PLANT SCIENCE TODAY
ISSN 2348-1900 (online)
Vol x(x): xx–xx
https://doi.org/10.14719/pst.3285

HORIZON
e-Publishing Group
HePG

RESEARCH ARTICLE

Phylogenomic analysis of glutamine synthetase gene family in *Helianthus annuus* L.

Vrinda Singh[1], Vivek Kumar[1], Mohd Razaq[1], Pooja Jain[1], Deepti Teotia[1], Aman Agrawal[1], Mukesh Kumar Bhardwaj[2], Hemant Sharma[3], Pranita Malik[4], Sachin Kumar[1] & Vijai Malik[1]*

[1]Department of Botany, Chaudhary Charan Singh University, Meerut, UP-250004, India
[2]Department of Botany, Dharma Samaj College, Aligarh, UP-202001, India
[3]Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, UP-250004, India
[4]Department of Botany, Maharaj Singh College, Saharanpur, UP-247001, India

*Email: gathwalajai@gmail.com

OPEN ACCESS

ARTICLE HISTORY

Received: 15 January 2024
Accepted: 30 June 2024

Available online
Version 1.0 : 24 October 2024

Check for updates

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc See https://horizonepublishing.com/journals/index.php/PST/indexing_abstracting

CITE THIS ARTICLE

Singh V, Kumar V, Razaq M, Jain P, Teotia D, Agrawal A, Bhardwaj M K, Sharma H, Malik P, Kumar S, Malik V. Phylogenomic analysis of glutamine synthetase gene family in *Helianthus annuus* L. . Plant Science Today (Early Access). https://doi.org/10.14719/pst.3285

Abstract

Glutamine synthetase is one of the predominant enzymes in nitrogen metabolism. In developing leaves, glutamine is mainly produced in chloroplasts by the activity of the GS2 isoenzyme. It catalyzes glutamine synthesis from glutamate and ammonia in an ATP-dependent reaction. The genes encoding glutamine synthetase play a crucial role in ammonia and glutamate detoxification, acid-base homeostasis, cell signaling, and cell proliferation. The gene family responsible for producing glutamine synthetase has been previously documented in model plants like *Arabidopsis*. Nevertheless, there has been no exploration into its existence and attributes in oilseed crops like sunflower (*Helianthus annuus* L.). This study thoroughly analyzes, gene structure, conserved motifs, chromosomal location, phylogenetic relationships, and expression patterns to identify the glutamine synthetase genes in *H. annuus*. Our findings unveiled 19 genes encoding glutamine synthetase within the *H. annuus* genome. These genes were distributed across 11 chromosomes of *H. annuus*. Furthermore, we examined the expression patterns of all the HaGS genes using RNA-seq datasets, specifically focusing on their response to biotic and abiotic stress conditions. Under biotic stress, *H. annuus* expresses genes for mycorrhizal fungi named *Rhizoglomus irregulare* at four days post inflorescence (dpi) and 16 dpi. Under abiotic stress, the effect of drought and hormones was investigated. In drought, one gene, HaGS6D, showed the highest expression in the leaf. Meanwhile, in roots, gene HaGS7B showed the highest expression. Under hormonal stress, the effects of auxin, brassinosteroid, and cytokinin were studied on the leaf. For auxin, the gene HaGS7C showed the highest expression. For brassinosteroid, the gene HaGS7D showed the highest expression; for cytokinin, the gene HaGS6E showed the highest expression. Thus, these findings can significantly contribute to our understanding of the arrangement of glutamine synthetase genes in *H. annuus* and offer valuable insights for developing of drought-resistant cultivars of this species.

Keywords

Glutamine synthetase; *Helianthus annuus*; gene expression; abiotic stress; biotic stress

Introduction

The Asteraceae are one of the largest flowering plant families. It has over 1600 genera and 2500 species worldwide (1). With the exception of

Antarctica, this family is found in various ecological habitats. Recent research indicates 193 genera and 1172 taxa, consisting of 1021 species, 27 subspecies, 117 varieties, and 07 forms in India (2). There are 96 cultivated taxa of Asteraceae in India (2). The chief hub of the diversity of the Asteraceae in India is the Himalayan biogeographic zone (3). The distribution of taxa of Asteraceae in the Indian region enables the classification of 12 phytogeographical zones (3). There are over 50 species in the genus *Helianthus,* most of which are endemic to North America. In Europe and other parts of the world, some species—most notably *H. annuus* and *H. tuberosus*—are grown for their edible qualities as well as their aesthetic value. Sunflower (*H. annuus* L.) is an important oilseed crop (4). The seeds of sunflower produce high-quality edible oil that is low in cholesterol and is equal in quality to olive oil. Sunflower seeds are small, but they have a good amount of healthy unsaturated fats, fiber, protein, and various essential nutrients like vitamin E, copper, selenium, folate, iron, zinc, and phytochemicals.

The foremost among plant nutrients, nitrogen is a crucial factor that markedly constrains plant productivity (5). Plants can exclusively assimilate nitrogen in the NH^{4+} ion (ammonia) form into organic compounds (6). As an indispensable nutrient, nitrogen significantly influences the quantity of protein stored in cereal grains. The process of accumulating storage proteins in seeds involves the remobilization of nitrogen from vegetative organs facilitated by the GS and GDH enzymes (5). Senescent leaves store nitrogen-rich amino acids for exportation. Young leaves have a higher concentration of ammonia due to increased nitrate reduction during photosynthesis and photorespiration. Additionally, these leaves have lower levels of soluble carbohydrates, which are known to limit the assimilation of ammonia (7). In higher plants, glutamine synthetase and glutamate dehydrogenase are the main enzymes involved in nitrogen metabolism (5). GS, a vital enzyme for plant autotrophy, plays a key role in incorporating ammonia into glutamate to produce glutamine while simultaneously hydrolyzing ATP. This enzymatic process is crucial for ammonia assimilation and re-assimilation during various stages of plant growth and development (8). GS is found in plants in two different isoforms, GS1 and GS2, which are categorized according to where they are found in the cell (9, 10). These isoforms, encoded by a small family of nuclear genes, undergo regulation in expression based on light and nitrogen availability in a development- and tissue-specific manner (11–15). GS2, the chloroplastic isoform is more susceptible to oxidative modification as compared to GS1 (16). There is a decrease in GS activity during leaf senescence because there is a notable drop in GS2 transcripts and enzyme activity. The term “phylogenomics” may be used to predict gene function using large-scale genomic data (17, 18). The aim of this study is to analyse the roles of glutamine synthetase gene family in *H. annuus* L. using phylogenomic approach under stress condition to facilitate development of stress tolerance cultivars.

Methodology

Analysis of gene sequence

Sequence and identification of glutamine synthetase gene family in *H. annuus*

H. annuus and *Arabidopsis thaliana* sequences of the glutamine synthetase gene are available in the database. The gene sequences for glutamine synthetase in *H. annuus* have been obtained from the Ensembl database. Additionally, the gene sequences of *Arabidopsis* were employed independently for a TblastX search against the available *H. annuus* genome assembly.

Characteristic features of the glutamine synthetase gene family

The structure of the gene was analyzed using the Ensembl plants database (https://plants.ensembl.org/Helianthus_annuus/Info/Index), meme suits, and TBtool. Repeat masker Version: open-4.0.9 has been utilized to identify SSR in gene sequences. All these were performed using default and modified parameters. Plant care (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) has been utilized to search the presence of cis-regulatory elements in gene sequence upstream of 1500 bp of the promoter region. A web-based server ps-RNA (https://www.zhaolab.org/psRNATarget/) target has been utilized with a default parameter to search putative micro-RNA and their targets in the gene of *H. annuus*. Here, a 0-3 e-value was used (19–21).

Evaluation of synteny, collinearity, and gene duplication

The Ensembl plant pipeline for gene tree (22, 23) was employed to elucidate the evolutionary relationships among glutamine synthetase genes. This involved using gene identifiers for each gene in *H. annuus*. Using the plant compara option, a gene tree containing homologs from the genomes of *H. annuus* and *A. thaliana* was created. This gene tree serves as a tool for identifying duplication and speciation events. Additionally, the synteny/collinearity between *H. annuus* and *A. thaliana* genes was determined. The Genomicus program version 49.01 was utilized to ascertain collinearity and synteny levels.

Peptide sequences analysis

Physical & chemical properties, conserved motifs, secondary structure in HaGS protein sequence

The Conserved Domain Databases (CDD) program at NCBI offers CD-search services. The primary domains in the peptide sequences of *H. annuus* were determined using this program. Expasy's Protparam tool was utilized to calculate the physical & chemical parameters. Similarly, the SOPMA program was utilized to compute the properties of protein at the secondary level. Meme Suite was used to find motifs. The annotation of the discovered patterns was done using the Inter Proscan database.

3D structure analysis (assessment and optimization)

Homology modeling was used to estimate the expected protein's three-dimensional structure. Swiss model template library was carried out against Pdb. The geometric

and energetic verification of the predicted 3D structure of a protein was done using the structure analysis and validation system. Saves v6.0 was used to find out the relative proportion of amino acids falling in the favoured region in comparison to another region (24). It also analysed the compatibility of an atomic model (3D) with its amino acid sequence (25) and the statistics of non-bonded interactions between different atom types (26).

Determination of superimposition, functional annotation of 3D structures, sub-cellular localization, and gene ontology (GO)

The 3D structures of the anticipated *H. annuus* protein and the protein encoded by different *A. thaliana* genes were compared using the Fat Cat Server. A globally optimized superimposition environment was used to compare the root mean square deviation (RMSD) value of the Cα atoms in the created structures with those of the corresponding 3-dimensional structures of the query genes in order to determine the identification of the generated 3-dimensional structures. g:profiler was used to analyze sub-cellular localization and gene ontology (GO). Using a string database, the protein-protein interaction (PPI) network was carried out to determine the unknown molecular roles of proteins (27).

Multiple sequence alignment and phylogenetic analysis

The multiple sequence alignment was utilized to analyze the conserved and coevolved amino acid residues of *H. annuus* and *A. thaliana* using Multalin software. The mutual information (MI) between two amino acid locations in multiple sequence alignment (MSA) was analyzed using the Mistic webserver. MI was used to identify coevolving residue. In a MSA, the MI measures the link between two places, especially two columns. It indicates how much information about the amino acid identity at one position can help in the prediction of the amino acid identity at another position. MI is a useful metric for finding compensatory and correlated mutations in related proteins. Using the Clustal Omega database, the amino acid sequences (aa) of the peptides were examined in order to conduct the phylogenetic tree analysis. The Newick format tree was edited using iTOL.

In silico expression profiling of glutamine synthetase gene

Transcriptome data for *H. annuus* was utilized from the Genevestigator database (https://genevestigator.com/) for *in silico* expression analysis. Heat maps of tissue-specific expression under abiotic and biotic were studied. Heat maps of perturbations showed the responses of genes during stress.

Results

Analysis of gene sequence

Identification and structure of *H. annuus* glutamine synthetase gene

A comprehensive search of the entire genome was conducted, resulting in the retrieval of 19 gene sequences. Among them, 6 sequences were found to be incomplete. Consequently, 13 glutamine synthetase genes with full-length sequences were identified and subjected to further study. Table 1 contains comprehensive information about the genes, cDNA, and CDS sequences of HaGS1-7. Information on the homology of *H. annuus* with *A. thaliana* is given in Table 1. Based on homologous genes found in *A. thaliana*, all 19 genes of *Helianthus* have been designated (28). The length of individual HaGS genes ranges from 390 – 15,549 bp. The number of exons and introns varies from 2 –17 and 1–16, respectively. All 19 HaGS genes have intron. The cDNA sequence varies from 333–2092 bp (Table 1). Variations in the coding sequence (CDS) are from 259–2484 bp (Table 1).

The structural characteristics of the glutamine synthetase enzyme in *H. annuus* were examined, focusing on the intron-exon features. This analysis yielded insights into their structural patterns. Exons are shown as solid yellow bars, introns as black lines, upstream and downstream areas as solid blue bars, and UTR or 3' – 5' regions as solid green bars. Additionally, the distribution of intron phases was observed, with 67.24% in Phase 0, 27.58% in Phase 1, and 5.17% in Phase 2 (Fig. 1).

Gene duplication and chromosome assignment

Eleven distinct chromosomes were attributed to the 19 HaGS genes. Only 11 chromosomes of *H. annuus* were found to contain genes. Chr16 is characterized by a maximum number of 5 genes. The minimum number of 1 gene was located on Chr05, Chr06, Chr09, Chr10, Chr12, and Chr17 (Fig. 2). Using gene trees that included the *Helianthus* gene in addition to genes from other plant species, orthology and paralogy among 19 HaGS genes were investigated. This tree was prepared using an ensembl plant compara pipeline (Supplementary Fig. 1).

Supplementary Table 1 provides detailed information on homologs, including orthologous and paralogous relationships of the HaGS gene with various taxa. An arrangement of 7 genes in *A. thaliana* has been used to study orthologs and paralogs among the *H. annuus* glutamine synthetase gene. *H. annuus* was likewise projected to follow the same duplication pattern as *A. thaliana*. Using plant compara gene tree in an ensembl plant database has distinguished orthologues from paralogues (Supplementary Table 1). Two orthologs were found in 7 genes of *H. annuus*. These genes are HaGS7E, HaGS7B, HaGS7A, HaGS7F, HaGS7C, HaGS3A and HaGS7D.

Synteny between *H. annuus* and *A. thaliana*

Synteny analysis of 19 genes containing HaGS genes with the corresponding *A. thaliana* gene is 100% (Fig. 3). However, the co-linearity was not found with any gene of *H. annuus*.Two gene pairs have been found in the peptides of *H. annuus* & *A. thaliana*. Ka and Ks ratios for all the two pairs were found to be less than one. These values indicate stabilizing selection, i.e., genes are constrained to maintain their current function and thus act against change, favoring conservation (Supplementary Table 2).

Table 1. Gene, cDNA (complementary DNA), and CDS (coding sequence) sequences for glutamine synthetase (GS) in *H. annuus* (Ha) and *A. thaliana* (At) (size in bp).

H. annuus				*A. thaliana*			
Gene name	**Size of gene**	**Size of cDNA**	**Size of CDS**	**Gene name**	**Size of gene**	**Size of cDNA**	**Size of CDS**
HaGS1	7069	2092	1638	AtGS1	2815	2036	1467
HaGS2	1999	508	360	AtGS2	2589	1279	819
HaGS3A	678	393	393	AtGS3	2395	1554	1065
HaGS3B	1889	711	711				
HaGS4	8696	2979	2484	AtGS4	4641	3079	2559
HaGS5A	4827	2172	1572	AtGS5	5561	2383	1569
HaGS5B	15549	2080	1554				
HaGS6A	4172	2090	1293	AtGS6	3114	1885	1293
HaGS6B	4172	2090	1293				
HaGS6C	556	351	351				
HaGS6D	1376	454	259				
HaGS6E	390	333	333				
HaGS6F	1203	398	378				
HaGS7A	3043	1743	1074	AtGS7	2600	1588	1071
HaGS7B	3475	1533	1074				
HaGS7C	5200	1472	1158				
HaGS7D	4449	1641	1065				
HaGS7E	390	306	306				
HaGS7F	3515	1319	1071				

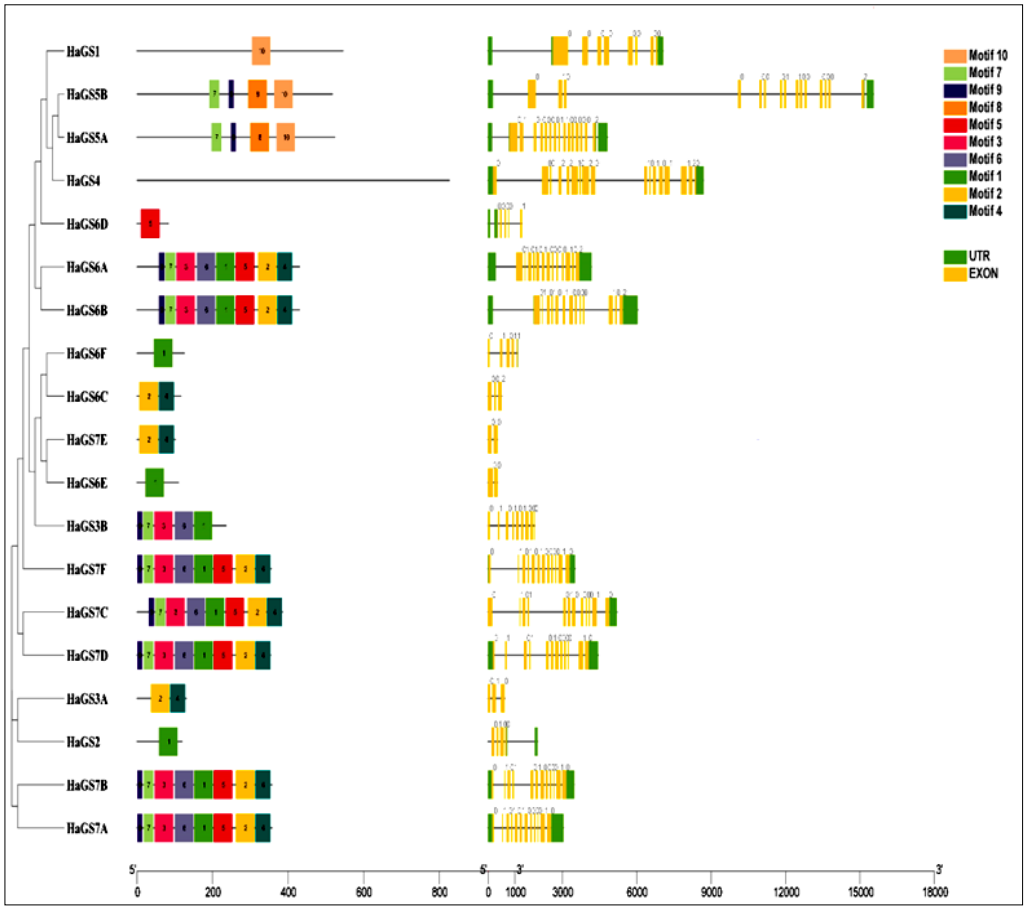

Fig. 1. Depiction of the HaGs genes in *H. annuus*, illustrating the arrangement of **exons** (solid yellow bars), **introns** (black lines), **upstream/downstream regions** (solid green bars), and intron phases denoted as 0, 1, and 2. The figure highlights the conserved motifs identified in HaGS proteins.

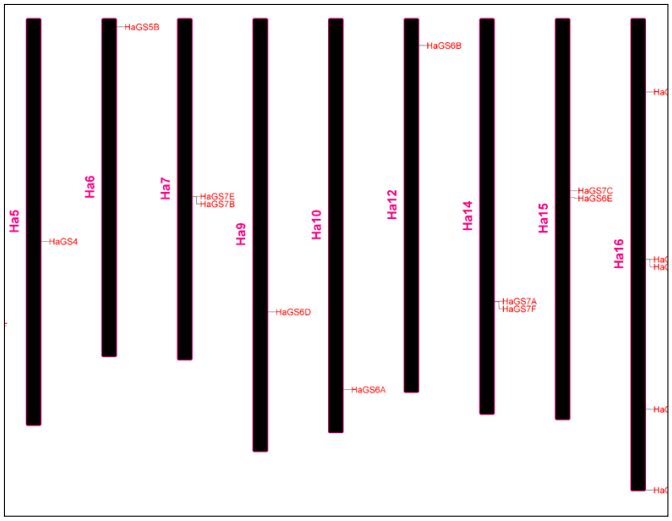

Fig. 2. Distribution of 19 HaGS genes on 11 chromosomes. On each chromosome, gene names are given on the upper side and their physical positions in megabases (Mb) are indicated on the left.

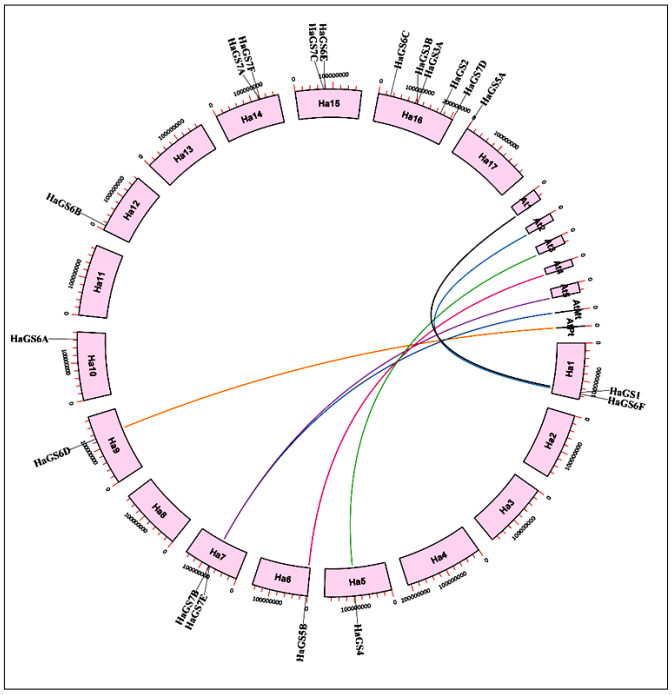

Fig. 3. A map depicting synteny among HaGS genes of *H. annuus* and *A. thaliana*.

SSR in HaGS genes

Out of the 19 genes, 51 SSRs were identified in 16 genes (84.21%). There is variation in number of SSR per gene, with 9 SSR in one gene (HaGS5B), 6 SSR in one gene (HaGS7B), 5 SSR in two genes (HaGS1 and HaGS7D), 4 SSR in two genes (HaGS6A and HaGS6B), 3 SSR in three genes (HaGS4, HaGS7A, and HaGS5A), 2 SSR in two genes (HaGS7C and HaGS7F) and 1 SSR in five genes (HaGS2, HaGS3B, HaGS6D, HaGS6E, and HaGS6F). The SSRs with mononucleotide motif (16) are most common, followed by pentanucleotide motifs (12), septanucleotide (5), trinucleotide (5), dinucleotide (4), tetranucleotide (4) and hexanucleotide (4) (Supplementary Table 3).

Analysis of promoter and related sequences

In each of the 19 HaGS genes, cis-regulatory elements were analyzed. The analysis was done at 1500 bp 5' upstream of the promoter sequence. The promoter region of all the genes contains many TCCC-motif, G-box, ARE, GT1-motif, TCA-element, LTR, G-box, O2-site, ABRE, TCT-motif, GATA-motif, CCAAT-box, CGTCA-motif, P-box, I -box, MBSI, TGA-element, MBS, TGACG-motif, HD-Zip 1, ACE, chs-CMA1a, AuxRR-core, Box 4, AE-box, Sp1, 3-AF1 binding site, CAT-box, RY-element, Unnamed_1, TATC-box, TC-rich repeats, Box II, MRE, 4cl-CMA1b, chs-CMA2a, GC-motif, circadian, GARE-motif, AT1-motif elements were present in a maximum number of genes (Supplementary Fig. 2).

Micro RNA and their targets in HaGS genes

All 19 genes of *H. annuus* were searched for mRNA. It has been observed that only a single gene, HaGS7A, had mRNA (Supplementary Table 4).

In silico expression analysis of HaGS Genes

The expression of the gene was examined in normal and stress conditions. In normal conditions, leaf development, inflorescence emergence, and flowering stages were studied (Fig. 4 and 5). Under stress conditions, we examined abiotic and biotic stress.

Tissue-specific expression

It was observed that under five organs or conditions (roots, leaf, shoot, inflorescence, and cell culture), tissue-specific expression shows variation in expression (Fig. 4) during three different developmental stages.

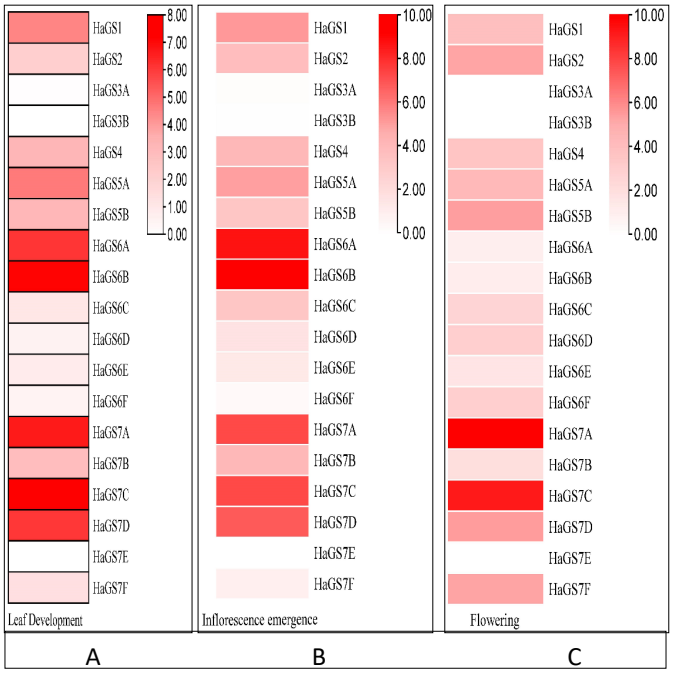

Fig. 4. HaGS gene expression profiles across various developmental stages, (**A**) Leaf development, (**B**) Inflorescence emergence, and (**C**) Flowering under normal conditions in *H. annuus*.

Abiotic stresses

Drought stress

Upregulation of two genes (HaGS6D & HaGS7B) were identified under stress conditions in *H. annuus*. Stress genes were found in the leaf and root of *H. annuus*. One gene (HaGS6D) showed the highest expression in the leaf at 1.82. Three genes in the leaf exhibited medium expression, i.e., 1.67-1.79 (HaGS6A, HaGS6C, and HaGS6E). The remaining genes showed very poor expression, i.e., 1-1.47. This suggests that the expression of these genes is very poor against drought stress. One gene (HaGS7B) in the

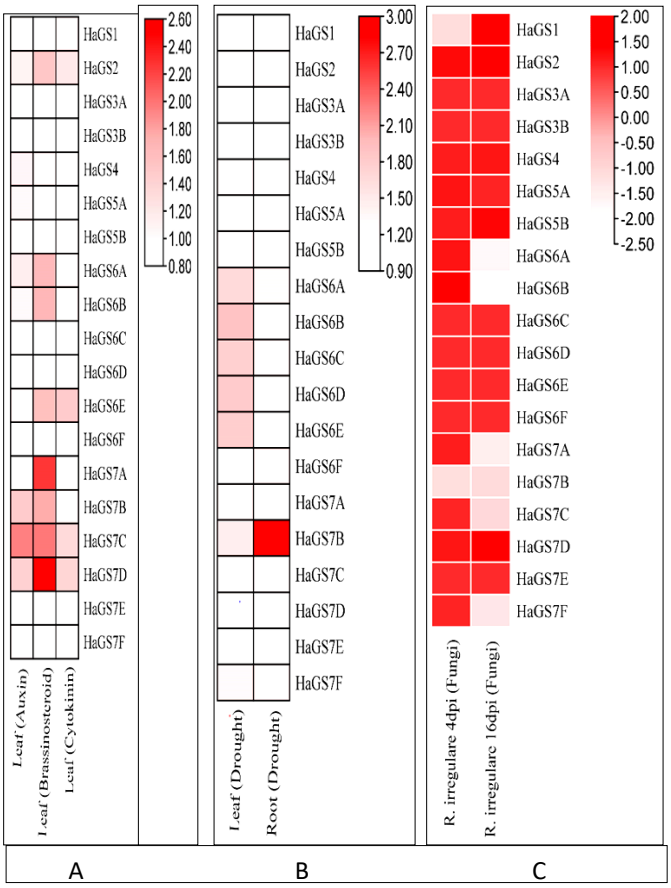

Fig. 5. *In silico* expression profiling of HaGS genes: (**A**) Hormone (**B**) Drought (**C**) *Rhizoglomus irregulare*.

root showed the highest expression at 2.98. The remaining genes showed poor expression, i.e., 1-1.28.

Hormone stress

The hormone stress was studied in the leaf of *H. annuus*. In auxin, one gene (HaGS7C) showed the highest expression at 1.95. Two genes, HaGS7B and HaGS7D, showed medium expression, at 1.62 and 1.58, respectively. The rest of the genes showed expression between 1 and 1.35. The expression of these genes is very poor against hormone stress. In brassinosteroid, we observed that one gene (HaGS7D) showed the highest expression at 2.5. Six genes (HaGS6E, HaGS6A, HaGS6B, HaGS7B, HaGS7C, and HaGS7A) showed medium expression between 1.7 and 2.2. The rest of the genes showed expression between 1 and 1.65. The expression of these genes is very poor against hormone stress. In cytokinin, we observed that one gene (HaGS6E) showed the highest expression at 1.62. Three genes (HaGS2, HaGS7C, and HaGS7D) showed medium expression between 1.4-1.54. The rest of the genes showed expression between 1 and 1.21. The expression of these genes is very poor against hormone stress.

Biotic stress

Under biotic stress, *H. annuus* expresses genes for fungi (*Rhizoglomus irregulare*) under two conditions, i.e., 4 dpi and 16 dpi. At 4 dpi and 16 dpi, one gene (HaGS6B) showed the highest expression at 1.41 and 2.47, respectively. At 4 dpi, three genes (HaGS2, HaGS5A, and HaGS6A) showed medium expression between 1.21–1.32. The rest of the genes showed expression between 1–1.19. At 16 dpi, five genes (HaGS1, HaGS2, HaGS6A, HaGS7A, and HaGS7D) showed medium expression between 1.4–1.71. The rest of the genes showed expression between 1–1.36.

Analysis of protein

Characterization of HaGS proteins

The HaGS protein exhibits a molecular weight range from 11649.13 (HaGS7E) to 91681.91 (HaGS4). The aliphatic index ranges from 57.77 (HaGS3A) to 104.64 (HaGS6E), and the isoelectric point (PI) ranges from 4.63 (HaGS6D) to 9.33 (HaGS3A). Among the 19 genes, 11 were identified as stable, and 8 were classified as unstable. Consequently, 57.89% of the genes were deemed stable, and 42.10% were characterized as unstable. The range of the grand average of hydropathy, or GRAVY, is -0.099 to 0.07 (Supplementary Table 5). In *H. annuus*, the peptides of glutamine synthetase are rich in 4 amino acids (arginine, glycine, isoleucine & leucine, ranging from 7.8% to 15.2%) (Supplementary Table 6).

Functional domains and motifs of HaGS proteins

In 19 HaGS proteins, the number of amino acids varies from 82 amino acids (HaGS6D) to 827 (HaGs4) amino acids, the mean of which is 321 amino acids (Table 2). The length of the subunit ranges from 1 to 815. Supplementary Fig. 3 illustrates a compilation of 10 distinct motifs, along with their corresponding e-values provided in Supplementary Table 7.

Table 2. Information on the amino acid counts in *H. annuus* and *A. thaliana*.

H. annuus		*A. thaliana*	
Name of the protein	**Amino acids**	**Name of the protein**	**Amino acids**
HaGs1	545	AtGS1	353
HaGs2	119	AtGS2	488
HaGs3A	130	AtGS3	354
HaGs3B	236		
HaGs4	827	AtGS4	852
HaGs5A	523	AtGS5	522
HaGs5B	517		
HaGs6A	430	AtGS6	430
HaGs6B	430		
HaGs6C	116		
HaGs6D	82		
HaGs6E	110		
HaGs6F	125		
HaGs7A	357	AtGS7	356
HaGs7B	357		
HaGs7C	385		
HaGs7D	354		
HaGs7E	101		
HaGs7F	356		

Conserved amino acids and multiple sequence alignment evaluation

A percent identity matrix table is used to examine the similarities between protein components (Supplementary

Table 8). The highest similarity is seen between the gene AtGS7 of *A. thaliana* and HaGS7A of *H. annuus* ((90.5%). The lowest similarity is observed between AtGS1 and HaGS1 (19.68%) and AtGS2 and HaGS2 (23%). Other genes show a similarity range between 63.7% and 89.9%. A similarity between amino acids was observed between *H. annuus* and *A. thaliana* genes. It had 5 clusters at 62 % identity (Supplementary Fig. 4 & 5).

Sub-cellular localization and function

The HaGS proteins are part of the cytoplasm. Gene ontology analysis and functional annotation indicate that these proteins are involved in various functions, including glutamate-ammonia ligase activity, ligase activity forming carbon-nitrogen bonds, glutamate-cysteine ligase activity, catalytic activity, acid-amino acid ligase activity, glutamine biosynthetic process, glutamine metabolic process, glutamine family amino acid biosynthetic process, amino acid metabolic process, alpha-amino acid biosynthetic process, amino acid biosynthetic process, carboxylic acid biosynthetic process, organic acid biosynthetic process, small molecule biosynthetic process, carboxylic acid metabolic process, oxoacid metabolic process, organic acid metabolic process, and glutathione biosynthetic process. We did protein interaction to know the unknown functions of proteins. We found 19 protein peptide nodes that correspond to several gene sequences (Supplementary Fig. 6). Based on the protein network and functional roles, we can see that the peptides retrieved show different types of functions (Supplementary Table 9).

Secondary and tertiary 3-D structure

A comparison of the secondary structure of the 19 HaGS proteins revealed a prevalence of α-helices, followed by the random coil structure. The random coil structure forms an irregular region that facilitates a distinctive folding of the polypeptide chain (Supplementary Table 10). All 19 proteins were utilized for the determination of *in silico* 3-D structures. The Global Model Quality Estimation (GMQE) for these proteins ranges from 0.33 to 0.89. Notably, 14 out of the 19 proteins exhibit a GMQE ranging from 0.71 to 0.89, indicating a high-quality protein model. The Qmean values range from 0.48±0.11 to 0.92±0.05, while the sequence identity spans from 28.67% to 90.11%. The quality factor, determined using ERRAT, varies from 79.0278 to 96.7336. Additionally, the 3D-1D score, determined through verified 3D, ranges from 37.88% to 95.87% (Supplementary Fig. 7 and Supplementary Table 11).

Alignment and functional annotation of 3D structure

The 3D structures, characterized by minimum energy, of *H. annuus* were superimposed onto the corresponding 3D protein structure of the reference protein from *A. thaliana* (Supplementary Table 12). The 3D structures of the 19 HaGS proteins exhibit similarity levels ranging from 9.93% to 91.71% with their respective *A. thaliana* proteins. The root mean square deviation (RMSD) value varies from 0.13 to 24.89.

Phylogenetic analysis

The phylogenetic tree was constructed using the amino acid sequences of both *H. annuus* and *A. thaliana*. Three different clades were observed showing the homology between the genes viz. HaGS2 and AtGS2, HaGS4 and AtGS4 and AtGS5 and HaGS5A and HaGS5B (Fig 6). Meme results show that motifs 1, 2, 3, 4, 5 & 9 are most conserved in 19 proteins of glutamine synthetase (Fig. 1).

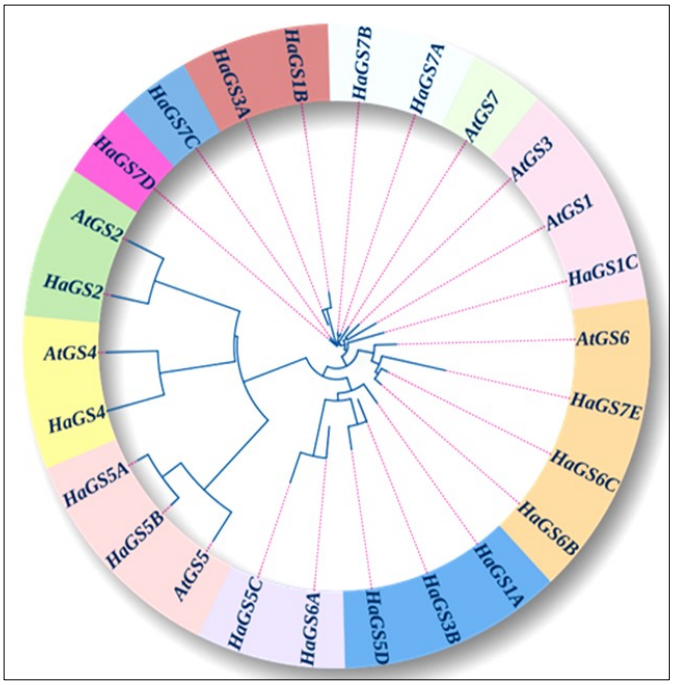

Fig. 6. Phylogenetic tree constructed using protein sequences of glutamine synthetase genes of *H. annuus* and *A. thaliana*.

Discussion

The genome sequences have been employed to investigate genes associated with diverse developmental stages and stress tolerance across numerous crops. Crops without sequenced genomes are benefiting from those with sequenced genomes. *A. thaliana* and *Oryza sativa* (21) have been employed in such investigations.

Identification of HaGS genes

This study is the first report in *H. annuus* to analyze the glutamine synthetase gene family. During this study, 19 HaGS genes were identified in *H. annuus* against 7 genes of *A. thaliana* (Fig. 7). The structural arrangement of exons and introns in most HaGS genes exhibits a consistent pattern. Variances in cDNA sequences are primarily attributed to differences in the size and number of introns within HaGS genes. Disparities in cDNA sequence lengths may

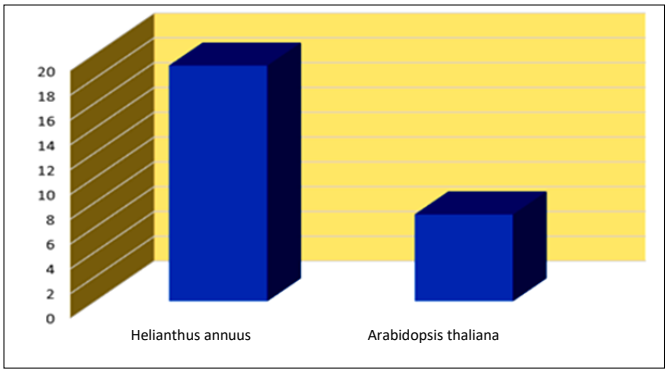

Fig. 7. The number of genes encoding glutamine synthetase in *H. annuus* (19) and *A. thaliana* (7)

also arise from the presence of UTRs flanking the cDNA borders. Phylogenomic analysis unveiled three clades in the relationship between the HaGS genes of *H. annuus* and AtGS genes of *A. thaliana*. There are 17 chromosomes in *H. annuus*, and the genes for glutamine synthetase are present on 11 of them. An uneven distribution of the genes is observed on the chromosomes. The variation in the distribution of orthologous genes among chromosomes could be attributed to processes such as inversion and translocation.

Duplication and synteny analysis in HaGS genes

The current analysis of the glutamine synthetase gene family in *H. Annuus* reveals no chromosomal duplication events. Chromosome 16 has 5 genes which are the highest in number, whereas chromosomes 5, 6, 9, 10, 12, and 17 have one gene each. Translocation events are not evident in the evolutionary process. The lack of duplication may result from deletion events. We examined the synteny between all chromosomes of *A. thaliana* and the 19 glutamine synthetase genes of *H. annuus*. Notably, all 19 genes exhibited synteny with *Arabidopsis* chromosomes, but no collinearity was observed. Each chromosome of *Arabidopsis* displayed synteny with certain genes of *Helianthus*. Chromosomes 1 and 5 of *Arabidopsis* demonstrated the highest synteny, each aligning with two genes of *Helianthus*. in contrast, other chromosomes of *Arabidopsis* showed synteny with one gene of *Helianthus*.

Promoter sequence analysis in HaGS gene

The glutamine synthetase genes take part in many biological processes like nitrogen metabolism and responses to different biotic and abiotic stress. Promoter analysis revealed that there are many cis-acting elements in HaGS genes. The elements present in the maximum number of genes are TCCC-motif, G-box, ARE, GT1-motif, TCA-element, LTR, G-box, O2-site, ABRE, TCT-motif, GATA-motif, CCAAT-box, CGTCA-motif, P-box, I -box, MBSI, TGA-element, MBS, TGACG-motif, HD-Zip 1, ACE, chs-CMA1a, AuxRR-core, Box 4, AE-box, Sp1, 3-AF1 binding site, CAT-box, RY-element, Unnamed_1, TATC-box, TC-rich repeats, Box II, MRE, 4cl-CMA1b, chs-CMA2a, GC-motif, circadian, GARE-motif, and AT1-motif (Fig. 2).

The TCCC-motif, GT1-motif, TCT-motif, GATA-motif, I-box, Sp1, Box-II, 4cl-CMA1b, chs-CMA2a, and 3-AF1 binding sites are light-responsive elements. The G-box, ACE, box 4, AE-box, and MRE are responsible for light responsiveness. The element ARE is responsible for anaerobic induction. The TCA element shows salicylic acid responsiveness. Low-temperature responsiveness is due to LTR. O2-site shows zein metabolism regulation. ABRE shows abscisic acid responsiveness. CCAAT-box is the MYBHv1 binding site. CGTCA-motif and TGACG-motif exhibit MeJA-responsiveness. P-box and TATC-box are gibberellin responsive elements. MBSI is involved in the regulation of flavonoid biosynthetic genes. TGA-element and AuxRR-core shows auxin responsiveness. MBS is responsible for drought inducibility. RY-element is responsible for seed-specific regulation. HD-Zip 1 is responsible for the differentiation of the palisade mesophyll cells. TC-rich repeats show defense and stress responsiveness.

Presence of SSR and miRNA in HaGS genes

In this study, SSRs were identified in 16 out of the 19 genes analyzed. A total of 51 SSRs were detected, indicating their prevalent occurrence within a significant portion of the gene family. The structural and functional features of SSRs have been shown in a wide range of genes. The frequency of mononucleotide was the highest (16), followed by pentanucleotide motifs (12), septanucleotide (5), trinucleotide (5), dinucleotide (4), tetranucleotide (4) and hexanucleotide (4).

This occurrence is typical because, generally, trinucleotide repeats are more prevalent compared to other SSRs. The presence of SSRs within the coding sequences of genes introduces polymorphism, serving as valuable genetic resources for developing functional markers associated with specific traits. The SSR identified in the gene responsible for encoding glutamine synthetase can be utilized to create functional markers in marker-assisted selections. Such markers have the potential to enhance tolerance against both biotic and abiotic stresses in plant studies. MicroRNAs play regulatory roles at post-transcriptional and translocational levels within cells (29). In this study, we identified one miRNA associated with the sequence of HaGS7A. The target sites for this miRNA are located between positions 1374-1393, and the identified miRNA is han-miR3630-5p.

Structural and functional features of HaGS genes

There are variations in the predicted proteins of HaGS subunit genes. They also differ from the protein length of genes similar to *A. thaliana*. In 19 HaGS proteins, the number of amino acids varies from 82 aa (HaGS6D) to 827 aa (HaGS4). The GRAVY (Grand Average of Hydropathy) ranges between -0.099 to 0.07. This indicates that the proteins with a negative GRAVY value are hydrophobic. This characteristic facilitates the appropriate folding of the protein, ensuring its biological activity and stability. Proteins are either polar or non-polar depending on the GRAVY value; a negative GRAVY value implies the former. As shown in Fig. 1, proteins with ten different motifs are encoded by each HaGS gene. The individual motifs range from 15 amino acids (motif 9) to 50 amino acids (motifs 6, 8, and 10). Notably, 8 of these motifs are identified as novel, and their molecular characteristics are yet to be characterized. The identified motifs, i.e., motif 8 and motif 10 are said to be involved in glutamate-cysteine ligase activity. It is expected that the majority of HaGS genes are present in the cytoplasm.

Phylogenetic analysis

The orthology among different proteins of glutamine synthetase genes of different species was observed. For this purpose, separate phylogenetic trees for glutamine synthetase of aforesaid species were constructed.

In silico expression analysis of HaGS genes

This study represents the first report in *H. annuus* on the phylogenomic analysis of genes encoding glutamine synthetase. In this study, we identified 19 genes for glutamine synthetase. Of these, one gene has been identified in the

root (HaGS7B) & leaf (HaGS6D) each for drought stress, one gene (HaGS6B) in the root (biotic stress), and two genes (HaGS7C & HaGS6E) in leaf (hormone stress). A phylogenetic tree was constructed using the Clustal Omega database between the HaGS genes of *H. annuus* and the AtGS genes of *A. thaliana*. Three different clades were observed showing the homology between the genes viz. HaGS2 and AtGS2, HaGS4 and AtGS4 and AtGS5 and HaGS5A and HaGS5B. Crop production is negatively impacted by abiotic stress, leading to losses. In the presence of abiotic stress, numerous genes are triggered, and their resulting products play crucial roles in stress responses and tolerance. Various plants, including rice and *Arabidopsis*, have been documented to possess a substantial number of abiotic stress-responsive genes. Drought inhibits plant growth and accelerates plant senescence or even death. Thus, drought stress has a significant impact on agricultural production. In drought stress, up-regulation of GS1 expression in leaves, roots, and calluses takes place. GS1 increases in leaves and roots during drought stress. GS1 increases drought tolerance (30, 31). We identified the up-regulation of two genes (HaGS6D & HaGS7B) under drought stress conditions in *H. annuus*.

The attack of a pathogen, such as fungal, bacterial, or viral attacks, leads to stress conditions in the plant. Plant infection by bacterial and fungal pathogens also alters the nitrogen metabolism of the host plant. Glutamine synthetase gene overexpression in the HaGS6B gene of *H. annuus* suggests its role in protecting the plant against fungal infection by *Rhizoglomus irregulare*. We identified upregulation of 1 gene (HaGS6B) under biotic stress conditions in *H. annuus*.

When hormone stress was studied in the leaves of *H. annuus*, 3 hormones were included in that study, viz., auxin, brassinosteroid, and cytokinin. It was observed that a single gene, HaGS7C, showed overexpression in the presence of both auxin and brassinosteroid. It suggests their role in maintaining the stress levels of plants. As for cytokinin, the gene HaGS6E showed high expression. We identified the upregulation of 2 genes (HaGS7C & HaGS6E) under hormone stress conditions in *H. annuus.* However, this is a preliminary study, and an in-depth future analysis is suggested. Genes activated during stress conditions play a crucial role in safeguarding plants against stress through synthesizing vital metabolic proteins. As a result, this study provides insightful knowledge that can be used to create *H. annuus* cultivars resistant to drought.

Conclusion

The analysis of glutamine synthetase (GS) genes in *H. annuus* offers valuable insights into their structural, functional, and evolutionary aspects. This study identified 19 HaGS genes that highlight a consistent exon-intron arrangement and variability due to intron size and number, as well as untranslated regions (UTRs). Phylogenomic analysis revealed three clades linking HaGS genes with *A. thaliana* genes, indicating evolutionary relationships. It can be noted that the HaGS genes are unevenly distributed across 11 of the 17 chromosomes in *H. annuus*, with no chromosomal duplications but possible deletions. Promoter analysis uncovered numerous cis-acting elements associated with various stress responses and developmental processes. Simple sequence repeats (SSRs) and a microRNA (hanmiR3630-5p) within HaGS genes present opportunities for developing genetic markers for stress tolerance. The structural and functional features, including variations in protein length and hydropathy, suggest distinct roles for HaGS proteins, with some novel motifs yet to be characterized. In silico expression analysis showed the upregulation of specific HaGS genes under drought, biotic, and hormone stress conditions, highlighting their potential roles in stress tolerance. Thus, this crucial study lays the groundwork for future research to enhance *H. annuus* cultivars' resilience to environmental stresses, contributing to improved crop production and sustainability.

Acknowledgements

The facilities (dev/1043;29/06/2022) provided by CCS University, Meerut, India, are acknowledged by the authors for enabling the Department of Botany to conduct this study.

Authors' contributions

VS & VK carried out the major studies, including data retrieval and analysis of gene expression. MR & PJ performed sequence alignment. HS carried out an analysis of the motif and conserved domain. DT designed the secondary structure of this study. AA & MKB retrieved data. VM, SK & PM drafted and edited the manuscript. All authors read and approved the final manuscript. Readers can get supplementary figures & tables from the first and corresponding authors.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None.

Supplementary data

Fig. 1. Ensembl plants gene tree pipeline showing the orthologous and paralogous relationship of the HaGS gene with other taxa. Duplication nodes are shown as red squares, whereas speciation nodes are in blue.

Fig. 2. Cis-regulatory elements with the location in 19 HaGS genes.

Fig. 3. The MEME algorithm was used to identify the logos of each of the ten motifs and related amino acids found in the sequences of HaGS proteins. The relative heights of the letters represent the frequencies and conservation levels of each motif. The length of the motif is represented by the X-axis, and the sequence conservation per site (bit score) for each letter is represented by the Y-axis.

Fig. 4. Amino acid sequence alignments of the glutamine synthetase genes of *H. annuus* and *Arabidopsis*. The conserved domains of both species are shown in different colours.

Fig. 5. The protein HaGS of *H. annuus* conserved and co-evolved in amino acid residues. Coevolving and conserved residues are displayed by mutual information networks. The amino acid locations of *H. annuus*'s HaGS proteins are shown by the labels on the outside of the second circle. The second circle's colored square boxes show the degree of conservation (highly conserved places are shown in red, while less conserved points are shown in blue). The proximity mutual information (PMI) is displayed in the third circle. Pairs of positions having a MI larger than 6.5 are connected by lines. According to MISTIC, red edges indicate the top 5%, black edges are between 70% and 95%, and grey edges make up the other 70%.

Fig. 6. Protein interaction network of 19 proteins of glutamine synthetase gene family in *H. annuus*.

Fig. 7. The three dimensional structures of 19 proteins of glutamine synthetase (GS) gene family in *H. annuus*. In all the 19 proteins, spirals represent helices, broad strips with arrow-head represent β-pleated sheets and thin loops represent coils.

Table 1. Number of exons and introns in the genes of glutamine synthetase family in *H. annuus*.

Table 2. Ka/Ks value of *H. annuus* & *A. thaliana* gene pair.

Table 3. Simple sequence repeats (SSRs) in HaGS genes in *H. annuus*.

Table 4. Predicted miRNA targets in HaGS genes.

Table 5. Physiochemical properties of 19 HaGS proteins encoded by all genes.

Table 6. Amino acid content in glutamine synthetase of *H. annuus.*

Table 7. List of identified motifs and their sequence in HaGS proteins.

Table 8. Amino acid sequence homology (per cent identity) among the glutamine synthetase genes of *H. annuus* and *A. thaliana.*

Table 9. Functional annotation of glutamine synthetase protein in *H. annuus.*

Table 10. Analysis of the secondary structure of HaGS proteins.

Table 11. 3D structure analysis of 19 HaGS proteins used for protein modelling.

Table 12. Predicted values of different parameters obtained after superimposition of 3D protein structures of HaGS proteins over 3D protein structure of AtGS1 gene of *Arabidopsis.*

References

1. Rolnik A, Olas B. The plants of the Asteraceae family as agents in the protection of human health. International Journal of Molecular Sciences. 2021 Mar 16;22(6):3009. https://doi.org/10.3390/ijms22063009
2. Karthikeyan S, Sanjappa M, Moorty S, Bhattacharjee B, Dash SS, Meena SL, Mastakar VK. Asteraceae. In: Mao, A. A. and Dash, S.S. Flowering Plants of India- An Annotated Checklist (eds.). Botanical Survey of India. I: 713-816. ISBN: 978-81-9452292-8; 2020.
3. SunitMitra SM, MuKherjee SK. Asteraceae of India: its diversity and phytogeographical affinity. In: Plant Biodiversity: Monitoring, Assessment and Conservation; Wallingford UK: CABI; 2017: pp. 36-70. https://doi.org/10.1079/9781780646947.0036
4. Stefansson BR. Oilseed crops. The Canadian Encyclopedia (Historica Foundation, Toronto) (Accessed April 26, 2023); 2007.
5. Grabowska A, Kwinta J, Bielawski W. Glutamine synthetase and glutamate dehydrogenase in triticale seeds: molecular cloning and genes expression. Acta Physiologiaeplantarum. 2012 Nov;34:2393-406. https://doi.org/10.1007/s11738-012-1085-9
6. Masclaux-Daubresse C, Carrayol E, Valadier MH. The two nitrogen mobilisation-and senescence-associated GS1 and GDH genes are controlled by C and N metabolites. Planta. 2005 Jun;221(4):580-88. https://doi.org/10.1007/s00425-004-1468-2
7. Morcuende R, Krapp A, Hurry V, Stitt M. Sucrose-feeding leads to increased rates of nitrate assimilation, increased rates of α-oxoglutarate synthesis and increased synthesis of a wide spectrum of amino acids in tobacco leaves. Planta. 1998 Aug; 206 (3):394-409. https://doi.org/10.1007/s004250050415
8. Unno H, Uchida T, Sugawara H, Kurisu G, Sugiyama T, Yamaya T, et al. Atomic structure of plant glutamine synthetase: a key enzyme for plant productivity. Journal of Biological Chemistry. 2006 Sep 29;281(39):29287-96. https://doi.org/10.1074/jbc.M601497200
9. Bielawski W. Distribution of glutamine synthetase isoforms in triticale seedling leaves. Acta Physiologiae Plantarum. 1993;15 (4).
10. Miflin BJ, Habash DZ. The role of glutamine synthetase and glutamate dehydrogenase in nitrogen assimilation and possibilities for improvement in the nitrogen utilization of crops. Journal of Experimental Botany. 2002 Apr 15;53 (370):979-87. https://doi.org/10.1093/jexbot/53.370.979
11. Edwards JW, Coruzzi GM. Photorespiration and light act in concert to regulate the expression of the nuclear gene for chloroplast glutamine synthetase. The Plant Cell. 1989 Feb 1;1(2):241-48. https://doi.org/10.1105/tpc.1.2.241
12. Cock JM, Brock IW, Watson AT, Swarup R, Morby AP, Cullimore JV. Regulation of glutamine synthetase genes in leaves of *Phaseolus vulgaris*. Plant Molecular Biology. 1991 Oct;17:761-71. https://doi.org/10.1007/BF00037059
13. Li MG, Villemur R, Hussey PJ, Silflow CD, Gantt JS, Snustad DP. Differential expression of six glutamine synthetase genes in *Zea mays*. Plant Molecular Biology. 1993 Oct;23:401-07. https://doi.org/10.1007/BF00029015
14. Lam HM, Coschigano KT, Oliveira IC, Melo-Oliveira R, Coruzzi GM. The molecular-genetics of nitrogen assimilation into amino acids in higher plants. Annual Review of Plant Biology. 1996 Jun;47(1):569-93. https://doi.org/10.1146/annurev.arplant.47.1.569
15. Finnemann J, Schjoerring JK. Translocation of NH^{4+} in oilseed rape plants in relation to glutamine synthetase isogene expression and activity. Physiologia Plantarum. 1999 Mar;105(3):469-77. https://doi.org/10.1034/j.1399-3054.1999.105311.x
16. Cabello P, Agüera E, De La Haba P. Metabolic changes during natural ageing in sunflower (*H. annuus*) leaves: expression and

activity of glutamine synthetase isoforms are regulated differently during senescence. Physiologia Plantarum. 2006 Sep;128 (1):175-85. https://doi.org/10.1111/j.1399-3054.2006.00722.x

17. Eisen JA. Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis. Genome Research. 1998 Mar 1;8(3):163-67. https://doi.org/10.1101/gr.8.3.163
18. DeSalle R, Tessler M, Rosenfeld J. Phylogenomics: a primer. CRC Press; 2020 Aug 18. https://doi.org/10.1201/9780429397547
19. Kumar V, Sharma H, Saini L, Tyagi A, Jain P, Singh Y, et al. Phylogenomic analysis of 20S proteasome gene family reveals stress-responsive patterns in rapeseed (*Brassica napus* L.). Frontiers in Plant Science. 2022 Oct 31;13:1037206.https://doi.org/10.3389/fpls.2022.1037206
20. Lu K, Wei L, Li X, Wang Y, Wu J, Liu M, et al. Whole-genome resequencing reveals *Brassica napus* origin and genetic loci involved in its improvement. Nature Communications. 2019 Mar 11;10 (1):1154.https://doi.org/10.1038/s41467-019-09134-9
21. Rensink WA, Buell CR. *Arabidopsis* to rice. Applying knowledge from a weed to enhance our understanding of a crop species. Plant Physiology. 2004 Jun 1;135(2):622-29.https://doi.org/10.1104/pp.104.040170
22. Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Current Biology. 2003 Sep 2;13(17):1512-17.https://doi.org/10.1016/S0960-9822(03)00558-X
23. Yu X, Han J, Wang E, Xiao J, Hu R, Yang G, He G. Genome-wide identification and homoeologous expression analysis of PP2C genes in wheat (*Triticum aestivum* L.). Frontiers in Genetics. 2019 Jun 12;10:561.https://doi.org/10.3389/fgene.2019.00561
24. Li XM, Chao DY, Wu Y, Huang X, Chen K, Cui LG, et al. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. Nature Genetics. 2015 Jul;47(7):827-33.https://doi.org/10.1038/ng.3305
25. Clavijo BJ, Venturini L, Schudoma C, Accinelli GG, Kaithakottil G, Wright J, et al. An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. Genome Research. 2017 May 1;27(5):885-96.https://doi.org/10.1101/gr.217117.116
26. Parmentier Y, Bouchez D, Fleck J, Genschik P. The 20S proteasome gene family in *A. thaliana*. FEBS Letters. 1997 Oct 27;416(3):281-85. https://doi.org/10.1016/S0014-5793(97)01228-3
27. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Research. 2023 Jan 6;51(D1):D638-46.https://doi.org/10.1093/nar/gkac1000
28. The *Arabidopsis* genome initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 2000;408:796-815. https://doi.org/10.1038/35048692
29. Malik V, Ikram FM, Singh Y, Kumar V, Malik P, Balyan P, et al. Structural and functional characteristics and expression profile of the 20S proteasome gene family in *Sorghum* under abiotic stress. Front Plant Sci. 2023 Nov 29;14:1287950. https://doi.org/10.3389/fpls.2023.1287950.
30. Santos C, Pereira A, Pereira S, Teixeira J. Regulation of glutamine synthetase expression in sunflower cells exposed to salt and osmotic stress. Scientia Horticulturae. 2004 Dec 31;103 (1):101-11.https://doi.org/10.1016/j.scienta.2004.04.010
31. Yin H, Sun Q, Lu X, Zhang L, Yuan Y, Gong C, et al. Identification of the glutamine synthetase (GS) gene family in four wheat species and functional analysis of Ta4D. GSe in *A. thaliana*. Plant Molecular Biology. 2022 Sep;110(1-2):93-106.https://doi.org/10.1007/s11103-022-01287-4