MOLECULAR CHARACTERIZATION OF *GBCHS* AND *GBGSTS* GENES INVOLVED IN FLAVONOIDS BIOSYNTHESIS IN (*GINKGO BILOBA* L.)

Rizk, Mokhtar S.
Department of Genetic Resources, Desert Research Center, El-Matareya, Cairo, Egypt
E-mail: drmokhtarsaid@yahoo.com

Ginkgo biloba is a well-known living gymnosperm fossil that has medicinal, biological, and economic value in the world. In this study, molecular characterization and bioinformatics analysis based on homology modeling were obtained by *GbGST*s and *GbCHS* genes, which are crucial in the pathway of plant flavonoids, anthocyanins, and other significant secondary metabolites in plants. The results indicated that a 687 bp open reading frame (ORF) encoding a 228 amino acids protein with a determined molecular weight of around 25.786 kDa were found in the full-length cDNA of the *GbGST*s gene sequence that was isolated from *G. biloba*. Whilst the full-length cDNA of the *GbCHS* gene sequence contained a 1176 bp (ORF) encoding a 391 amino acids protein with a predicted molecular weight of about 43.078 kDa. Additionally, phylogenetic analyses were carried out using the amino acid sequences of *GbGST*s and *GbCHS* with other recognized plants that were obtained from NCBI. Multiple Sequence Alignment (MSAs) of chosen (30) amino acids with high identity and similarity using MEGA7 software, ProtParam software was used to the molecular weight and grand average of hydropathicity (GRAVY), and Protscale software was used to the hydrophilicity scales based on various. Finally, the molecular characterization and bioinformatics analysis of *GbGSTs* and *GbCHS* genes encoding key enzymes is the first step to fully understanding the regulatory mechanisms controlling flavonoid and anthocyanin biosynthesis in *G. biloba*.

Keywords: Ginkgo, *GbGST*s gene, *GbCHS* gene, phylogenetic tree, flavonoid

INTRODUCTION

A kind of deciduous tree in the Ginkgo family called *Ginkgo biloba* L. has a long tradition of cultivation and is commonly used in food, medicine,

and health products in addition to garden landscapes. A well-known long-living tree with significant economic, decorative, and research values is *G. biloba* (Yan et al., 2021). However, there have been very few findings on the systematic selection of optimal reference genes based on transcriptome data in *G. biloba* (Zhou et al., 2020). Furthermore, *G. biloba* is the unique and famous gymnosperm and the sole surviving member of the genus Ginkgoales, which is considered to have started thousands of years ago (Zhou and Zheng, 2003). Also, the well-known long-living tree species *G. biloba* can live for hundreds or even thousands of years (Wang et al., 2020). It is a popular extinct plant known as a "living fossil" in the kingdom Plantae. However, the temperature range of *G. biloba's* natural environment, which ranges from 4 to 40 degrees Celsius, limits the plant's ability to spread geographically (Cao et al., 2012). Furthermore, it is a considerable commercial tree species that are extensively used in both production and day-to-day life. *G. biloba* extract, which is a key source of its economic value and contains flavonoids, is frequently utilized in clinical therapy (Efferth and Koch, 2011 and Zhao et al., 2019). Flavonoids are one of the many categories of secondary metabolites that are present in plants and are essential for scavenging free radicals, halting oxidation, and preserving plant growth and development (Ku et al., 2020). Ni et al. (2017) revealed that post-harvest *G. biloba* leaves treated with 200 mmol/L NaCl significantly boosted the concentration of flavonoids. Also, numerous biological processes are mediated by flavonoids, including defense against phytopathogens and herbivores, ultraviolet filtering for tissue protection, anthocyanin production to draw insect pollinators, pollen germination, biological communication in the rhizosphere, regulation of auxin transport and catabolism, and most importantly, antioxidant activity that prevents the production of reactive oxygen species, which occurs in living things (Mouradov and Spangenberg, 2014).

Plant glutathione S-transferees (GSTs) are versatile enzymes that help proceed with anthocyanin, plant GSTs are significant non-catalytic carriers of protein for the absorption of anthocyanin by vacuoles in plants. Additionally, almost all aerobic species including plants, microbes, and humans, have a huge, old, and diversified set of multifunctional proteins known as glutathione S- transferee's (GSTs) (Kitamura et al., 2004 and Fang et al., 2020). Also, the GST enzyme family is a member of the multi-gene family and is highly complicated. Furthermore, it performs a crucial regulatory role in a variety of metabolic processes (Wagner et al., 2002 and Edwards and Dixon, 2005). For instance, plants GSTs effectively remove heavy metals, harmful lipid peroxides, and unusual microbes, and finally play an important role in stress response (Lan et al., 2009 and Csiszár et al., 2014). The essential function of GST*s* is to catalyze the addition of *GSH* to heterocyclic organic anions (Vaish et al., 2020). Additionally, two conserved domains are present in the GST protein. At the N-terminus, there is a conserved *GSH*-binding domain (G-site), and at the C-terminus, there is a substrate-binding domain (H-site). Through

the three-dimensional (3D) structure, these two domains are close to one another to create catalytic sites with particular functionalities (Dixon et al., 2002). Apart from essential regulatory components including ACE elements, silencers, H-box sequences, and AT-rich units, the inducible gene Chalcone Synthase (*CHS)* has several cis-acting elements linked to adversity, hormones, tissue specificity, and other systems with inducible expression (Kiba et al., 1995). Overall, the molecular analysis of the *CHS* genes revealed that most of them were split into two or more subfamilies and that the majority of the *CHS* genes were composed of two exons and one intron (Durbin et al., 2000). Furthermore, the major enzyme on the path used to synthesize flavonoids and iso-flavonoids is *CHS*, as well plays an essential role in sustaining plant growth and development (Manaf, 2013 and Kong et al., 2020). Numerous genes involved in the synthesis of flavonoids, including *GbPAL*, *GbCHI*, *GbF3H*, *GbFLS*, and *GbANS*, have been cloned and identified in *G. biloba* since the *GbCHS* gene was isolated. Additionally, most of these genes are expressed across the plant and are controlled by several internal and external variables. Moreover, it was discovered that *GbMYBF2* gene is a negative regulator of flavonoid synthesis. Lastly, the expression levels of these genes, which form the basis for flavonoid synthesis, can explain how the flavonoid content of *G. biloba* increases in response to environmental cues (Xu et al., 2014).

Identifying the cDNA genes for glutathione S-transferase and chalcone synthase from *G. biloba* was the main objective of this study. Other objectives included investigating homology modeling, functional and structure analysis prediction (showing ligands, global quality estimate, local quality estimate, sequence alignment, and phylogenetic relationship analysis with *GST*s and *CHS*s genes from other plant species), and improving understanding of the enzymatic activities of *GbGST*s and *GbCHS* from *G. biloba*, molecular mechanisms and ascertain its regulatory function in the biosynthesis of flavonoids and anthocyanin's, it was necessary to look into homology modeling, functional and structure analysis prediction (showing ligands, global quality estimate, local quality estimate, sequence identity percentage, and perform several necessary bioinformatics analysis).

MATERIALS AND METHODS

1. Plant Materials

A few young, fresh leaves of *G. biloba* were obtained from the El-Orman Botanical Garden in Giza, Egypt, which is near the end of El-Dokki Street. Samples were frozen immediately using liquid nitrogen and kept at -80°C until they were used for Real Time Polymerase Chain Reaction (RT-PCR).

2. Total RNA Extraction

G. biloba leaves were used to extract total RNA using Direct-zol™ RNA MiniPrep (Zymo Research, USA). Sample preparation and RNA purification were the two components of this technique (Catalog Nos. R2050).

3. Sample Preparation

Each leaf tissue sample was crushed to a powder under liquid nitrogen, then immediately transferred and homogenized in a tube containing 600 µl of TRI Reagent®. The supernatant was then placed into an RNase-free tube after being centrifuged at 10,000 x g for 1 min to eliminate the particle debris.

4. RNA Purification

A sample that had been lysed in TRI Reagent® was added and carefully mixed with an equal volume of ethanol (95–100%). The mixture was then centrifuged after being placed in a Zymo-Spin™ IICR Column in a collection tube. The flow-through was discarded after moving the column into a new collection tube. DNase digestion step: Using DNase/RNase-Free Water (Cat. # E1009-A, ZYMO Research Crop. Set 250 U) to remove genomic DNA contaminated by DNase I. The column was filled with around 400 µl of Direct-zol™ RNA Prewash, and it was centrifuged, continuous after discarding the flow-through. To confirm that, all the wash buffers had been completely removed, 700 µl of RNA Wash Buffer was added to the column and centrifuged for 1 min, carefully the column was transferred into an RNase-free tube. The column matrix was directly injected with 50 µl of DNase/RNase Free Water, and the mixture was centrifuged to elute the RNA. The eluted RNA was either utilized right away or put in a freezer until needed.

5. Estimation of RNA Concentration and cDNA Library Preparation

With a NanoDrop (ND-1000) spectrophotometer, total RNA was quantified, and the yield of total RNA was determined by measuring absorbance at 260 nm (A_{260}/A_{230} and A_{260}/A_{280} ratios) (NanoDrop, Technologies Inc.). By doing a separate assay using electrophoresis on a 1.2% agarose gel, the integrity of total RNA was confirmed. With the aid of mixing equal volumes from the three RNA replications in one tube, RNA pools were prepared for cDNA libraries. Using TransScript® First-Strand cDNA synthesis kits (Super Script III Reverse Transcriptase) in accordance with the manufacturer's instructions (Invitrogen™, Cat No. 18080044), two micrograms of total RNA (roughly 800 ng) per sample was used for the synthesis of total cDNA. Following that, cDNA synthesis using PCR was carried out at 42°C for 15 min, then 85°C for 5 min. The cDNA synthesis reaction was then stored at -20°C and used for the second PCR step.

6. Primers Design and RT-PCR Amplification

PCR was carried out in a 50 ml reaction mixture using specific primers to obtain the full length of *GbGSTs* gene (GST_Fwd:

ATGTCGAACGAAGAACAAGTGAAGG) with GST_Rev1: TCAATCGGTAACA AATTTCTTCCGG and for partial length *GbGST*s gene (GST_Fwd: ATGTCG AACGAAGAACAAGTGAAGG) with GST_Rev2: CCACGCCTCGTCGATGT ATTG according to accession no, AY987385.1 with *G. biloba*. Also, using specific primers to obtain the full length of *GbCHS* gene (CHS_Fwd: ATGGAAGACTTGGAGGCATTC) with CHS_Rev1: GGGTTTA CTTATTGCAAGGTACGC and for partial length *GbCHS* gene (CHS_Fwd: ATGGAAGACTTGGAGGCATTC) with CHS_Rev2: CCGTCAAGTACATGTA TCTCTTC according to *G. biloba* with accession No. DQ054841.1. The High-Fidelity DNA polymerase, Phusion® Taq (Thermo Scientific, Product codes: F-530L, 500 Unit) was used to amplify the cDNA. The reaction was done in a 50 µl total volume. Reaction contained 2 µl cDNA, 10 µl 5_X Phusion HF Buffer, 1 µl 10 mM dNTP mix, 2.5 µl primer 1 (10 µM), 2.5 µl primer 2 (10 µM), 0.5 µl Phusion DNA polymerase, 31.5 µl DEPC H_2O and spined for 15 s. PCR program was used for the amplification of cDNA GST and CHS genes. The PCR conditions were one cycle 60 s of preheated at 98ºC, 30 cycles; 30 s of denaturation at 98ºC, 30 s of annealing at 56ºC, 1 min of extension at 72ºC, followed by a final extension at 72ºC for 7-10 min. A volume of 40 µl of each sample was analyzed using 1.2% agarose gel electrophoreses with DNA ladder sizes ranged from 100 to 3000 bp and stained with ethidium bromide (Eth-Br). The PCR fragments of each sample were excised and purified from the agarose gel with a clean, sharp scalpel. The gel slice was weighed in a colorless tube and the QIAquick® Gel Extraction Kit (Qiagen, cat. no. 28706) was used according to the manufacturer's procedure to elute the PCR product from the gel for sequence.

7. Bioinformatics Analysis of *GbGST*s and *GbCHS* from *Ginkgo biloba*

Software for bioinformatics was used to examine the proteins encoded by the putative *G. biloba GbGST*s and *GbCHS* genes. Basic Local Alignment Tool (BLAST), homology, and domain searches were used to find *GbGST*s and *GbCHS*-related sequences in public databases, such as GenBank (www.ncbi.ncbi.nlm.nih.gov). *GbGST*s protein sequence from *G. biloba* with the accession number AAY54294.1 was used for BLASTp and homology searches with other plant species. For BLASTp and homology searches with other plant species, *GbCHS* proteins sequence from *G. biloba* with accession number AAY52458.1 was used. Phylogenetic tree analysis of 30 amino acid sequences of the involved *GbGST*s and *GbCHS* genes with other plant species was carried out in the MEGA 7.0 software program by the Maximum Likelihood method. Multiple Sequence Alignment (MSA) and the JalView program with total protein sequences were used to compare and perform using the software online Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

8. Primary and Secondary Structure Prediction and Functional Analysis

This study used an online server to assess the secondary structure of the *G. biloba GbGST*s and *GbCHS* proteins based on the gene sequence. This protein's secondary protein structure was predicted and examined using the SOPMA server, which is available online (http://npsa-pbil.ibcp.fr/cgi-bin/npsa automat. pl?Page=/NPSA/npsa sopma.html). Whilst, for the hydrophilicity prediction study, the profile created by the *GbGST*s and *GbCHS* amino acid scale on a particular protein was represented using the Protscale software online https://web.expasy.org/protscale/ (Kyte and Doolittle, 1982; Wilkins et al., 1999 and Pang et al., 2004). Several physical and chemical parameters for *GbGST*s and *GbCHS* query user entered target protein and a specified protein saved in Swiss-Prot or TrEMBL or for a sequence could be computed using the ProtParam software online at https://web.expasy.org/protparam/ (Gasteiger et al., 2005 and Tanaka and Brugliera, 2013). The SWISS-MODEL template library has been used for template selection and search with BLAST and HHBlits (Camacho et al., 2009; Remmert et al., 2011 and Steinegger et al., 2019). The quality of each detected template has been projected based on the characteristics of the target-template alignment. Afterward, for model building, the greatest quality templates have been chosen.

9. Models Building

Models Building was created based on the alignment of the target and template using ProMod3 and the template's conserved coordinates were transferred to the model. Swiss Model's online server (https://swissmodel.expasy.org/interactive) was used to estimate the three-dimensional (3D) structural properties of *GbGST*s and *GbCHS* proteins. The three-stage structure model is shown by Homology Modeling with SWISS-MODEL. National Center for Biotechnology Information (NCBI-CDD), the CASTp 3.0 server (http://sts.bioe.uic.edu/castp/index.html?1w27) (Schwede et al., 2003), and the Pfam server (http://pfam.xfam.org/family/PF00221) were used for the functional analysis (Bertoni et al., 2017 and Studer et al., 2020).

RESULTS

The results revealed that cDNA of *GbGST*s gene from *G. biloba* contains an open reading frame (ORF) of 687 bp which encoding a 228-amino-acid protein with a calculated molecular weight of about 25.786 kDa and isoelectric point (*pI*) was predicted by using the pI/Mw tool at www.expasy.org to be about 6.23, whereas, the full-length cDNA of *GbCHS* gene was 1176 bp, encoding a 391 amino acid protein with a calculated molecular weight of about 43.078 kDa and *pI* was predicted to be about 6.04**.** The obtained PCR products of partial length cDNA *GbGST*s and *GbCHS* gene

were 240 bp and 220 bp, respectively. Both fragments represent the full-length and partial cDNA of *GbGST*s and *GbCHS* genes, as shown in Fig. (1).

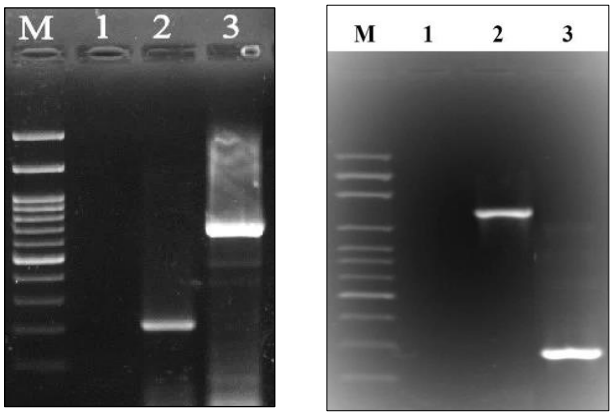

Fig. (1). RT-PCR product of full length and partial cDNA using specific primer pair to amplify *GbCHI* and *GbGST* ORF. 1: Negative control, 2 and 3 leaf samples of *Ginkgo biloba*, M: DNA size marker (100 bp DNA Ladder).

1. Analysis of *GbGST*s Gene from *Ginkgo biloba*

For search *GSTs* cDNA amino acid sequence of *G. biloba (*AAY54294.1), *Taxus chinensis* (KAH9290580.1), *Pinus densata (*WAA68344.1), *P. tabuliformis (*AGC13136.1), *P. tabuliformis* (AAT69969.1), *Larix kaempferi* (AHA46516.1), *Platycladus orientalis* (UZD10839.1), *Picea mariana* (ABA25922.1), *Phoenix dactylifera* (XP_038987897.1), *P. dactylifera* (XP_008775488.2), *Chenopodium quinoa* (XP_021715456.1), *Glycine max* (NP_001236486.2), and *Glycine max* (NP_001237686.1) were found in NCBI database. These sequences, which include the *G. biloba GbGST*s cDNA sequence, were maintained in a FASTA file (current study). Moreover, MSA of the selected *GST*s from various plant species and the deduced polypeptide sequence of *GbGST*s were performed. Using various E-values, it was observed that *GbGST*s ranged from 93 to 100% identity (Table 1).

Whilst, for search *CHS*s cDNA amino acid sequence of *G. biloba (*AAY52458.1), *Larix kaempferi* (QDF21381.1), *Picea sitchensis* (AEN84253.1), *Picea sitchensis* (AEN84248.1), *Picea glauca* (AEN84264.1), *Picea abies* (AEN84239.1), *Abies alba (*ABD24230.1), *Pinus pinaster* (CAP09644.1), *Pinus densiflora* (BAA94594.1), *Gossypium hirsutum* (NP_001314048.1), *Gossypium raimondii* (XP_012440802.1), *Hibiscus syriacus (*XP_039071121.1), *Glycine max* (KAH1236760.1), *Glycine max* (NP_001358311.1) were obtained from NCBI database. These sequences were maintained in a FASTA file including the *GbCHS* cDNA sequence of *G. biloba* (current study). MSA of the deduced polypeptide sequence of *GbCHS* and other selected *CHS*s from several plant species was

carried out. It was initiated that *GbCHS* ranged from 99 to 100% identity with 0 E-value (Table 2).

Table (1). The homology of amino acid sequences for 15 selected accession lists and its related *Ginkgo biloba* for the sequenced *GbGST*s, BLAST top hits with GenBank protein database, similarity score, accession length, and accession no.

No.	Scientific Name	Accession	Per. identify	Max score	Total score	Query cover	Acc. length	E-value
1	*Ginkgo biloba*	Current study						
2	*Ginkgo biloba*	AAY54294.1	100.00	461	461	100	228	8e-164
3	*Taxus chinensis*	KAH9290580.1	68.00	317	317	98	225	1e-106
4	*Pinus densata*	WAA68344.1	65.78	310	310	97	228	8e-104
5	*Pinus tabuliformis*	AGC13136.1	65.20	310	310	98	233	1e-103
6	*Pinus tabuliformis*	AAT69969.1	65.78	308	308	97	228	2e-103
7	*Larix kaempferi*	AHA46516.1	65.45	293	293	95	228	4e-97
8	*Platycladus orientalis*	UZD10839.1	59.91	283	283	98	226	3e-93
9	*Platycladus orientalis*	UZD10838.1	58.15	273	273	98	223	3e-89
10	*Picea mariana*	ABA25922.1	62.39	283	283	98	232	4e-93
11	*Phoenix dactylifera*	XP_038987897.1	59.82	266	266	98	221	9e-87
12	*Phoenix dactylifera*	XP_008775488.2	57.66	261	261	97	221	6e-85
13	*Chenopodium quinoa*	XP_021715456.1	55.11	263	263	98	220	3e-85
14	*Glycine max*	NP_001236486.2	53.46	236	236	99	219	3e-78
15	*Glycine max*	NP_001237686.1	54.68	232	232	93	221	7e-77

Table (2). The similarity of the amino acid sequences for the related *Ginkgo biloba* for the *GbCHS* sequenced in this investigation, as well as the top matches from BLAST searches with the GenBank protein database, similarity score, accession length, and accession no.

No.	Scientific Name	Accession	Per. identify	Max score	Total score	Query cover	Acc. length	E-value
1	*Ginkgo biloba*	Current study						
2	*Ginkgo biloba*	AAY52458.1	100.00	791	791	100.	391	0.0
3	*Larix kaempferi*	QDF21381.1	85.30	690	690	100	395	0.0
4	*Picea sitchensis*	AEN84253.1	85.04	691	691	100	395	0.0
5	*Picea sitchensis*	AEN84248.1	85.04	689	689	100	391	0.0
6	*Picea glauca*	AEN84264.1	85.04	690	690	100	395	0.0
7	*Picea abies*	AEN84239.1	84.78	688	388	100	396	0.0
8	*Abies alba*	ABD24230.1	84.78	689	689	100	395	0.0
9	*Pinus pinaster*	CAP09644.1	83.99	684	684	100	395	0.0
10	*Pinus densiflora*	BAA94594.1	84.25	686	686	100	396	0.0
11	*Gossypium hirsutum*	NP_001314048.1	84.70	679	679	99	389	0.0
12	*Gossypium raimondii*	XP_012440802.1	84.70	677	677	99	389	0.0
13	*Hibiscus syriacus*	XP_039071121.1	84.96	679	679	99	389	0.0
14	*Glycine max*	KAH1236760.1	82.63	665	665	99	388	0.0
15	*Glycine max*	NP_001358311.1	82.89	665	665	99	388	0.0

With the use of software (MEGA 7.0), a neighbor-joining phylogenetic tree was created using 30 amino acid sequences, 15 *GbGST*s genes, and 15 *GbCHS* genes from several plant species (Gymnosperms and Angiosperms). The tree with the highest log likelihood comprised *GbGST*s and *GbCHS* gene sequences (-4642.89). The phylogenetic analysis showed two branches, the first branch contains *GST*s cDNA amino acid sequence of *G. biloba* (current study), *G. biloba (*AAY54294.1), and other *GST*s sequences and the other branch contains *CHS* cDNA amino acid sequence of *G. biloba* (current study), *G. biloba (*AAY52458.1), and other *CHS*s sequences. The results revealed that *G. biloba GbGST*s cDNA in this investigation was close to *G. biloba (*AAY54294.1), and *G. biloba GbCHS* cDNA to *G. biloba (*AAY52458.1). The genetic relationship between the *GST* and *CHS* cDNA is consistent with the phylogenetic tree. Finally, a search of a database using MAS of amino acids (http://www.ncbi.nlm.nih.gov/) revealed that the deduced *GbGST*s and *GbCHS* gene had been regarded to have a high similarity with other plant's GSTs and CHSs gene families, as shown in Fig. (2).

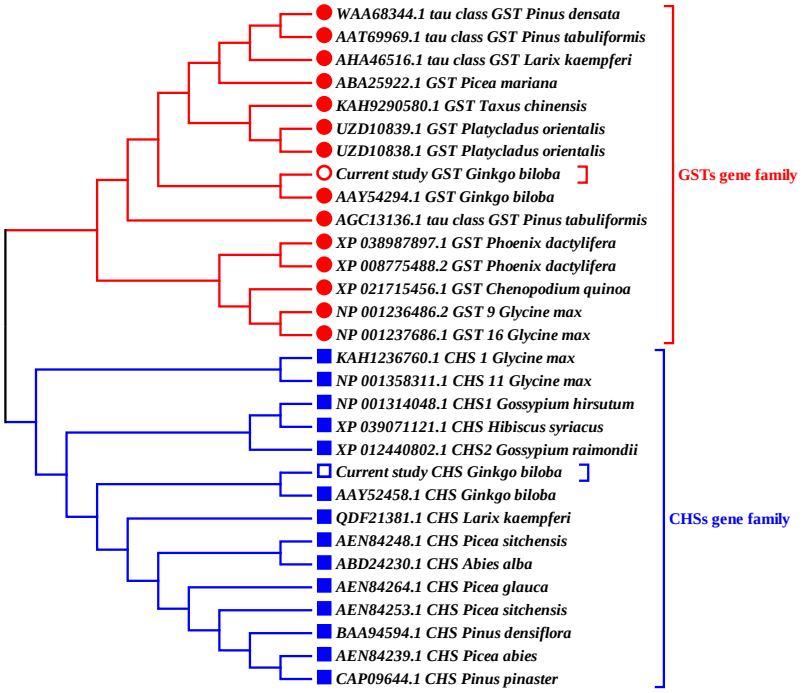

Fig. (2). The molecular phylogenetic analysis involving 30 amino acid sequences, 15 *GST*s genes and 15 *CHS*s genes from several plant species (Gymnosperms and Angiosperms) including *GbGST*s and *GbCHS* gene sequences, conducted in MEGA 7.0 software program by Maximum Likelihood method, the tree with the highest log likelihood (-4642.89).

2. Similar Motifs were Found in Motif

Based on homology modeling (Untitled Project | Models expasy.org), *GbGSTs* and *GbCHS* genes were significant and engaged in the biosynthesis pathways of flavonoids, anthocyanin, and other significant secondary metabolites in *G. biloba*. The SOPMA tool predicted the secondary structure of the *GbGSTs* protein (228aa). The results showed the majority of *GbGSTs* (Hh, 129, 56.58%) and random coils (Cc, 60, 26.32%), along with a few extended strands (Ee, 29, 12.72%) and beta turns (Tt, 10, 4.39%) as shown in Fig. (3 and 4). Furthermore, the secondary structure of *GbCHS* (391aa) was predicted. The results indicated that *GbCHS* consists mainly of α-helices (Hh) (165 is 42.20%) and random coils (Cc) (130 is 33.25%) as well as a few extended strands (Ee) (68 is 17.39%) and beta turns (Tt) (28 is 7.16%) as shown in Fig. (5 and 6).

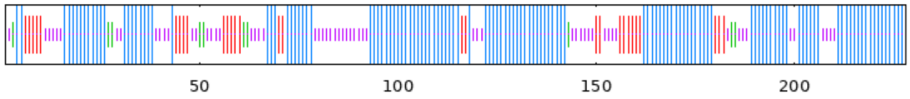

Fig. (3). Prediction of *GbGSTs* secondary structure from *Ginkgo biloba*.

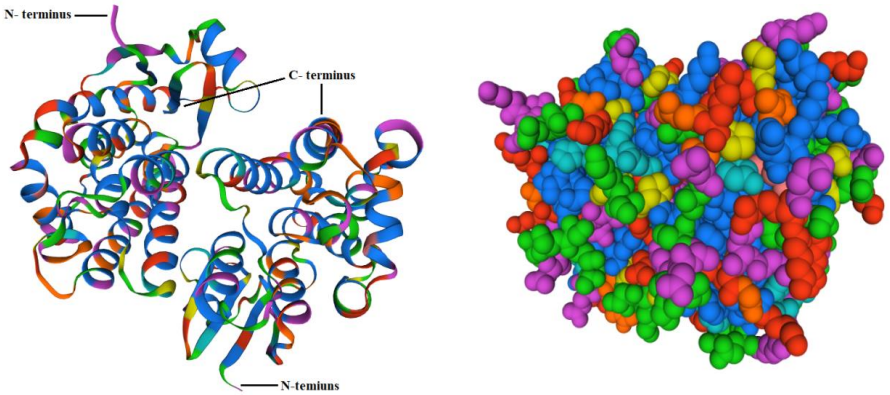

Fig. (4). The three-dimensional model of *GbGSTs* protein from *Ginkgo biloba* Prediction of *GbGSTs* secondary structure: ά–helicase in red and green and β-sheets are indicated by patches in blue by SOMPA program. Turns and loops are indicated by lines. (https://swissmodel.expasy.org/interactive).

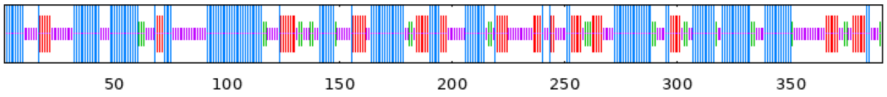

Fig. (5). Prediction of *GbGHS* secondary structure from *Ginkgo biloba*.

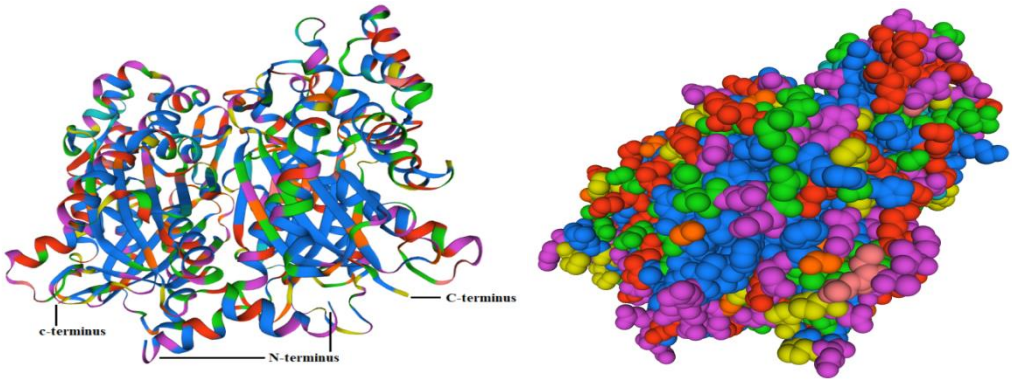

Fig. (6). The three-dimensional model of chalcone synthase *GbCHS* protein from *Ginkgo biloba*. Prediction of *GbCHS* secondary structure: ά–helicase in red and green and β sheets are indicated by patches in blue by the SOMPA program. Turns and loops are indicated by lines (https://swissmodel.expasy.org/interactive).

The Protscale software was used to assign a numerical value to each type of amino acid, which was measured based on various chemical and physical properties of the amino acids. This value serves as the basis for the amino acid scale. The computer pI/Mw Tool software was used to determine the hydrophilicity of *GbGST*s protein from *G. biloba*. Using the ProtScale tool, the hydrophilicity of *G. biloba GbGST*s protein was predicted with 228 amino acids. The findings demonstrated that the majority of *G. biloba GbGST* protein sites with scores ranging from 2.311 to -2.244 were in the hydrophilic area as shown in Fig. (7).

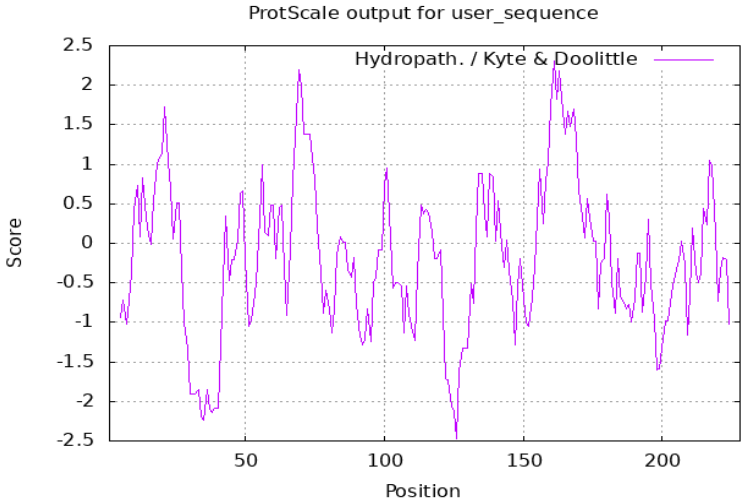

Fig. (7). Hydrophilicity profile of flavonoid 3-hydroxylase protein *GbGST*s from *Ginkgo biloba* using (https://web.expasy.org/protscale/).

Furthermore, it was concluded that the *GbGST*s protein is a hydrophilic protein, the parameters calculated by ProtParam software online were obtained with a calculated a molecular weight of about 25.786 kDa and *pI* of 6.23 and amino acid composition: 15 (A) alanine (Ala) 6.6%, 8 (R) arginine (Arg) 3.5%, 8(N) asparagine (Asn) 3.5%, 9 (D) aspartic acid (Asp) 3.9%, 1 (C) cysteine (Cys) 0.4%, 6 (Q) glutamine (Gln) 2.6%, 23 (E) glutamic acid (Glu) 10.1%, 18 (G) glycine (Gly) 7.9%, 2 (H) histidine (His) 0.9%, 16 (I) isoleucine (Ile) 7.0%, 26 (L) leucine (Leu) 11.4%, 23 (K) lysine (Lys) 10.1%, 5 (M) methionine (Met) 2.2%, 14 (F) phenylalanine (Phe) 6.1%, 11 (P) proline (Pro) 4.8%, 16 (S) serine (Ser) 7.0%, 5 (T) threonine (Thr) 2.2%, 4 (W) tryptophan (Trp) 1.8%, 7 (Y) tyrosine (Tyr) 3.1%, 11 (V) valine (Val) 4.8%, 0 (O) pyrrolysine (Pyl) 0.0%, and 0 (U) selenocysteine (Sec) 0.0%), Atomic composition: C: 1183, H: 1851, N: 297, O: 335, S: 6, extinction coefficients: 32430, estimated half-life: 30 hours (mammalian reticulocytes, *in vitro*). The instability index (II) is computed to be 39.87, Aliphatic index is 92.41, Grand average of hydropathicity (GRAVY) is -0.227. The findings suggest that *GbGST*s protein under study was hydrophobic in nature due to the presence of high non-polar residues content. GbGST protein has a high percentage of Leu (11.4%), Glu (10.1%), and Lys (10.1%). Results also showed that the maximum number of amino acid present in the sequence was found to be Leu (11.4%) and the least was for Cys (0.4%) and Trp (1.8%). The total number of negatively charged residues (Asp + Glu) is 32 and the total number of positively charged residues (Arg + Lys) is 3. On the other hand, the hydrophilicity of *G. biloba GbCHS* protein was predicted with 391 amino acids utilizing the program of ProtScale. The results showed that most sites of *G. biloba GbCHS* protein with a score of 2.211 to -2.600 in the hydrophilic region as shown in Fig. (8).

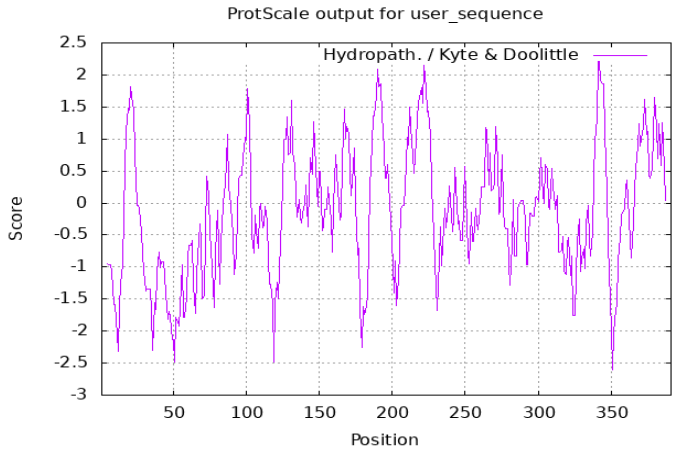

Fig. (8). Hydrophilicity profile of flavonoid 3-hydroxylase protein *GbCHS* from *Ginkgo biloba* using https://web.expasy.org/protscale/.

It was concluded that the *GbCHS* protein is a hydrophilic protein, the parameters computed by ProtParam software online were obtained with a molecular weight of about 43.078 kDa and *pI* of 6.04 and amino acid composition: 35 (A) Ala 9.0%, 16 (R) Arg 4.1%, 11(N) Asn 2.8%, 20 (D) Asp 5.1%, 8 (C) Cys 2.0 %, 11 (Q) 2.8%, 29 (E) Glu 7.4%, 31 (G) Gly 7.9%, 9 (H) His 2.3%, 20 (I) Ile 5.1%, 38 (L) Leu 9.7%, 28 (K) Lys 7.2%, 10 (M) Met 2.6%, 16 (F) Phe 4.1%, 22 (P) Pro 5.6%,21 (S) Ser 5.4%, 23 (T) Thr 5.9%, 6 (W) Trp 1.5%, 10 (Y) Tyr 2.6%, 27 (V) Val 6.9%, 0 (O) Pyl 0.0%, and 0 (U) Sec 0.0%), Atomic composition: C: 1931, H: 3043, N: 513, O: 566, S: 18, extinction coefficients: 48400, estimated half-life: 30 hours (mammalian reticulocytes, *in vitro*). The instability index (II) is computed to be 45.84 (this classifies this protein as unstable), the Aliphatic index is 86.83, Grand average of hydropathicity (GRAVY) is -0.161. The findings suggest that *GbCHS* protein under study was hydrophobic in nature due to the presence of high non-polar residues content. *GbCHS* protein has a high percentage of Leu 9.7%, Ala 9.0%, Gly 7.9%, Glu 7.4% and Lys 7.2%. Results also showed that the maximum number of amino acid present in the sequence was found to be Leu 9.7% and the least was for Trp 1.5% and Cys 2.0%. Total number of negatively charged residues (Asp + Glu): 69, total number of positively charged residues (Arg + Lys): 44.

3. Advanced Structure of *Ginkgo biloba* GbGST and GbCHS Protein

The main important task in the field of protein research was structure prediction from primary to advanced structure prediction. The three-dimensional structure model of *GbGST* and *GbCHS* protein from *G. biloba* were predicted by the Swiss-Model server, by homology modeling based on the available structures. Additionally, functional analysis of proteins by classification of protein families and predicting domains and important sites were provided by several different databases. Template search in either FASTA or Clustal format with the highest quality for the model building have then been selected from BLAST and HHBlits database has been performed against SWISS-MODEL Template Library (SMTL-ID) for evolutionary-related structures matching the target sequence. Furthermore, HHblits (a database of HMMs) first converts the query sequence (or MSA) to an HMM. This is conventionally done by adding pseudo counts of amino acids that are physic chemically similar to the amino acid in the query. For each identified template, the template's quality has been predicted from features of the target-template alignment. Models were built based on the target-template alignment using ProMod3 online. In case loop modeling with ProMod3 fails, an alternative model is built with PROMOD-II. For Model Quality Estimation: The global and per-residue model quality has been assessed using the QMEAN scoring function. Ligands present in the template structure are transferred by homology to the model. For Oligomeric State Conservation: The quaternary structure annotation of the template is used to model the target

sequence in its oligomeric form and based on other template features to provide a Quaternary Structure Quality Estimate (QSQE). QSQE score is a number between 0 and 1, reflecting the expected accuracy of the inter-chain contacts for a model built based on a given alignment and template. Higher numbers indicate higher reliability. This complements the Global Model Quality Estimation (GMQE) score which estimates the accuracy of the tertiary structure of the resulting model. The homologous sequence of *GbGST*s protein from *G. biloba* with more than 1107 templates is available in databases by name using the PDB ID format such as SMTL-ID: 7y55.1.A (for PdGST synthase from *Pinus densata*) with the bio-unit oligomeric state: homo-dimer, QMEAN: -0.86, GMQE: 0.89, sequence identify: 65.47%; sequence similarity: 0.51%; SMTL-ID: 4top.1.A (for 2,4-D inducible glutathione S-transferase (GmGST) synthase from *Glycine max*) with the bio-unit oligomeric state: monomer, QMEAN: 0.77, sequence identify: 52.56%, sequence similarity: 0.46% and other species with high homology and three-dimensional structure were one model built successfully as template alignment (Table 3).

Table (3). The top 8 filtered templates of protein from SMTL with high quality for GST model building with GbGSTs protein using X-ray according to the parameters (Sequence Identity, Oligo-state matching prediction), QSQE, found by (HHblits or BLAST, Resolution, Sequence Similarity, Coverage and Description). Further, more than 1107 templates were found which were less suitable for modeling than the filtered list.

No.	Template from	Seq. identity	Oligo state	QSQ	Found by	Resolution	Seq. similarity	cover	Description
1	7y55.1.A	65.47	homo-dimer	0.99	HHblits	2.19 Å	0.51	0.89	Glutathione S-transferase Tau1 from *Pinus densata*
2	7y55.1.A	66.52	homo-dimer	0.99	BLAST	2.19 Å	0.51	0.89	Glutathione S-transferase Tau1 from *Pinus densata*
3	4top.1.A	52.56	homo-dimer	1.00	HHblits	2.35 Å	0.46	0.77	Glutathione S-transferase *Glycine max*
4	4chs.1.A	52.31	homo-dimer	0.99	HHblits	1.6 Å	0.46	0.77	Tau class glutathione transferase 10 from *Glycine*
5	5g5a.1.A	56.34	homo-dimer	1.00	BLAST	1.95 Å	0.47	0.76	Glutathione transferase U25 from
6	6ep7.1.A	53.75	homo-dimer	0.98	HHblits	1.95 Å	0.45	0.75	Glutathione S-transferase U23 *Arabidopsis thaliana*
7	5j4u.1.A	55.50	homo-dimer	0.93	HHblits	1.25 Å	0.47	55.50	Glutathione S-transferase PtGSTU30 from
8	5j4u.1.A	55.19	homo-dimer	0.95	BLAST	1.25 Å	0.48	0.74	Glutathione S-transferase PtGSTU30 from

The template of matching 5 - 220 to *Pinus densata*, *Glycine max*, and other GSTs was used for homology modeling since the N and C terminals of *GbGST*s from *G. biloba* have weak homology to less than 4 amino acids. The

findings were near to the protease real space conformation, as illustrated in Fig. (9 and 10).

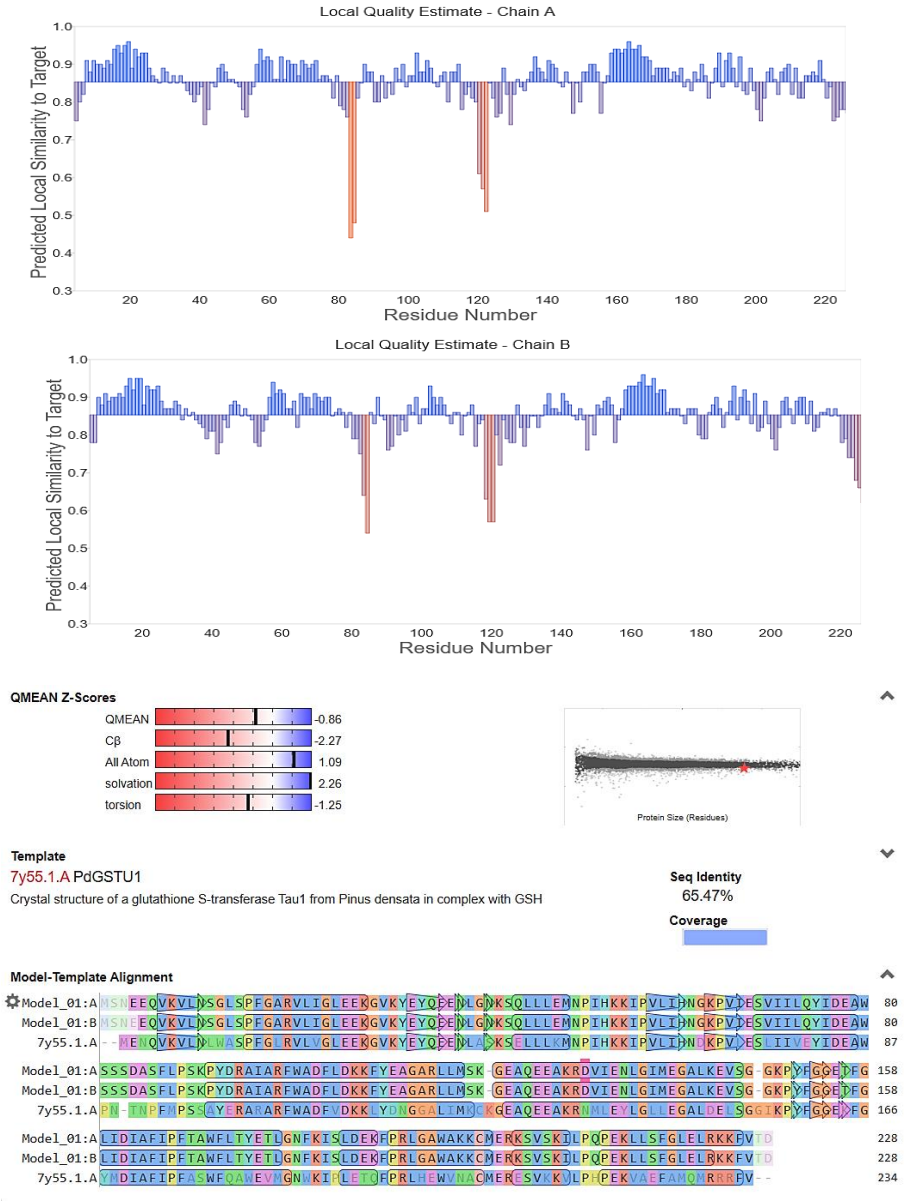

Fig. (9). Homology model of *GbGSTs* from *Ginkgo biloba* showing quality estimate (global quality estimate, local quality estimate, comparison non-redundant set of PDB structures) and model-template alignment with STML-ID: 7y55.1A form *Pinus densata.*

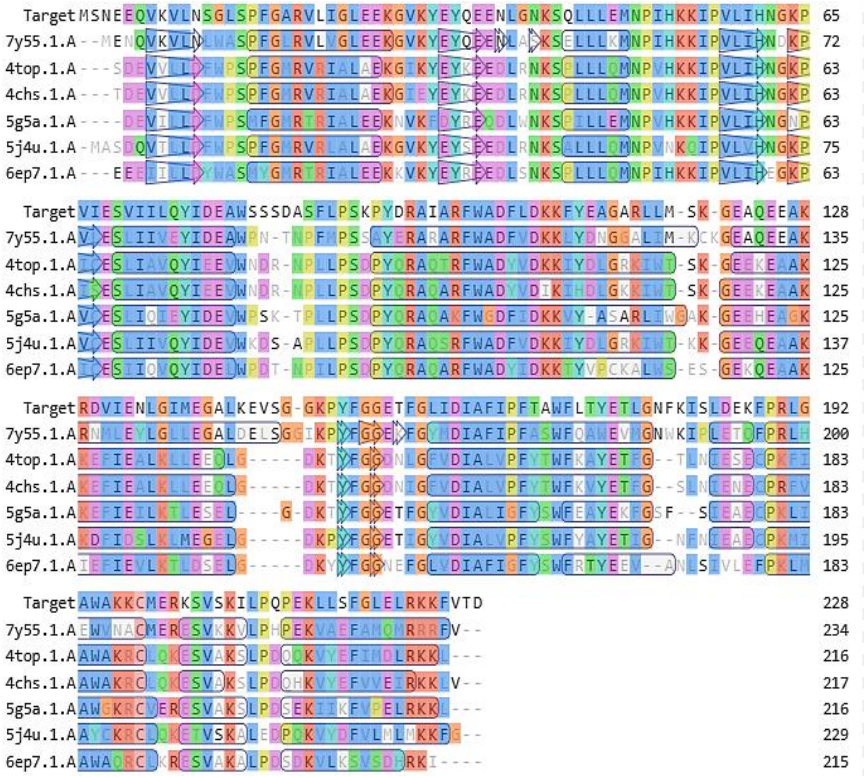

Fig. (10). Sequence alignment and homology modeling protein analyses. An amino acid sequence alignment of *GbGST*s (Target) and *GST*s.

On other hand, the homologous sequence of *GbCHS* protein from *G. biloba* with more than 1605 templates available in databases by named using the PDB-ID format such as SMTL-ID: 6dxa.1.A (for chalcone synthase from *Pinus sylvestris* with Biounit oligomeric state: homo-dimer, none ligands, QMEAN: 0.87, GMQE: 0.91, sequence identify: 85.31% and sequence similarity: 0.57%. Also with SMTL-ID: 5uc5.1.A (for chalcone synthase (MdCHS) from *Malus domestica* with biounit oligomeric state: homo-dimer, none ligands, QMEAN: 0.87, GMQE: 0.91, sequence identify: 81.91%; sequence similarity: 0.55% and other species with high homology and three-dimensional structure were one model built successfully as template alignment (Table 4).

Because the N and C terminal of *G. biloba* CHS protein is poor homology to *Pinus sylvestris, Malus domestica* and other than 4 amino acids, the template of matching 4 - 383 and some amino acid inside not conserve sequence to *them* CHSs protein was selected for homology modeling as shown in Fig. (11 and 12). The results were close to the protease real space conformation. Local estimates of the model quality based on the QMEAN

scoring function are shown as a per-reside plot and as a global score concerning a set of high-resolution PDB structures (Z-score). Based on the results obtained, the homology model can be considered a reliable model. A high similarity was observed for *GbGST*s and *GbCHS* protein with model template alignment, but the N terminal and C terminal regions showed some variability in length and composition. It was clear from the multi-sequence alignment that *GbGST*s and *GbCHS* protein from gymnosperm plants were more like each other than to those of angiosperm plants, as confirmed by the phylogenetic analysis. The stringent conservation among evolutionary diverse plant species may indicate the functional significance of these amino acids. Additionally, Homology modeling was used as a useful tool for the prediction of protein structure when the model protein with a known sequence and an unknown structure is related with high/identify to at least one other protein with both a known sequence and a known structure.

Table (4). The top 8 filtered templates for protein from (SMTL) with high quality for (CHS) chalcone synthase model building with *GbCHS* protein using X-ray according to the parameter (Sequence Identity, Oligo-state matching prediction), QSQE, found by (HHblits or BLAST, resolution, sequence similarity, coverage and description). Further, more than 1605 templates were found which were less suitable for modeling than the filtered list.

No.	Template from SMTL-ID	Seq. identity	Oligo state	QSQ	Found by	Resolution	Seq. Similarity	Cover	Description
1	6dxa.1.A	85.31	homo-dimer	1.00	BLAST	2.01 Å	0.57	0.93	Chalcone synthase from *Pinus sylvestris*
2	5uc5.1.A	82.17	homo-dimer	0.96	BLAST	2.10 Å	0.55	0.93	Chalcone synthase from *Malus domestica*
3	5uc5.1.A	81.91	homo-dimer	0.96	HHblits	2.10 Å	0.55	0.93	Chalcone synthase from *Malus domestica*
4	4wum.1.A	79.84	homo-dimer	0.97	BLAST	1.77 Å	0.55	0.92	Chalcone Synthase from *Freesia hybrida*
5	6dxa.1.A	84.97	homo-dimer	1.00	HHblits	2.01 Å	0.51	0.93	Chalcone synthase from *Pinus sylvestris*
6	7bur.1.A	82.12	homo-dimer	0.93	BLAST	1.82 Å	0.56	0.93	Chalcone synthase from *Glycine max (L.)*
7	6dxb.1.A	69.43	homo-dimer	0.94	BLAST	1.55 Å	0.55	0.92	Chalcone synthase from *Arabidopsis thaliana*
8	4yjy.1.A	76.62	homo-dimer	0.89	BLAST	1.86 Å	0.54	0.92	Chalcone synthase 1 from *Oryza sativa*

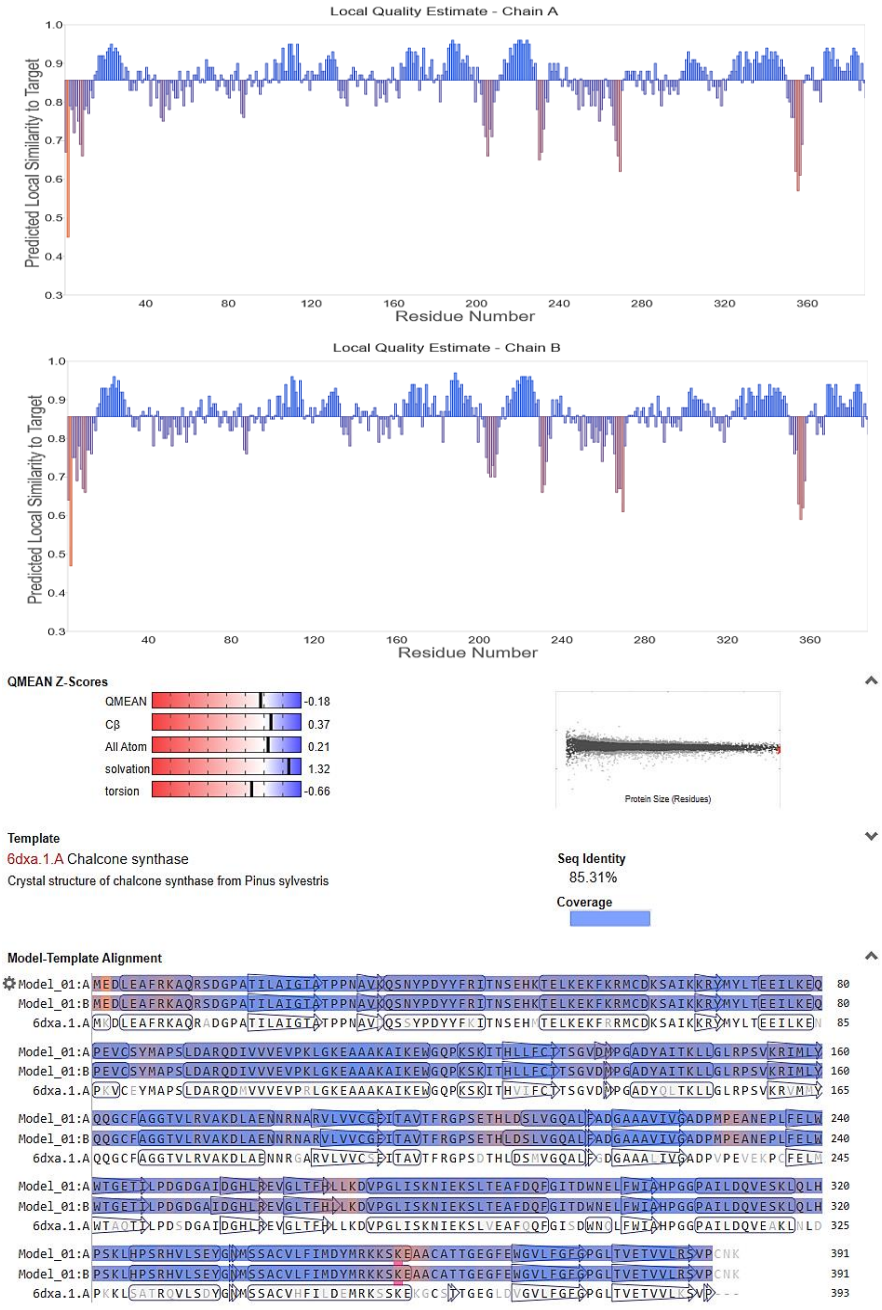

Fig. (11). Homology model of chalcone synthase (*GbCHS*) from *Ginkgo biloba* showing quality estimate (global quality estimate, local quality estimate, comparison non-redundant set of PDB structures) and model-template alignment with STML-ID: 6dxa.1A from *Pinus sylvestris.*

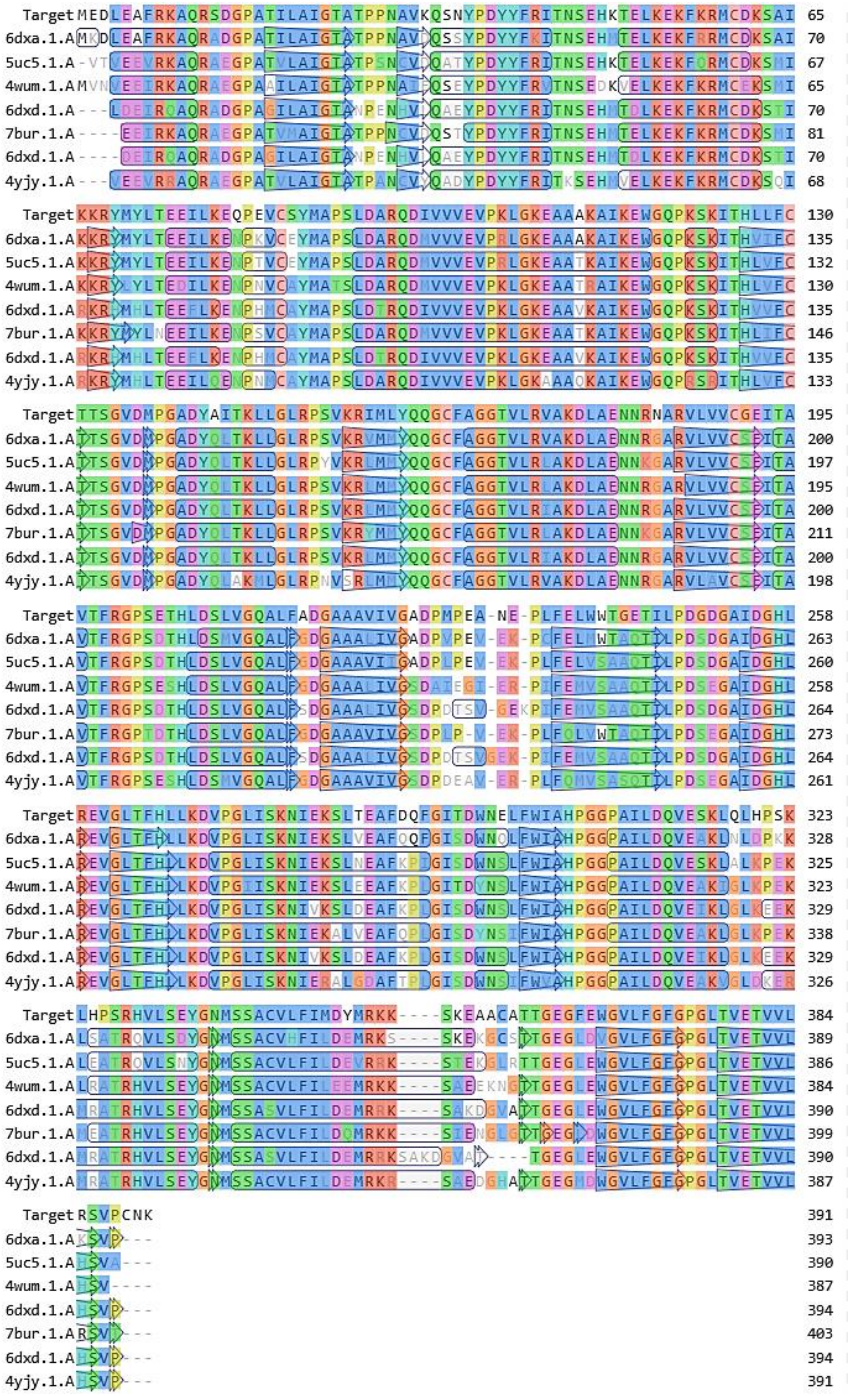

Fig. (12). Sequence alignment and homology modeling protein analyses, an amino acid sequence alignment of *GbCHS* (Target) and CHSs.

DISCUSSION

Anthocyanins are produced in the cytoplasm by flavonoid metabolic pathways and subsequently transported to vesicles for storage (Gu et al., 2019). Moreover, *G. biloba*'s main physiologically active ingredients are flavonoids, which are essential for neutralizing free radicals and protecting plant maturation and development (Li et al., 2009). Also, the most of flavonoids in *G. biloba* are in glycosylated types (Liu et al., 2015). Flavonoids can also increase a plant's tolerance to biotic and abiotic stress. Recent research has shown that flavonoids and some biotic, abiotic stress have a strong relationship (Walia et al., 2005). In earlier investigations, it was discovered how anthocyanin is transported intracellularly, *GST* mediation, membrane transport, or vesicle trafficking are all necessary for anthocyanins to enter vacuoles from the cytoplasm (Zhao, 2015). In recent years, *GST*s have been found in a broad range of plant types, and the number and make-up of the *GST* family vary from plant to plant (Fang et al., 2020). Additionally, being an essential part of plant growth and development, CHS is an essential enzyme in the production of derivatives of the flavonoid. Most *CHS* genes were found to be grouped into two or more subfamilies, and all *GbCHS*s were classified into five classes based on their evolutionary relationships, according to molecular evolution studies.

Many researchers have identified and investigated the expression of the flavonoid biosynthesis genes using the *CHS* gene as a model (Kubasek et al., 1998). The distribution of motif types and the examination of gene structure both supported the phylogenetic relationship. Moreover, the majority of *CHS* genes have two exons and one intron (Durbin et al., 2000). For instance, 3 exons and 2 introns constitute the *GbCHI* gene from the *CHS*s families, which codes for a 244 amino acid peptide with an estimated molecular weight of 26.29 kDa (Cheng et al., 2011). Additionally, most of the *GbCHS* genes were found to have two exons and one intron, and 20 motifs were indeed detected. In addition, 12 *GbCHS* genes included 25 pairs of duplicated events, including 23 pairs of segmental duplications and 2 pairs of tandem duplications. This suggests that segmental duplication events are the main cause of the *GbCHS* gene family's sluggish evolution (Kong et al., 2020). The structural biology and the 3D structures produced by the current work can be useful for further research into the distribution of amino acid residues in each fold, the prediction of active sites, the molecular mechanism of function, and structure-based phylogeny. For predicting protein function, structural information is frequently more important than sequencing alone (Waterhouse et al., 2009). Consequently, the degree of similarity between the model and template sequences determines the quality of the projected structure that is achieved by homology modeling. Homology modeling of the query protein does not produce a significant result if the similarity was very low. The biosynthesis pathways of flavonoids from *G. biloba* were important and

involved in homology modeling and bioinformatics analysis of *GbGST*s and *GbCHS* genes. *GST*s and other *CHS*s had a common evolutionary origin, according to phylogenetic analysis, and they also had a relationship with other angiosperm species. Even though the *G. biloba* genome contains a sizable number of potential *GbCHS* and *GbGST*s genes, only a few of these genes have been functionally described. We will learn more about the evolution of flavonoids in the plant world by analyzing the *G. biloba* genome sequence that is now accessible.

One of the common categories of secondary plant metabolites known as flavonoids offers several health advantages. The precise mechanism of flavonoid production in plants is still mostly unknown, in many plants, the enzyme class known as chalcone isomerase (*CHI*) plays an essential role in the metabolism of flavonoids. A new CHI from safflower that encodes 217 amino acids was identified in its full-length cDNA (1161 bp), *CtCHI* is extremely similar to other plants, including the characteristic polyadenylation signals AATAA and Poly A tail, according to the results of Sanger sequencing and phylogenetic research (Liu et al., 2019). Gene architectures, motifs, and the distribution of WD40 genes across chromosomes are examined. Five *GbWD40* genes may be involved in the control of flavonoid synthesis in *G. biloba*, according to promoter analysis, which reveals that their promoters contain the MYB binding site involved in the regulation of flavonoid metabolism (Zheng et al., 2021). In addition, The *G. biloba* plant's GbMADS gene architectures varied within the same gene family or subfamily, but conserved protein motifs were distributed uniformly throughout (Yang et al., 2022). A significant increase in the accumulation of flavonol glycosides and higher expression of many genes involved in the synthesis of flavonoids, including *CHS*, *FLS*, *F3'H*, *DFR*, and *LAR*, was also found utilizing metabolomics analysis. By overexpressing *GbCHS* in *G. biloba calli*, its major contribution to flavonoid synthesis was demonstrated (Lu et al., 2022).

CONCLUSION

In this study, *GbGST*s and *GbCHS* cDNA genes from *G. biloba* were extracted and sequenced. Moreover, MSA was performed using 30 amino acid sequences from *GbGST*s and *GbCHS* genes, showing high identity and similarity between the two gene families. The amino acid sequences of *GbGST* and *GbCHS* were used in a phylogenetic study together with other known plant-specific *GST*s and *CHS*s. Furthermore, two *G. biloba* genes (*GbGST*s and *GbCHS*) have been studied for their enzymatic activity using homology modeling and structural analysis. Additionally, *GbGST*s and *GbCHS* theoretical 3D models were predicted using homology modeling to show ligands, global quality estimate, local quality estimate, sequence identity percentage, and model template alignments. Finally, the findings suggested that the first step in completely understanding the regulatory mechanisms

governing flavonoid and anthocyanin biosynthesis in *G. biloba* is molecular identification, phylogenetic analysis, homology modeling, and structure analysis predictions of many genes encoding important enzymes.

REFERENCES

Bertoni, M., F. Kiefer and M. Biasini (2017). Modeling protein quaternary structure of homo-and hetero-oligomers beyond binary interactions by homology. Scientific Reports, 7, Article no.: 10480.

Camacho, C., G. Coulouris and V. Avagyan (2009). BLAST^{+}: architecture and applications. BMC Bioinformatics, 10: 421-430.

Cao, F., H. Cheng, S. Cheng, L. Li., F. Xu, W. Yu and H. Yuan (2012). Expression of selected *Ginkgo biloba* heat shock protein genes after cold treatment could be induced by other abiotic stress. International Journal of Molecular Sciences, 13 (5): 5768-5788.

Cheng, H., L. Li, S. Cheng, F. Cao, Y. Wang and H. Yuan (2011). Molecular cloning and function assay of a chalcone isomerase gene (*GbCHI*) from *Ginkgo biloba*. Plant Cell Reports, 30: 49-62.

Csiszár, J., E. Horváth, Z. Váry, Á. Gallé, K. Bela, S. Brunner and I. Tari (2014). Glutathione transferase supergene family in tomato: salt stress-regulated expression of representative genes from distinct GST classes in plants primed with salicylic acid. Plant Physiology and Biochemistry, 78: 15-26.

Dixon, D.P., A. Lapthorn and R. Edwards (2002). Plant glutathione transferases. Genome Biol., 3 (3): reviews3004.

Durbin, M.L., B. Mccaig and M.T. Clegg (2000). Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol., 42: 79–92.

Edwards, R and D.P. Dixon (2005). Plant Glutathione Transferases. In: 'Sies, H. and L. Packer (Eds.)'. Methods Enzymol; Academic Press: Cambridge, MA, USA, pp. 169–186.

Efferth, T. and E. Koch (2011). Complex interactions between phytochemicals. The multi-target therapeutic concept of phytotherapy. Current Drug Targets, 12 (1): 122-132.

Fang, X, Y. An, J. Zheng, L. Shangguan and L. Wang (2020). Genome-wide identification and comparative analysis of GST gene family in apple (*Malus domestica*) and their expressions under ALA treatment. 3 Biotech, 10 (307): 1-16.

Gasteiger, E., C. Hoogland and A. Gattiker (2005). Protein identification and analysis tools on ExPASy server, In: 'Walker, J.M. (Ed.)'. The Proteomics Protocols Handbook, Humana Press, pp. 571-607.

Gu, K.D., C.K. Wang, D.G. Hu and Y.J. Hao (2019). How do anthocyanins paint our horticultural products? Sci. Hortic., 249: 257–262.

Kiba, A., K. Toyoda, Y. Ichinose, T. Yamada and T. Shiraishi (1995). Specific inhibition of cell wall-bound ATPases by fungal suppressor from *Mycosphaerella pinodes*. Plant and Cell Physiology, 36 (5): 809-817.

Kitamura, S., N. Shikazono and A. Tanaka (2004). TRANSPARENT TESTA 19 is involved in the accumulation of both anthocyanins and proanthocyanidins in *Arabidopsis*. The Plant Journal, 37 (1): 104-114.

Kong, X., A. Khan, Z. Li, J. You, F. Munsif, H. Kang and R. Zhou (2020). Identification of chalcone synthase genes and their expression patterns reveal pollen abortion in cotton. Saudi Journal of Biological Sciences, 27 (12): 3691-3699.

Ku, Y.S., M.S. Ng, S.S. Cheng, A.W. Lo, Z. Xiao, T.S. Shin and H.M. Lam (2020). Understanding the composition, biosynthesis, accumulation and transport of flavonoids in crops for the promotion of crops as healthy sources of flavonoids for human consumption. Nutrients, 12 (6): 1717.

Kubasek, W.L., F.M. Ausubel and B.W. Shirley (1998). A light-independent developmental mechanism potentiates flavonoid gene expression in *Arabidopsis* seedlings. Plant Molecular Biology, 37 (2): 217.

Kyte, J. and R.F. Doolittle (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157: 105-132.

Lan, T., Z.L. Yang, X. Yang, Y.J. Liu, X.R. Wang and Q.Y. Zeng (2009). Extensive functional diversification of the Populus glutathione S-transferase supergene family. Plant Cell, 21 (12): 3749–3766.

Li, L., J.D. Stanton, A.H. Tolson, Y. Luo and H. Wang (2009). Bioactive terpenoids and flavonoids from *Ginkgo biloba* extract induce the expression of hepatic drug-metabolizing enzymes through pregnane X receptor, constitutive androstane receptor, and aryl hydrocarbon receptor-mediated pathways. Pharmaceutical Research, 26: 872-882.

Liu, X.G., S.Q. Wu, P. Li and H. Yang (2015). Advancement in the chemical analysis and quality control of flavonoid in *Ginkgo biloba*. Journal of Pharmaceutical and Biomedical Analysis, 113: 212-225.

Liu, X., N. Ahmad, L. Yang, T. Fu, J. Kong, N. Yao and H. Li (2019). Molecular cloning and functional characterization of chalcone isomerase from *Carthamus tinctorius*. Amb. Express, 9: 1-12.

Lu, Z., L. Zhu, J. Lu, N. Shen, L. Wang, S. Liu and J. Lin (2022). Rejuvenation increases leaf biomass and flavonoid accumulation in *Ginkgo biloba*. Horticulture Research, 9 (18): 1-18.

Manaf, A (2013). Molecular cloning and expression profiling of a chalcone synthase gene from *Ginkgo biloba*, *Scutellaria viscidula* Bunge and Tree peony (*Paeonia suffruticosa*). International Journal of Biosciences, 3 (10): 187-202.

Mouradov, A. and G. Spangenberg (2014). Flavonoids: a metabolic network mediating plants adaptation to their real estate. Front. Plant Sci., 5: 620.

Ni, J., J. Hao, Z. Jiang, X. Zhan, L. Dong, X. Yang, Z. Sun, W. Xu, Z. Wang and M. Xu (2017). NaCl Induces flavonoid biosynthesis through a putative novel pathway in post-harvest *Ginkgo* leaves. Front. Plant Sci., 8: 920.

Pang, Y., G.A. Shen, C. Liu, X. Liu, F. Tan, X. Sun and K. Tang (2004). Molecular cloning and sequence analysis of a novel chalcone synthase cDNA from *Ginkgo biloba* DNA Sequence. Journal of DNA Sequencing and Mapping, 15: 283-290.

Remmert, M., A. Biegert, A. Hauser and J. Söding (2011). HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nature Methods, 9: 173-175.

Schwede, T., J. Kopp, N. Guex and M.C. Peitsch (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucl. Acids Res., 31: 3381–3385.

Steinegger, M., M. Meier and M. Mirdita (2019). HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinformatics, 20: 473.

Studer, G., C. Rempfer and A.M. Waterhouse (2020). QMEANDisCo - distance constraints applied on model quality estimation. Bioinformatics, 36: 1765-1771.

Tanaka, Y and F. Brugliera (2013). Flower colour and cytochromes P450. Philos. Trans. R. Soc. Lond B. Biol. Sci., 368 (1612): 20120432.

Vaish, S., D. Gupta, R. Mehrotra, S. Mehrotra and M.K Basantani (2020). Glutathione S-transferase: A versatile protein family. 3 Biotech, 10: 321.

Wagner, U., R. Edwards, D.P. Dixon and F. Mauch (2002). Probing the diversity of the *Arabidopsis* glutathione S-transferase gene family. Plant Mol. Biol., 49: 515–532.

Walia, H., C. Wilson, P. Condamine, X. Liu, A.M. Ismail, L. Zeng and T. J. Close (2005). Comparative transcriptional profiling of two contrasting rice genotypes under salinity stress during the vegetative growth stage. Plant Physiology, 139 (2): 822-835.

Wang, L., J. Cui, B. Jin, J. Zhao, H. Xu, Z. Lu, W. Li, X. Li, L. Li and E. Liang (2020). Multifeature analyses of vascular cambial cells reveal longevity mechanisms in old *Ginkgo biloba* trees. Proc. Natl. Acad. Sci. USA, 117 (4): 2201–2210.

Waterhouse, A.M., J.B. Procter and D.M.A. Martin (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25: 1189-1191.

Wilkins M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel and A. Bairoch (1999). Protein Identification and Analysis

Tools in the Expasy Server. In: 'Clifton, N.J. (Ed.)'. Methods in Molecular Biology, 112: 531-552.

Xu, F., Y. Ning, W. Zhang, Y. Liao, L. Li and H. Cheng (2014). An R2R3-MYB transcription factor as a negative regulator of the flavonoid biosynthesis pathway in *Ginkgo biloba*. Funct. Integr. Genomics, 14: 177–189.

Yan, J., S. Zhang, M. Tong, J. Lu, T. Wang, Y. Xu, W. Li and L. Wang (2021). Physiological and genetic analysis of leaves from the resprouters of an old *Ginkgo biloba* tree. Forests, 12 (9): 1255.

Yang, K., Z. Liu, X. Chen, X. Zhou, J. Ye, F. Xu and Q. Wang (2022). Genome-Wide identification and expression analysis of the mads-box family in *Ginkgo biloba*. Forests, 13 (11): 1953.

Zhao, J. (2015). Flavonoid transport mechanisms: how to go, and with whom. Trends in Plant Science, 20 (9): 576-585.

Zhao, J., J. Yang and Y. Xie (2019). Improvement strategies for the oral bioavailability of poorly water-soluble flavonoids: An overview. Int. J. Pharm., 570: 118642.

Zheng, J., L.I. Yongling, X.U. Feng, Z.H. Xian, Y.E. Jiabao, F.U. Mingyue and W. Zhang (2021). Genome-wide identification of WD40 superfamily genes and prediction of WD40 gene of flavonoid-related genes in *Ginkgo biloba*. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 49 (2): 12086-12086.

Zhou, Z. and S. Zheng (2003). The missing link in Ginkgo evolution. Nature, 423 (6942): 821–822.

Zhou, T., X. Yang, F. Fu, G. Wang and F. Cao (2020). Selection of suitable reference genes based on transcriptomic data in *Ginkgo biloba* under different experimental conditions. Forests, 11 (11): 1217.

التوصيف الجزيئي لجينات *GbCHS* وGbGSTs المشاركة في التخليق الحيوي لمركبات الفلافونويد في الجنكة

مختار سعيد رزق
قسم الأصول الوراثية، مركز بحوث الصحراء، المطرية، القاهرة، مصر

يعرف نبات الجنكة أنه من عاريات البذور وله قيمة طبية وبيولوجية واقتصادية في العالم ككل. في هذه الدراسة تم التوصيف الجزيئي باستخدام المعلوماتية الحيوية على أساس نمذجة التماثل بواسطة جينات *GbCHS* and *GbGST*s حيث أن هذه الجينات تلعب دورًا هامًا في التخليق الحيوي لمركبات الفلافونويد والأنثوسيانين وغيرها من المستقبلات الثانوية المهمة في النباتات. أشارت النتائج لعزل جين *GbGST*s كامل التتابع منcDNA لنبات الجنكة بطول ٦٨٧ من ازواج القواعد النيوكلوتيدية والتى نشفر إلى ٢٢٨ حمض أميني بوزن ٢٥.٧٨٦ كيلو دالتون، بينما تم عزل جين *GbCHS* كامل التتابع من cDNA بطول ١١٧٦ من أزواج القواعد النيوكليوتيدية والتي تشفر إلى ٣٩١ حمض أميني بوزن جزيئي ٤٣.٠٧٨ كيلو دالتون. بالإضافة إلى ذلك، تم تحليل تتابعات المحاذاة المتعددة والعلاقات التطورية لتتابعات الأحماض الأمينية (٣٠ حمض أميني) لـ *GbCHS* و*GbGST*s وذلك باستخدام برنامج MEGA7، حيث وجد تطابق عالي مع تلك الأحماض الأمينية للنباتات المتحصل عليها من قاعدة بيانات NCBI، كما تم استخدام برنامج ProtParam لتحديد الوزن الجزيئي وتم استخدام برنامج GRAVY لتحديد الجزيئات المحبة والكارهة للماء، وبرنامج Protscale لتحديد الجزيئات المحبة والكارهة للماء على أساس قياسات مختلفة. أخيرًا يعد التوصيف الجزيئي وتحليل المعلوماتية الحيوية لجينات GbCHS و*GbGST*s والتي تشفر إلى إنزيمات تعتبر المفتاح الرئيسي كخطوة أولى لفهم الآليات التنظيمية التي تتحكم في التخليق الحيوي للفلافونويد والأنثوسيانين في الجنكة.