements were associated with hormonal signaling pathways, circadian rhythm and cell cycle regulation, seed-specific regulation, and stress-responsive functions. Apart from that, strong structural similarity among the identified *PgCGT* proteins (TM-align score > 0.8) together with high pLDDT confidence scores of AlphaFold-predicted 3D protein structures, indicates strong conservation and reliability of the predicted protein models. Synteny analysis revealed several collinear gene pairs between pearl millet and other cereals, including rice (15), sorghum (18), foxtail millet (20), and maize (15). All collinear gene pairs exhibited Ka/Ks values below 1, indicating purifying selection and functional conservation during evolution, although functional validation remains required.

C-Glycosylflavonoids possess stable C–C glycosidic bonds that confer resistance to enzymatic and acidic hydrolysis (Dorjjugder et al. 2021). The low-rancid genotype (DS-SP-15) identified based on RPV exhibited higher flavonoid levels during early storage (days 0 and 10) along with lower peroxide values, indicating stronger antioxidant buffering capacity against lipid peroxidation. Correspondingly, several *PgCGT* genes showed elevated expression in DS-SP-15 during these stages, suggesting their involvement in maintaining basal flavonoid levels and oxidative stability. In contrast, the high-rancid genotype (DS-SP-9) showed lower flavonoid accumulation during early storage and higher peroxide values, while the later increase in flavonoids may reflect a compensatory response to oxidative stress rather than a direct preventive mechanism against rancidity. A comparable stage-specific modulation of flavonoid biosynthesis has also been reported in metabolomic studies of pearl millet (Yogendra et al. 2024).

A previous genome-wide study in pearl millet identified 191 UDP-glycosyltransferase genes and reported developmental-stage-specific expression of selected genes during grain development, indicating roles in secondary metabolite biosynthesis and seed development (Kumar et al. 2025). In contrast, the present study focused specifically on the CGT subgroup and revealed genotype- and storage-stage-specific expression patterns associated with flavonoid accumulation and oxidative stability as a measure of peroxide accumulation, suggesting a specialized role in post-harvest grain quality regulation. Similarly, differential expression patterns of glycosyltransferase genes have been reported in other plant species. In rice, members of this gene family show transcriptional regulation associated with pathogen defense (Dauda et al. 2022), whereas in cotton they exhibit developmental-stage-specific expression during seed development (Cao et al. 2024). Likewise, glycosyltransferase genes in maize and tobacco exhibit stress-responsive expression under heat and salinity, respectively (Li et al. 2024; Sun et al. 2013). These observations, together with genotype- and stage-specific expression patterns of *PgCGT* genes in pearl millet, suggest that this gene family is transcriptionally regulated by developmental and environmental cues.

Several gene families, including phospholipases (Moin et al. 2024), TAG-lipases (Aher et al. 2022), and lipoxygenases (LOXs) (Palakolanu et al. 2026), play important roles in regulating rancidity development, and the present study contributes to a broader understanding of lipid deterioration mechanisms. Further elucidation of the roles of these identified genes in the metabolic pathways associated with lipid oxidation and antioxidant defense is essential for understanding their coordinated regulation during post-harvest storage. Comprehensive metabolomic profiling studies are further essential to identifying and quantifying specific C-glycosylflavonoid compounds regulated by *PgCGT* genes during storage. Functional validation through enzymatic assays and transgenic or reverse-genetic approaches will be an important future direction to establish the direct involvement of these genes in regulating lipid stability. Therefore, putative functional candidates identified in this study provide a valuable resource for future genome editing and molecular breeding efforts aimed at enhancing flour shelf life and reducing rancidity.

Protein–protein interaction enrichment suggests that these genes function as a coordinated network rather than as independent components. Their association with flavonoid biosynthesis, oxidative-stress responses, and hydrogen peroxide pathways supports a shared antioxidant function. Flavonoids can reduce oxidative damage by scavenging reactive oxygen species and limiting lipid peroxidation, and glycosyltransferase-mediated modification can further enhance flavonoid stability and activity. In particular, C-glycosylation yields more stable compounds, potentially extending antioxidant effects. Enrichment of peroxide-detoxification pathways also points to combined enzymatic and non-enzymatic redox control. Overall, CGT-mediated flavonoid metabolism likely supports oxidative stability and may influence peroxide accumulation and the shelf life of pearl millet flour.

Conclusion

A genome-wide analysis of the *CGT* gene family in pearl millet identified 32 PgCGT genes, characterized by their physicochemical properties, structures, motifs, and regulatory elements. These genes are distributed across all chromosomes except chromosome 3. Synteny analysis showed collinear pairs with rice, maize, sorghum, and foxtail millet, indicating conservation through purifying selection. Their conservation suggests their importance for grain quality and potential for breeding. Promoter analysis identified cis-regulatory elements associated with seed development

and endosperm metabolism. Expression profiling revealed stage- and genotype-specific patterns: some genes, such as *PgCGT7.6*, *PgCGT7.7*, and *PgCGT1.4*, showed higher expression in the low-rancid line at days 0 and *PgCGT2;1, PgCGT6.7, PgCGT6.9, PgCGT6.5, PgCGT6.6, PgCGT1.1* and *PgCGT7.5* at day10, possibly related to shelf-life. Others, such as *PgCGT2.4* and *PgCGT6.11*, showed higher expression in the high-rancid genotype, suggesting a link to rancidity, but further validation is needed. Overall, the study offers a resource for understanding the CGT family and identifying candidate genes for grain quality improvement through breeding, transgenics, and editing.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10142-026-01911-2.

Author contributions Kommineni Jagadeesh: Investigation, data curation, formal analysis, visualization, writing original draft preparation; Eswari. B. Kancharla: Conceptualization, writing original draft preparation, writing – review and editing; Vetriventhan mani : Conceptualization, writing original draft preparation, writing – review and editing; Kuldeep Singh : Supervision, writing – review and editing; Roopa Banerjee: Supervision, writing – review and editing; Kalyani Prasad: Formal analysis; Sudhakar Reddy Palakolanu: Conceptualization, methodology, supervision, writing – review and editing. All authors approved the final version.

Funding This study was undertaken as part of the ICRISAT Research Program on Accelerated Crop Improvement and the Consultative Group on International Agricultural Research (CGIAR) Genebanks Accelerator. J. K. acknowledges fellowship support from the Council of Scientific and Industrial Research in the form of a JRF and SRF fellowship (08/1285(15150)/2022-EMR-I).

Data availability All required data are provided as part of the manuscript. Researchers can contact the corresponding author for any additional information required.

Declarations

Competing interests The authors declare no competing interests.

References

Aher RR, Reddy PS, Bhunia RK et al (2022) Loss-of-function of triacylglycerol lipases are associated with low flour rancidity in pearl millet [*Pennisetum glaucum* (L.) R. Br]. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.962667

Akingbala' J (1991) Effect of Processing on Flavonoids in Millet (*Pennisetum americanum*) Flour. Cereal Chem 68(2):0

Ali A, Kumar RR, Vinutha T et al (2023) Characterization of biochemical indicators and metabolites linked with rancidity and browning of pearl millet flour during storage. J Plant Biochem Biotechnol 32(1):121–131. https://doi.org/10.1007/s13562-022-00787-0

An J, He Q, Xi J et al (2025) Genome-wide identification of the *UGT* gene family in poplar *Populus euphratica* and functional analysis of PeUGT110 under drought stress. Forests 16(8). https://doi.org/10.3390/f16081214

Bhargavi HA, Singh SP, Goswami S et al (2024) Deciphering the genetic variability for biochemical parameters influencing rancidity of pearl millet (*Pennisetum glaucum* LR Br.) flour in a set of highly diverse lines and their categorization using rancidity matrix. J Food Compos Anal 128:106035

Boncompagni E, Orozco-Arroyo G, Cominelli E et al (2018) Antinutritional factors in pearl millet grains: phytate and goitrogens content variability and molecular characterization of genes involved in their pathways. PLoS ONE 13(6). https://doi.org/10.1371/journal.pone.0198394

Cao Y, Han Z, Zhang Z et al (2024) UDP-glucosyltransferase 71C4 controls the flux of phenylpropanoid metabolism to shape cotton seed development. Plant Commun 5(8). https://doi.org/10.1016/j.xplc.2024.100938

Chen D, Chen R, Wang R et al (2015) Probing the Catalytic Promiscuity of a Regio- and Stereospecific C-Glycosyltransferase from Mangifera indica. Angew Chem 127:12869–12873. https://doi.org/10.1002/ange.201506505

Chen C, Chen H, Zhang Y et al (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13(8):1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Dauda WP, Shanmugam V, Tyagi A et al (2022) Genome-Wide Identification and Characterisation of Cytokinin-O-Glucosyltransferase (*CGT*) Genes of Rice Specific to Potential Pathogens. Plants 11(7):917. https://doi.org/10.3390/plants11070917

Dorjjugder N, Hatano M, Taguchi G (2021) Production of flavonol and flavone 6-C-glucosides by bioconversion in *Escherichia coli* expressing a C-glucosyltransferase from wasabi (*Eutrema japonicum*). Biotechnol Lett 43(9):1913–1919. https://doi.org/10.1007/s10529-021-03165-3

Ferreyra MLF, Rodriguez E, Casas MI et al (2013) Identification of a bifunctional Maize C- and O-glucosyltransferase. J Biol Chem 288(44):31678–31688. https://doi.org/10.1074/jbc.M113.510040

Gharabli H and Welner DH (2024) The sugar donor specificity of plant family 1 glycosyltransferases. Front Bioeng Biotechnol 12:1396268. https://doi.org/10.3389/fbioe.2024.1396268

Ghimire BK, Seo JW, Yu CY et al (2021) Comparative study on seed characteristics, antioxidant activity, and total phenolic and flavonoid contents in accessions of *sorghum bicolor* (L.) moench. Molecules 26(13):3694. https://doi.org/10.3390/molecules26133964

Goswami S, Asrani P, Ansheef Ali TP et al (2020) Rancidity Matrix: Development of Biochemical Indicators for Analysing the Keeping Quality of Pearl Millet Flour. Food Anal Methods 13(11):2147–2164. https://doi.org/10.1007/s12161-020-01831-2/Published

Govindaraj M, Rai KN, Kanatti A et al (2020) Exploring the genetic variability and diversity of pearl millet core collection germplasm for grain nutritional traits improvement. Sci Rep 2020 10:121177. https://doi.org/10.1038/s41598-020-77818-0

Govindaraj M, Kanatti A, Rai KN et al (2022) Association of grain iron and zinc content with other nutrients in pearl millet germplasm, breeding lines, and hybrids. Front Nutr 8. https://doi.org/10.3389/fnut.2021.746625

Isla MI, Salas A, Danert FC et al (2014) Analytical methodology optimization to estimate the content of non-flavonoid phenolic compounds in Argentine propolis extracts. Pharm Biol 52(7):835–840. https://doi.org/10.3109/13880209.2013.871638

Ito T, Fujimoto S, Suito F et al (2017) C-Glycosyltransferases catalyzing the formation of di-C-glucosyl flavonoids in citrus plants. Plant J 91(2):187–198. https://doi.org/10.1111/tpj.13555

Kanatti A, Rai KN, Radhika K et al (2014) Grain iron and zinc density in pearl millet: combining ability, heterosis and association with grain yield and grain size. Springerplus 3(1):763. https://doi.org/10.1186/2193-1801-3-763

Kaur R, Kaur K (2016) Appraisal of maize germplasm for identification of nutritionally rich nutraceutical genotypes. Cereal Res Commun 44(3):501–512. https://doi.org/10.1556/0806.44.2016.008

Kim B, Woo S, Kim MJ et al (2018) Identification and quantification of flavonoids in yellow grain mutant of rice (*Oryza sativa* L). Food Chem 241:154–162. https://doi.org/10.1016/j.foodchem.2017.08.089

Krishnan R, Meera MS (2017) Assessment of inhibitory factors on bioaccessibility of iron and zinc in pearl millet (*Pennisetum glaucum* (L.) R. Br.) cultivars. J Food Sci Technol 54(13):4378–4386. https://doi.org/10.1007/s13197-017-2911-2

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 429(2):198–203. https://doi.org/10.1016/j.abb.2004.06.021

Kumar A, Tun TT, Kumar V (2025) Genome-wide identification and expression analysis of UDP-glycosyltransferases genes associated with secondary metabolism during grain development in pearl millet (*Pennisetum glaucum*). Plant Gene 44:100541. https://doi.org/10.1016/j.plgene.2025.100541

Li B, Fan R, Sun G et al (2021) Flavonoids improve drought tolerance of maize seedlings by regulating the homeostasis of reactive oxygen species. Plant Soil 461(1):389–405. https://doi.org/10.1007/s11104-020-04814-8

Li F, Fan K, Guo X et al (2022) Genome-wide identification, molecular evolution and expression analysis of the non-specific lipid transfer protein (nsLTP) family in *Setaria italica*. BMC Plant Biol 22(1):547. https://doi.org/10.1186/s12870-022-03921-1

Li H, Wang J, Li M et al (2024) The *ZmHSF08-ZmUGT92A1* module regulates heat tolerance by altering reactive oxygen species levels in maize. Crop J 12(5):1437–1446. https://doi.org/10.1016/j.cj.2024.09.007

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. Methods 25(4):402–408. https://doi.org/10.1006/meth.2001.1262

Louveau T, Orme A, Pfalzgraf H et al (2018) Analysis of two new arabinosyltransferases belonging to the carbohydrate-active enzyme (CAZY) glycosyl transferase family1 provides insights into disease resistance and sugar donor specificity. Plant Cell 30(12):3038–3057. https://doi.org/10.1105/tpc.18.00641

Mamoon Rehman H, Amjad Nawaz M, Bao L et al (2016) Genome-wide analysis of Family-1 UDP-glycosyltransferases in soybean confirms their abundance and varied expression during seed development. J Plant Physiol 206:87–97. https://doi.org/10.1016/j.jplph.2016.08.017

McNally DJ, Wurms KV, Labbé C, Bélanger RR (2003) Synthesis of C-glycosylflavonoid phytoalexins as a site-specific response to fungal penetration in cucumber. Physiol Mol Plant Pathol 63(6):293–303. https://doi.org/10.1016/j.pmpp.2004.03.005

Moin M, Bommineni PR, Tyagi W (2024) Exploration of the pearl millet phospholipase gene family to identify potential candidates for grain quality traits. BMC Genomics 25(1):581. https://doi.org/10.1186/s12864-024-10504-x

Nagatomo Y, Usui S, Ito T et al (2014) Purification, molecular cloning and functional characterization of flavonoid C-glucosyltransferases from *Fagopyrum esculentum* M. (buckwheat) cotyledon. Plant J 80(3):437–448. https://doi.org/10.1111/tpj.12645

Ouyang L, Liu Y, Yao R et al (2023) Genome-wide analysis of UDP-glycosyltransferase gene family and identification of a flavonoid 7-O-UGT (*AhUGT75A*) enhancing abiotic stress in peanut (*Arachis hypogaea* L). BMC Plant Biol 23(1):626. https://doi.org/10.1186/s12870-023-04656-3

Palakolanu SR, Bommineni PR, Tyagi W (2026) Genome-Wide Identification and Expression Analysis of Lipoxygenase *(LOX)* Gene Family in Pearl Millet (*Pennisetum Glaucum* L). Plant Mol Biol Rep 44(1):7. https://doi.org/10.1007/s11105-025-01634-x

Pujar M, Gangaprasad S, Govindaraj M et al (2020) Genome-wide association study uncovers genomic regions associated with grain iron, zinc and protein content in pearl millet. Sci Rep 10(1):19473. https://doi.org/10.1038/s41598-020-76230-y

Punia S, Sandhu KS, Siroha AK (2019) Difference in protein content of wheat (*Triticum aestivum* L.): Effect on functional, pasting, color and antioxidant properties. J Saudi Soc Agricultural Sci 18(4):378–384. https://doi.org/10.1016/j.jssas.2017.12.005

Pushparaj FS, Urooj A (2014) Antioxidant activity in two pearl millet (*Pennisetum typhoideum*) cultivars as influenced by processing. Antioxidants 3(1):55–66. https://doi.org/10.3390/antiox3010055

Reddy PS, Reddy DS, Sharma KK et al (2015) Cloning and validation of reference genes for normalization of gene expression studies in pearl millet [*Pennisetum glaucum* (L.) R. Br.] by quantitative real-time PCR. Plant Gene 1:35–42. https://doi.org/10.1016/j.plgene.2015.02.001

Rehana R, Anwar M, Arshad SF et al (2025) Genome-Wide Identification and Functional Characterization of the Glycosyltransferase 43 (*GT43*) Gene Family in Sorghum bicolor for Biofuel Development: A Comprehensive Study. Processes 13(3):709. https://doi.org/10.3390/pr13030709

Satyavathi CT, Ambawat S, Khandelwal V, Srivastava RK (2021) Pearl Millet: A Climate-Resilient Nutricereal for Mitigating Hidden Hunger and Provide Nutritional Security. Front Plant Sci 12:659938. https://doi.org/10.3389/FPLS.2021.659938/BIBTEX

Satyavathi CT, Ambawat S, Thirunavukkarasu N et al (2024) Global production, status, and utilization pattern of pearl millet. In: Pearl Millet: A Resilient Cereal Crop for Food, Nutrition, and Climate Security. Wiley Online Library, pp 1–21

Sharma R, Rawat V, Suresh CG (2014) Genome-wide identification and tissue-specific expression analysis of UDP-glycosyltransferases genes confirm their abundance in *Cicer arietinum* (Chickpea) genome. PLoS ONE 9(10):e109715. https://doi.org/10.1371/journal.pone.0109715

Sharma B, Chugh L, Singh VK et al (2022) Effect of storage on fat hydrolysis system in pearl millet flour. In: AIP Conference Proceedings. AIP Publishing

Shen Y, Jin L, Xiao P et al (2009) Total phenolics, flavonoids, antioxidant capacity in rice grain and their relations to grain color, size and weight. J Cereal Sci 49(1):106–111. https://doi.org/10.1016/j.jcs.2008.07.010

Singh S, Yadav D, Beckmann M et al (2024) Variation in protein and amino acids in global collection of pearl millet (*Pennisetum glaucum*) germplasm. J Food Compos Anal 134. https://doi.org/10.1016/j.jfca.2024.106557

Sun YG, Wang B, Jin SH et al (2013) Ectopic Expression of Arabidopsis Glycosyltransferase UGT85A5 Enhances Salt Stress Tolerance in Tobacco. PLoS ONE 8(3):e59924. https://doi.org/10.1371/journal.pone.0059924

Tomar M, Bhardwaj R, Kumar M et al (2021) Nutritional composition patterns and application of multivariate analysis to evaluate indigenous Pearl millet ((*Pennisetum glaucum* (L.) R. Br.) germplasm. J Food Compos Anal 103:104086. https://doi.org/10.1016/j.jfca.2021.104086

Wu A, Hao P, Wei H et al (2019) Genome-Wide Identification and Characterization of Glycosyltransferase Family 47 in Cotton. Front Genet 10:824. https://doi.org/10.3389/fgene.2019.00824

Yin Q, Shen G, Chang Z et al (2017) Involvement of three putative glucosyltransferases from the *UGT72* family in flavonol glucoside/rhamnoside biosynthesis in Lotus japonicus seeds. J Exp Bot 68(3):597–612. https://doi.org/10.1093/jxb/erw420

Yogendra K, Sanivarapu H, Avuthu T et al (2024) Comparative Metabolomics to Unravel the Biochemical Mechanism Associated with Rancidity in Pearl Millet (*Pennisetum glaucum* L). Int J Mol Sci 25(21):11583. https://doi.org/10.3390/ijms252111583