THAI NGUYEN UNIVERSITY

UNIVERSITY OF EDUCATION

LE THI HONG TRANG

STUDY THE EXPRESSION OF *GmCHI1A* GENES IN RELATION TO THE SYNTHESIS OF ISOFLAVONES ISOLATED FROM SOYBEAN PLANTS [*Glycine max* (L.) Merill]

Speciality: Genetics

Code: 9420121

DISSERTATION SUMMARY

THAI NGUYEN - 2020

The dissertation was finished at:

THAI NGUYEN UNIVERSITY – UNIVERSITY OF EDUCATION

Supervisor: Prof.Dr. Chu Hoang Mau

Reviewer 1: …….

Reviewer 2: ……..

Reviewer 3: ……..

The dissertation will be defended in the university committee:

THAI NGUYEN UNIVERSITY – UNIVERSITY OF EDUCATION

At ……, 2020

The dissertation can be read at:

1. National Library of Vietnam.
2. Thai Nguyen University - Learning Resource Center
3. Library of University of Education

THE AUTHOR'S PUBLICATIONS RELATED TO THE DISSERTATION TOPIC

1. Huu Quan Nguyen, Thi Hong Trang Le, Thi Ngoc Lan Nguyen, Thu Giang Nguyen, Danh Thuong Sy, Quang Tan Tu, Thi Thu Thuy Vu, Van Son Le, Hoang Mau Chu, Thi Kim Lien Vu (2020), "Overexpressing *GmCHI1A* increases the isoflavone content of transgenic soybean (*Glycine max* (L.) Merr.) seeds", *In Vitro Cellular & Developmental Biology-Plant,* (SCIE, Q2) **https://doi.org/10.1007/s11627-020-10076-x**
2. Le Thi Hong Trang, Chu Hoang Mau, Nguyen Huu Quan (2019), "Agrobacterium – mediated transformation with *Glycine max chalcone isomerase 1A* gene in tobacco: a model for overexpression of *GmCHI1A* gene in soybean plants", Journal of Science & Technology of Thai Nguyen University Volume 207 (14), page 195-200.
3. Le Thi Hong Trang, Ho Manh Tuong, Le Van Son, Chu Hoang Mau (2018), "Design of plant transgenic vectors carrying *GmCHI* gene isolated from soybean planst", Proceedings of the National Biotechnology Conference 2018, Publishing House of Natural Sciences and Technology p. 83-87.
4. Le Thi Hong Trang, Tran Thi Thanh Van, Ho Manh Tuong, Pham Thanh Tung, Le Van Son, Chu Hoang Mau (2016), "The characteristics of *GmCHI* gene isolated from soybean cultivars with different isoflavone content", Journal of Biology 38 (2), p. 236-242.

The gene sequences registered on the International Gene Bank

1. Le,T.H.T., Ho,T.M., Hoang,H.P., Le,S.V. and Chu,M.H.(2016), "*Glycine max* mRNA for chalcone isomerase RNA (chalcone isomerase(CHI) gene), cultivar DT26", *GenBank:* LT594994.1.
2. Le,T.H.T., Ho,T.M., Hoang,H.P., Le,S.V. and Chu,M.H.(2016), **"***Glycine max* mRNA for chalcone isomerase RNA (chalcone isomerase (CHI) gene), cultivar DT51**",** *GenBank*: LT594995.1.
3. Le,T.H.T., Ho,T.M., Hoang,H.P., Le,S.V. and Chu,M.H.(2016), **"***Glycine max* mRNA for chalcone isomerase RNA (chalcone isomerase (CHI) gene), cultivar DT84**",** *GenBank:* LT594993.1**.**
4. Le,T.H.T., Ho,T.M., Hoang,H.P., Le,S.V. and Chu,M.H.(2016), **"***Glycine max* mRNA for chalcone isomerase RNA (chalcone isomerase (CHI) gene), cultivar DT2008**",** *GenBank:* LT594996.1.

INTRODUCTION

1. Problem statement

Flavonoids are an important natural product that helps protect plants and human health. Isoflavones are a type of flavonoid, abundant in soybean seeds and exhibit antioxidant, anti-cancer, antibacterial and anti-inflammatory properties. Soybean isoflavones are easy to use for humans, while some compounds with the same composition like isoflavones in clover, alfalfa, arrowroot are very difficult to use.

Isoflavones are synthesized from a branch of the phenylpropanoid pathway. Isoflavone synthesis involves many enzymes, of which CHI is the key enzyme that catalyzes the reaction of open-chain naringenin chalcone to be closed to form naringenin. Naringenin is converted into many main flavonoids such as flavanone, flavonol and anthocyanin. CHIs are classified into two main types, CHI type I and CHI type II. Type I CHIs are found in most plants, but Type II CHIs are found only in legumes. The *GmCHI1A* gene in soybean belongs to CHI type II located on chromosome 20, which encodes the CHI1A enzyme. Research results of CHI gene expression confirmed that the overexpression of *CHI* gene increased the total isoflavonoid content in transgenic plants many times compared to non-transgenic plants. Thus, the action on CHI enzyme can increase the accumulation of isoflavones and other flavonoids. So far, only Lyle Ralston et al (2005) studied on *GmCHI* gene expression in yeast and Vu et al (2018) analyzed *GmCHI1A* gene expression in *Talinum paniculatum*; there is no research addressing the results of *GmCHI1A* gene expression analysis in soybean plants in the direction of creating a transgenic line with high isoflavone content.

Soybean (*Glycine max* (L.) Merrill) is an important crop in the agricultural production of many countries around the world. Soybean seeds have high nutritional value. In addition, soybean is also a crop of economic value and is a soil improvement crop. It is worth noting that soybean seeds contain isoflavones, especially aglucones, which are

quickly absorbed by the human digestive system, but the content is very low. This is the reason for the research interest in improving the isoflavone content in soybean seeds.

Based on the above issues, we selected and conducted the research project: ***"Study the expression of GmCHI1A genes in relation to the synthesis of isoflavones isolated from soybean plants (Glycine max (L.) Merill)"*** to clarify the relationship between the enhanced *GmCHI1A* gene expression and an increase in isoflavone content in transgenic soybean seed germs.

2. Research objectives

Express *GmCHI1A* genes in transgenic soybean and create *GmCHI1A* transgenic soybean lines with higher isoflavone content than non-transgenic control plants.

3. Research contents

3.1. Study the characteristics of *GmCHI1A* gene in soybean plants

i) Investigate isoflavone content of some common soybean cultivars in Northern Vietnam.

ii) Study information on *GmCHI* gene of soybean, design PCR primers and duplicate the *GmCHI1A* coding segment from high isoflavone soybean cultivars.

iii) Clone, sequence nucleotides and analyze the characteristics of GmCHI1A gene isolated from soybean

3.2. Design plant transgenic vector carrying the *GmCHI1A* gene and evaluate the performance of the designed transgenic vectors.

3.3. Analyse GmCHI1A gene expression in transgenic soybean

i) Study on the transfer of *GmCHI1A* transgene structure into DT2008 soybean cultivar.

ii) Analyze the incorporation of the *GmCHI1A* transgene into the genome of soybean by PCR and Southern blot.

iii) Analyze the expression of GmCHI1A recombinant protein in transgenic soybean using Western blot and ELISA.

iv) Evaluate the change in isoflavone content in *GmCHI1A* transgenic plants compared to the non-transgenic control.

4. New contributions of the thesis

The thesis is a new research project in Vietnam and in the world that has demonstrated that the overexpression of *GmCHI1A* gene can increase isoflavone content in transgenic soybean seed germs. The thesis is a systematic project with the contents presented from gene isolation to design of transgenic vector, analysis of gene expression and creation of high-isoflavone transgenic lines.

Specifically:

1) The *GmCHI1A* gene isolated from Vietnamese soybean is 657 nucleotides in the size of the coding region, encodes 218 amino acids, belongs to subfamily II, and is located on chromosome 20 of soybean.

2) For the first time, the expression of *GmCHI1A* gene was analyzed and the overexpression of the *GmCHI1A* transgene increased the content of CHI enzyme in soybean.

3) Four T2 generation transgenic soybean lines were created with daidzein content increasing from 166.46% to 187.23% and genistein content increasing from 329.80% to 463.93% compared to that of non-transgenic plants.

5. Scientific and practical significance of the thesis topic

The scientific results of the thesis have shown that the overexpression of the gene encoding the key enzyme in the isoflavone biosynthesis pathway of soybean has increased the isoflavone content in soybean seed germs. The results of the study are the scientific basis for improving the content of secondary compounds in plants by gene expression techniques.

The research results published in scientific papers and gene sequences registered on GenBank are valuable references in research and teaching.

Practically, *GmCHI1A* transgenic soybean lines can be used as materials to select high-isoflavone soybean cultivars. The results of the thesis can be applied to legumes and other plant species in the direction of improving isoflavone content in seed germs to do research on functional foods for community health care.

6. The structure of the thesis

The thesis has 139 pages (including appendices), divided into chapters and sections: Introduction (5 pages); Chapter 1: Literature Review (36 pages); Chapter 2: Materials and Research methods (15 pages); Chapter 3: Results and Discussion (43 pages); Conclusions and Recommendations (2 pages); Published works related to the thesis (2 pages); References (16 pages); Appendixes (6 pages). The thesis has 14 tables, 36 pictures, 3 appendices, 126 references documents and some websites.

Chapter 1. LITERATURE REVIEW

The thesis has consulted and summarized 126 documents and some websites, including 17 Vietnamese documents, 109 English documents on three basic issues, namely: (1) Soybean and isoflavones in soybean seeds; (2) CHI Enzyme and CHI encoding gene; (3) Transgene in soybean and CHI gene expression analysis.

Soybean seeds (*Glycine max* (L.) Merrill) contain high content of protein and lipid, lots of non-replaceable amino acids, mineral salts Ca, Fe, Mg, P, K, Na and vitamins B1, B2 , C, E, K ... necessary for human and animal bodies. It is worth noting that soybean seeds contain isoflavones. Isoflavones are secondary metabolites with diverse biological functions. Isoflavones and compounds similar to isoflavones are found in soybeans and some plants such as clover, alfalfa, arrowroot, etc. Isoflavones in soybean are easy to use for humans, while isoflavones derived from other plants are difficult to use. Isoflavones in soybean have antioxidant and anti-cancer activities, prevent cardiovascular diseases, improve women's health and can positively impact other physiological processes. The

isoflavone content in soybeans is low, so the research direction to improve the isoflavone content in soybeans, especially in seeds, is a matter of concern.

The content of isoflavones in soybean seeds is relatively low, about 50 - 3000 μg/g and exists in two main forms: β-glucoside (daidzin, genistin, glycitin) and aglucone (daidzein, genistein, glycitein). The glycoside form, which has a large molecular weight, can be limitedly absorbed in the human digestive system, while the aglucone form can be absorbed faster, but the content is very low. Isoflavones are synthesized from the phenylpropanoid pathway found in all plants, and chalcone isomerase (CHI) is an important enzyme because it catalyzes the reaction of naringenin chalcone and open-chain isoliquiritigenin to to be closed to form naringenin and liquiritigenin. These are two precursors of many flavonoid and isoflavonoid compounds. CHI enzymes in soybeans are classified into 4 categories based on homogeneity and specific substrates, namely CHI1, CHI2, CHI3, CHI4. Type I CHIs are found in most plants, while Type II CHIs are found only in legumes. CHI consists of about 220 amino acids, including 7 α-helical chains and 7 β folded plates. The active site of the enzyme CHI is mostly non-polar amino acids from the β3a folded plate, βfolded plate, α4 helical chain and α6 helical chain. Clarifying the key location of the chalcone isomerase enzyme in the phenylpropanoid pathway as well as its structure and active position plays an important role in improving flavonoid and isoflavonoid content in plants. 12 CHI genes in the soybean genome have been initially identified and placed in 4 gene subfamilies. Gen *CHI1A* in soybeans is classified in CHI type II. The *CHI1A* gene in soybeans has four exons and three introns; the 657-bp coding segment encodes 218 amino acids. The *CHI* gene encoding chalcone isomerase is the key enzyme for flavonoid biosynthesis by catalyzing open-chain naringenin chalcone and isoliquiritigenin to be closed to form naringenin and liquiritigenin - two precursors of many flavonoid and isoflavonoid compounds. The approach of enhancing the expression of genes encoding key enzymes in the phenylpropanoid

synthesis pathway is a technique used to increase isoflavone content in many different plant species.

Transgenic studies using *A.tumefaciens* in soybeans all used ripe seed cotyledon as the gene receiving material. The ability of soybean to receive genes by damaging the axillary shoots and recombinant *A.tumefaciens* infection has been studied and confirmed to be more effective than other transformation methods. Many researchers have applied this technique in the direction of improving the content of secondary substances, enhancing drought tolerance, enhancing resistance to pests and viruses, etc. Studies of transfering *CHI* gene from one species to another have resulted in transgenic plants with increased flavonoid accumulation, many times higher than that of non-transgenic plants. Research on enhanced expression of the CHI gene of that species has not been mentioned much.

The expression of *GmCHI1A* gene of soybean has been analyzed in yeast, Boerhaavia Diffusa, however, the application of *GmCHI1A* gene transfer technique to improve recombinant CHI1A content in the direction of improving isoflavone content in soybean seed germs has not been studied. Research directions of overexpression of *GmCHI1A* gene in soybean help create materials for selecting soybean cultivars with high isoflavone accumulation, creating raw materials for production of probiotics to meet the growing demand for the care and protection of human health in our country.

Chapter 2. MATERIALS AND RESEARCH METHODS

2.1. MATERIALS, CHEMICALS, RESEARCH EQUIPMENT

Soybean cultivars used in the study: Five soybean cultivars DT51, DT26, DT90, DT84 and DT2008 were used in the experiments of the thesis. Two cultivars, DT51 and DT26, were supplied by the Center for Research and Development of Beans, Vietnam Academy of Agricultural Sciences; three cultivars DT90, DT84 and DT2008 were provided by the Agricultural Genetics Institute.

Vectors and bacterial strains: The vectors used in the study included: pBT cloning vector, pRTRA7/3 vector containing 35S promoter and cmyc tag, pCB301 gene transfer vector. Strains of *E.coli* DH5*α* bacteria were used in cloning and *Agrobacterium tumefaciens* CV58 strains were used in gene transfer. The vectors and bacterial strains are provided by the Division of Plant Cell Technology - Institute of Biotechnology, Vietnam Academy of Science and Technology.

The PCR primers used in the study included *CHI-NcoI-F/CHI-NotI-R; CHI-NcoI-F/CHI-SacI-R; nptII-F/nptII-R; pUC18-F/pUC18-R*

Table 2.1. The nucleotide sequence of primer pairs used in PCR and expected DNA product size

Primer pairs	**Nucleotide sequences (5'- 3')**	**Product size (bp)**
CHI-NcoI-F/	ATGCCATGGATGGCAACGATCACCGCGGTT	677
CHI-NotI-R	TTGCGGCCGCGACTATAATGCCGTGGCTC	(cDNA)
CHI-NcoI-F/	*CATGCCATGG*ATGGCAACGATCAGCGCGGTT	722
CHI-SacI-R	*CGAGCTCGTC*ACTATAATGCCGTGGCTC	
nptII-F/	GAGGCTATTCGGCTATGACTG	963
nptII-R	ATCGGGAGCGGCGATACCGTA	
pUC18-F/	GTAAAACGACGGCCAGT	838
pUC18-R	CAGTATCGACAAAGGAC	

Chemicals: Molecular manipulators purchased from Fermentas and Bio-Neer. Enzymes purchased from Fermentas: *BamH*I, *Not*I, *Nco*I, *Hind*III, *Sac*I, T4 ligase .... Chemicals: Bacto pepton, Yeast extract, Agarose, Sucrose, Glucose, Trypton, X-gal, KCl, Tris HCl, EDTA, NaOH, MgSO4, MgCl2, Glycerol, CaCl2. Antibiotics like kanamycin, rifamycine, cefotaxime, carbenicillin ... purchased from Fermentas, Invitrogen, Sigma, Amersham and some other companies.

Equipment: PCR System 9700 (Appied Biosystem, USA), Powerpac300 electrophoresis machine (Bio-Rad, USA), DNA scanner (Mini-transllumminatior, Bio-Rad, USA), Voltex machine (Mimishaker, IKA, Germany), centrifuge, Plulser electric pulse machine, NanoDrop

nucleic acid determination machine, ABI PRISM @ 3100 Advant Genetic Analyzer (Applied Biosystem) and other modern devices.

2.2. RESEARCH METHODS

The thesis used the following research method groups: 1) methods of analysing isoflavone content; 2) methods of isolating genes; 3) methods of designing vectors for plant gene transfer; 4) methods of creating transgenic plants and analyzing transgenic plants; and 5) analysing and processing data. The diagram of experiments performed in the thesis is shown in Figure 2.1.

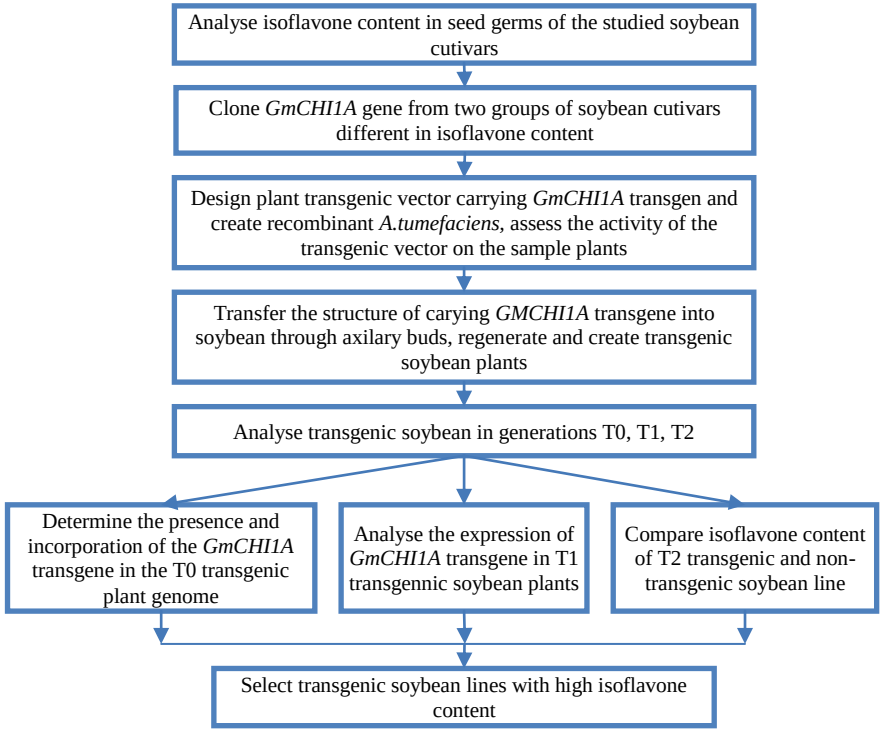


Figure 2.1. General diagram of experiments performed in the thesis

2.2.1. Methods of analysing isoflavone content

Soybean seeds germinated at 3 days of age were collected as raw material to extract daidzein and genistein. Quantification of daidzein and genistein was performed using the method of AOAC Official 2008.03 and Chen et al (2001).

2.2.2. Methods of isolating genes

Design PCR primer pairs for cloning GmCHI1A gene: From the information about the CHI gene sequence of soybean coded NM_001248290 on GenBank, the CHI-NcoI-F/CHI-NotI-R primer pair was designed to clone the coding segment of *GmCHI1A* gene.

Extract total RNA and synthesize cDNA: Total RNA was extracted from soybean germs using the Trilzol® Regents kit (Invitrogen), following the manufacturer's instructions. Use Fermantas Maxima® First Strand cDNA Synthesis to synthesize cDNA from extracted total RNA according to the manufacturer's instructions.

Clone GmCHI1A gene: *GmCHI1A* gene was amplified from cDNA using PCR technique with primers *CHI-NcoI-F/CHI-NotI-R*.

Electrophoresis test: The PCR products must experience electrophoresis on 1% agrose gel in 1X TAE buffer. The gel is dyed in ethidium bromide solution at a concentration of 0.1 µg/ml.

Clone and determine the sequence of the GmCHI1A gene: The gene cloning technique was performed according to Sambrook et al.

Analyse gene sequences: Using the BLAST software in NCBI, BioEdit, Lasergene, MEGA for analysing data on *GmCHI1A* genes. Diverse analysis was based on nucleotide sequences and inferred amino acid sequences.

2.2.3. Methods of designing GmCHI1A transgene vector

Experiments on vector design carried out were shown in Figure 2.3.

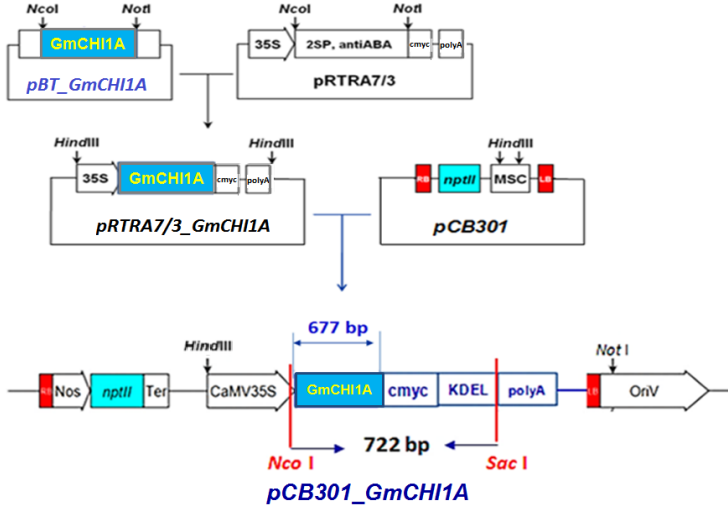


Figure 2.3. Diagram of designing *pCB301_GmCHI1A* transgene vector

2.2.4. Methods of analysing the activities of the transgene vector on tobacco plants

The transformation of the structure carrying *GmCHI1A* transgene through *A. tumefaciens* into leaf fragments and the regeneration of transgenic tobacco plants were performed as described by Topping (1998). The extraction of total DNA from tobacco leaves was performed as described by Saghai-Maroof et al. (1992). The presence of *GmCHI1A* transgene in the genome of transgenic tobacco plants in the T0 generation was analyzed using PCR. The analysis of *GmCHI1A* transgene expression at transcription level was done with RT-PCR techniques.

2.2.5. Methods of transforming and analysing transgenic soybean plants

The method of transgene in soybean by *A.tumefaciens* through axillary shoots was carried out based on the research of Olhoft et al (2006) [128] and Nguyen Thu Hien (2014) [3]. The transgenic plants regenerated from in vitro shoots, transplanted into pots and then grown in net houses are called T0; the seeds of the transgenic T0 germinating into plants are called T1; the seeds of T1 transgenic plants germinating into plants are called T2.

Verify the presence and incorporation of transgene in soybean genome: Verify the presence of the *GmCHI1A* transgene in T0 transgenic plants with PCR and specific primers *CHI-NcoI-F/CHI-SacI-R.* Verify the incorporation of GmCHI1A transgene into transgenic plant genome with Southern blot technique

Analyse the expression of recombinant rCHI1A protein: The total protein extracted from transgenic soybean leaves and non-transgenic plants was separated by electrophoresis using 10% SDS-PAGE gel (Laemmli 1970). The determination of recombinant protein expression was done using Western blot and the quantification of recombinant CHI protein was done using ELISA as described by Sun et al. (2006).

2.2.6. Process biological data

The data on isoflavone content of the research samples were processed with Microsoft Excel software, Statistical Package for the Social Science (SPSS) software at the significance level $\alpha = 0.05$. Test of statistical values was done according to Duncan at significance level $\alpha=0.05$.

2.3. RESEARCH LOCATION

The experiments were conducted from August 2015 to November 2018.

The analysis of isoflavone content in soybean germs was conducted at the Food Technology Division - Hanoi National Institute of Food Safety and Hygiene, Ministry of Healthcare. Genetic amplification experiments, molecular cloning, gene transfer, and analysis of transgenic plants were conducted at the Laboratory of Genetics and Plant Cell Technology, Department of Biology, Thai Nguyen University of Education. Transgenic vector design experiments, Southern blot analysis, Western blot, ELISA were conducted at the Department of Applied DNA Technology, Plant Cell Technology Division and Key Laboratory of Gene Technology Division of the Institute of Biotechnology - Vietnam Academy of Science and Technology.

Chapter 3. RESULTS AND DISCUSSION

3.1. CHARACTERISTICS OF *GmCHI1A* GENES ISOLATED FROM SOYBEAN PLANTS

3.1.1. Daidzein and genistein content in seed germs of some common soybean cultivars in Northern Vietnam

The investigation of isoflavone content (daidzein and genistein) of 5 soybean cultivars (DT26; DT51; DT2008; DT84; DT90) by HPLC chromatography showed that, the three-day-old seed germs of DT26 soybean cultivar had the highest content of daidzein and genistein (64.27 mg/100 g) while those of DT2008 had the lowest content (26.17 mg/100 g). The content of daidzein and genistein was different between 5 soybean cultivars at significance level $\alpha = 0.001$. Isoflavone content (daidzein + genistein) of the 5 studied soybean cultivars can be ranked in descending order as follows: DT26> DT51> DT90> DT84> DT2008

3.1.2. Clone and determine the nucleotide sequence of the GmCHI1A gene from soybean

Results of cloning and testing *GmCHI1A* gene cloning products by PCR with specific primers are shown in Figure 3.3 and Figure 3.4.

Selecting recombinant plasmid lines carrying *GmCHI1A* gene of four soybean cultivars DT26, DT51, DT2008 and DT84 and conducting nucleotide sequencing, the results showed that the DNA fragment is 657 nucleotide in size as expected when designing primers. Online analysis by the BLAST program in NCBI showed that the isolated *GmCHI1A* gene sequences had coefficients similar to the NM_001248290 sequence on GenBank used in PCR primer design: 98.93% (DT51); 98.93% (DT84); 98.78% (DT2008); 97.87% (DT26).

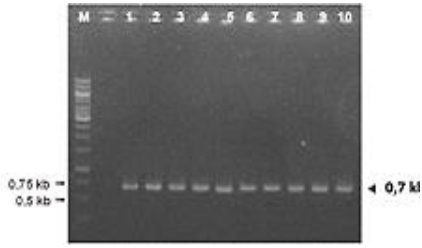


Figure 3.3. A. Image of electrophoresis testing PCR products with *GmCHI1A* gene cloning. (M: 1kb DNA Ladder; 1, 2: DT26; 3, 4: DT51; 5, 6: DT84; 7,8 DT90; 9, 10: DT2008);

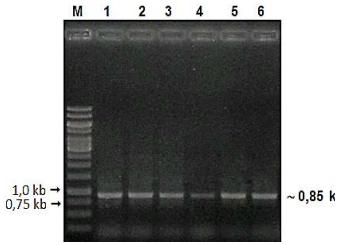


Figure 3.4. Image of electrophoresis testing colony-PCR product with primers pUC18F/pUC18R. (M: 1kb DNA Ladder; 1, 2, 3, 4, 5, 6: colonies with white phenotype were tested by colony-PCR)

Thus, the BLAST analysis results showed that the DNA fragment isolated from mRNA of the four soybean cultivars DT26, DT51, DT2008, and DT84 is the segment encoding *GmCHI1A* gene of soybean. The *GmCHI1A* (cDNA) gene of the studied soybean cultivars has 657 nucleotides and encode 218 amino acids. *GmCHI1A* gene sequences published on GenBank have the following codes: LT594994.1, LT594995.1, LT594993.1 and LT594996.1 respectively.

3.1.3. The diversity of nucleotide sequence and amino acid sequence of *GmCHI1A* gene

The four *GmCHI* gene sequences on GenBank bearing the codes AF276302, DQ191401, DQ835284 and NM_001248290 along with the 4

sequences isolated from soybean cultivars DT26, DT51, DT84 and DT2008 were selected to analyze the diversity based on nucleotide sequences and amino acid sequences. The tree diagram in Figures 3.8 and 3.9 was established based on the nucleotide sequence by UPGMA method using MEGA7 software. The analysis results in Figure 3.8 show that, based on the nucleotide sequence of the *GmCHI1A* gene, soybean cultivars are distributed in two branches: DT26 soybeans are distributed in one branch and the other 7 cultivars are distributed in the second branch, with a genetic distance of 1.2%. In Figure 3.9, the tree diagram established by the UPGMA method based on the inferred amino acid sequence of the *GmCHI1A* gene shows that the soybean cultivars are distributed in two main branches, the first main branch contains only DT26 and the second main branch includes the remaining 7 cultivars, with a genetic distance of 3.0%.

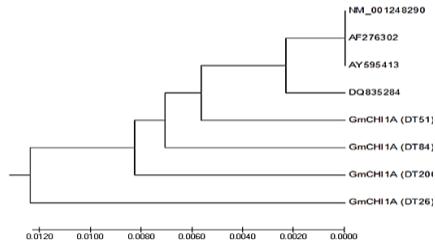


Figure 3.8. Tree diagram of the relationship between soybean cultivars based on the nucleotide sequence of *GmCHI1A* gene established by UPGMA method

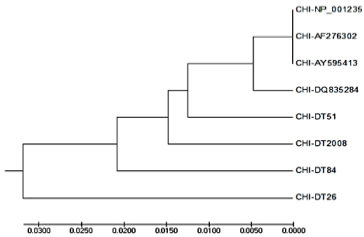


Figure 3.9. Tree diagram of the relationship between soybean cultivars based on the deducted amino acid sequence of *GmCHI1A* genes established by UPGMA method

3.2. DESIGN PLANT TRANSGENE VECTOR CARRYING *GmCHI1A* GENE

In order to transfer the *GmCHI1A* gene into plants and be able to check the expression of the protein product of the gene, the pCB301 gene expression vector carrying the *CaMV35S* promoter was designed to control *GmCHI1A* gene expression in plants.

3.2.1. Create a structure carrying *GmCHI1A* transgene

The pRTRA7/3 vector contains the CaMV35S transcription promoter, the nucleotide sequence identifying the p-myc peptide sequence and the

nucleotide sequence identifying the KDEL segment. Open the pRTRA7/3 vector ring with the pair of enzymes *NotI/NcoI* to create 2 DNA segments with sizes of 0.9 kb and 3.3 kb, of which the DNA fragment of 3.3 kb had a *35S_cmyc_KDEL sequence*. The *GmCHI1A* gene from the *pBT_GmCHI1A* cloning vector was cleaved with a pair of enzymes *NotI* and *NcoI* to produce two DNA segments with sizes of about 0.67 kb and 2.7kb. In particular, the 0.66 kb DNA segment is the target *GmCHI1A* gene that needs to be collected (Figure 3.10). Purify the *GmCHI1A* gene segment and bind it to pRTRA7/3 vector through ligation reaction under the catalysis of the T4 ligase enzyme to create recombinant *pRTRA7/3_GmCHI1A* vector structure carrying *CaMV35S_GmCHI1A-cmyc-polyA* structure. Clone in E.coli DH5α and check by colony-PCR (Figure 3.11).

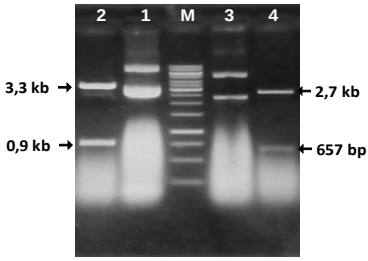


Figure 3.10. Electrophoresis image of *pRTRA7/3* cut products and pBT-GmCHI1A cut products using *NcoI/NotI* enzyme pairs. (M: 1kb DNA Ladder; 1: *pRTRA7/3* Vector not cut with enzyme *NotI* and *NcoI*; 2: *pRTRA7/3* vector product cut with enzymes *NcoI* and *NotI*; 3: recombinant *pBT-GmCHI1A* vector not cut with *NotI* and *NcoI* enzymes; 4: Recombinant *pBT-GmCHI1A* vector products cut with *NcoI* and *NotI* enzymes)

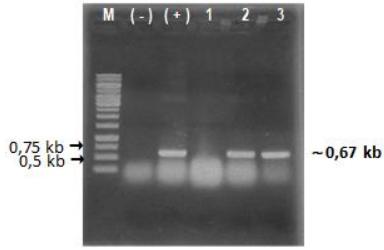


Figure 3.11. Electrophoretic image of colony-PCR product cloning *GmCHI1A* gene from colonies. (M: 1kb DNA Ladder; (-): Colony-PCR from *E.coli* colonies with non-transformed *pRTRA7/3_GmCHI1A*; (+): PCR of the cloned *GmCHI1A* gene from *pBT_GmCHI1A* vector; 1-3: Colony-PCR from colonies with transformed *pRTRA7/3_GmCHI1A*

3.2.2. Generate *pCB301_GmCHI1A* transgenic vector

Performing the *pRTRA7/3_GmCHI1A* vector cut reaction using *HindIII*, the structure *35S_GmCHI1A_cmyc_KDEL_polyA* (1.5 kb) and

DNA segment with a size of 2.4 kb were shown in Figure 3.12. Opening the pCB301 gene transfer vector ring with *Hind*III, there are two 5,502-kb DNA fragments (Figure 3.13). Attach the *35S_GmCHI1A_cmyc_KDEL* structure to the pCB301 vector to generate *pCB301_GmCHI1A* transgenic vector (Figure 3.14).

Transform *pCB301_GmCHI1A* and clone in *E.coli* DH5α and select colonies using colony-PCR. The *pCB301_GmCHI1A* plasmid was extracted from PCR-positive lines.

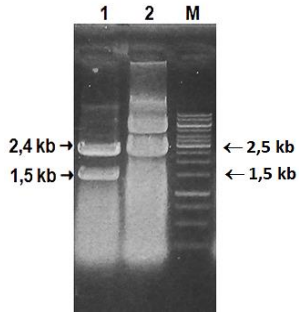


Figure 3.12. Electrophoresis image of cutting *pRTRA7/3_GmCHI1A* plasmid product by *HindIII*. (M: 1kb DNA Ladder; Electrophoresis lane 1: *pRTRA7/3_GmCHI1A* plasmid cut by *HindIII*; Electrophoresis lane 2: uncut *pRTRA7/3_GmCHI1A* plasmid)

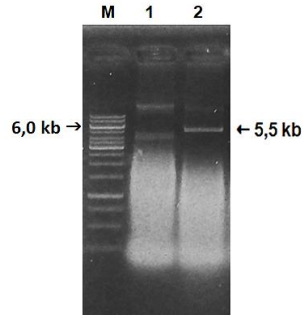


Figure 3.13. Electrophoresis image of testing pCB301plasmid cutting product. (M: 1kb DNA Ladder; Electrophoresis lane 1: plasmid not cut by *HindIII*; Electrophoresis lane 2: Open target DNA product from pCB301 vector)

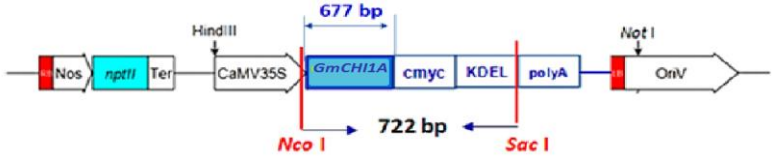


Figure 3.14. Diagram of *pCB301_GmCHI1A* transgenic vector structure. *(nptII*: kanamycin resistance gene; *CaMV35S*: promoter 35S; *GmCHI1A*: *Glycine max chalcone isomerase 1A (GmCHI1A)* gene isolated from soybean; *cmyc*: nucleotide sequence encoding c-myc peptide; *KDEL*: nucleotide sequence encoding the KDEL peptide

3.2.3. Create *A. tumefaciens* CV58 containing *pCB301_GmCHI1A* transgenic vector

pCB301_GmCHI1A plassmid extracted from purified colony-PCR-positive *E.coli* strains was transformed into *A.tumefaciens* CV58. Raise for 48 hours at 28°C and when colonies appear on agar, check the specific primers *CHI-NcoI-F/CHI-NotI-R* colony-PCR to select colonies carrying *GmCHI1A* transgene vector (Figure 3.16).

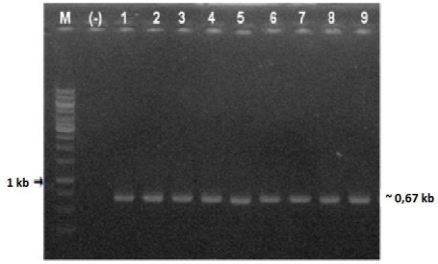


Figure 3.16. Electrophoresis image of testing colony-PCR products by specific primers CHI-NcoI-F/CHI-NotI-R from A.tumefaciens CV58 colonies. (M: 1kb DNA Ladder; (-): negative control - A.tumefaciens with non-transformed pCB301_GmCHI1A; 1-9: nine colonies of A.tumefaciens CV58 containing pCB301_GmCHI1A vector)

3.2.4. Analyse the activity of *pCB301_GmCHI1A* transgenic vector on tobacco plants

The transformation of *GmCHI1A* transgene structure was carried out by *A.tumefaciens* infection into tobacco leaf tissue (Figure 3.17). The results of 3 times of transformation were presented in Table 3.5. Table 3.5 shows that after three times of transformations in the experimental batch, 83 samples were generated for shoot cluster and through antibiotic selection, there were 206 shoots surviving. In rooting environment, there were 163 root shoots and 98 plants were selected to be transferred in potting soil. The final result is that 30 plants survived in net house conditions.

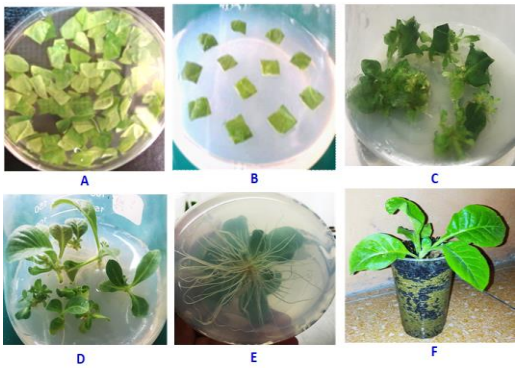
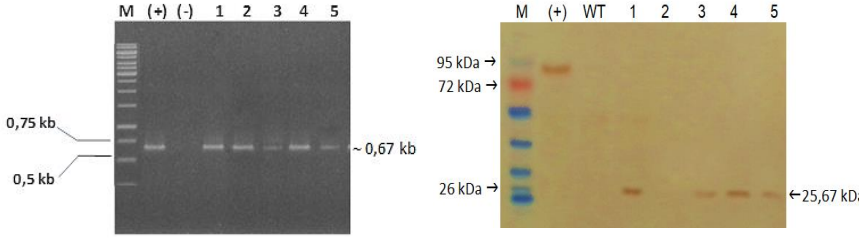


Figure 3.17. Transformation and regeneration of *GmCHI1A* transgenic tobacco plants. (A: cotyledons submerged into the bacterial suspension; B: co-cultivate in CCM medium; C: regeneration of multiple shoots in a selective medium containing kanamicin; D: shoot elongation; E: Initiation of roots in the RM medium; F: Transgenic tobacco plants grown on stands.

Collecting the cotyledons of 30 transgenic tobacco plants and then extracting total DNA and analyzing the presence of transgenic *GmCHI1A* gene by PCR with primers *CHI-NcoI-F/CHI-NotI-R*, the results showed that the DNA band has a size of about 0.67 kb in 22 electrophoresis lanes, which are plants 4, 5, 6, 7, 8, 9, 11, 12, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29 while plants 1, 2, 3, 10, 14, 15, 26, 30 have no DNA band.

Selecting randomly seven T0 tobacco plants positive for PCR, with normal growth and development for Southern blot analysis, the results show that 5/7 transgenic T0 tobacco plants T01, T02, T04, T05, T06 appears DNA band. Thus, the *GmCHI1A* transgene has been incorporated into the transgenic tobacco genome.

Total RNA extracted from the cotyledons of the 5 transgenic tobacco plants (T01, T02, T04, T05, T06) were positive for Southern hybrid, with normal growth and development in T0 generation were used to generate cDNA and implement PCR reaction with primers *CHI-NcoI-F/CHI-NotI-R.* The results of *GmCHI1A* (cDNA) transgene cloning from mRNA of 5 transgenic tobacco plants showed that all the 5 electrophoresis lanes have DNA band with a size of about 0.67 kb (Figure 3.19A). This results demonstrated that the *GmCHI1A* transgene exhibits mRNA synthesis transcription.



A B

Figure 3.19. A- Electrophoresis image of testing RT-PCR product cloning *GmCHI1A* gene (cDNA) from mRNA of 5 transgenic tobacco plants in generation T0. M: 1kb DNA Ladder; (+):*pBT_GmCHI1A* plasmid (positive control); (-) Non-transgenic plants (WT-negative control); 1-5: T0 transgenic tobacco plants)

B- Results of Western blot analysis on transgenic tobacco plants of generation T0. (+): C-myc-tagged HA protein; WT: Protein obtained from non-transgenic plants; 1, 2, 3, 4, 5 protein samples collected from transgenic tobacco plants positive for southern blot hybrids

However, in Figure 3.19B, Western blot analysis results obtained 4/5 T0 plants with protein band in size of about 25.67 kDa. Thus, the *GmCHI1A* transgene decoded the recombinant rCHI1A protein synthesis in 4 T0 tobacco plants (T01, T04, T05, T06).

The results of tobacco analysis imply that the *pCB301_GmCHI1A* transgenic vector works well in transgenic tobacco plants and can be used to transfer into soybean and other crops.

3.3. ANALYSE THE EXPRESSION OF *GmCHI1A* GENE IN TRANSGENIC SOYBEAN

3.3.1. Transform *pCB301_GmCHI1A* structure into soybean through *A.tumefaciens*

Carry out an experiment to transfer *pCB301_GmCHI1A* structure into the DT2008 soybean line through 3 times of transformation with 390 cotyledons (Figure 3.20). Out of 390 transformed samples, 26 transgenic plants were grown on the substrate.

3.3.2. Analyse presence and incorporation of *GmCHI1A* transgene in T0 transgenic soybean plants

Using PCR technique to check the presence of *GmCHI1A* transgene in 26 transgenic soybean plants in T0 generation. Total DNA extracted from the cotyledons of T0 transgenic soybean and non-transgenic control plants was used for PCR with *CHI-NcoI-F/CHI-SacI-R* primers. The results of *GmCHI1A* transgenic PCR product electrophoresis showed that on the electrophoresis gel plate, there were 8 lanes running the band of DNA. They are lanes 1, 3, 4, 5, 21, 22, 24 and 25 with a size of approximately 0.72 kb corresponding to the size of the *GmCHI1A* transgene. Transgenic soybean

plants positive for PCR in the T0 generation of DT2008 cultivars were labelled as T0-1; T0-3; T0-4; T0-5; T0-21; T0-22; T0-24; T0-25. In the electrophoresis analysis of PCR products from DNA of non-transgenic control plants, there was no visible DNA band.

PCR-positive soybean plants were tested for incorporation of the *GmCHI1A* transgene into the transgenic genome by Southern blot. Total DNA extracted from the leaves of transgenic soybean and non-transgenic control plants was purified and treated with *Sac*I restriction enzyme to collect *nptII_CaMV35S_GmCHI1A_cmyc* fragments containing *nptII* and *GmCHI1A* genes. The results of Southern blot analysis shown in Figure 3.23 show that 7 T0 plants T0-1, T0-3, T0-4, T0-21, T0-22, T0-24, T0-25 produce DNA band while T0-5 and WT plants did not produce Southern hybrid results. Transformation efficiency up to the time of Southern blot analysis was 7/390 = 1.79%. After the membrane appeared, on the hybrid membrane of each DNA band there was a corresponding copy. The WT samples showed negative results, indicating a specific hybrid reaction where the probe was not associated with endogenous genes.

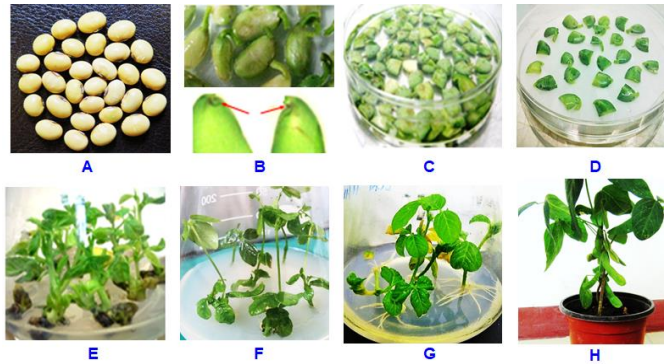


Figure 3.20. Results of generating *GmCHI1A* transgenic soybean plants from DT2008 by recombinant *A. tumefaciens* infection through ripe seed axillary cotyledon. (A: DT2008 soybean seeds after disinfection with chlorine gas; B: Damaged cotyledons obtained from germinated seeds in GM medium to produce transformation materials; C: Damaged axillary cotyledons submerged into the bacterial suspension for 30 minutes; D: Cultivating cotyledons in co-cultivation medium (CCM) in dark conditions for 5 days; E: Multi-shoot induction in SIM, supplement with BAP 2 mg L^{-1} + kanamycin 50 mg L^{-1}; F: Cut off the cotyledons, transfer to SEM shoot elongation medium for 2 weeks, adding GA3 0.5 mg L^{-1} + IAA 0.1 mg L^{-1} + kanamycin 50 mg L^{-1}); G: Rooting in RM medium, supplementing with 0.1 mg L^{-1} IBA for 20 days; H: Transgenic plants grown in pots containing husk ash and golden sand with a ratio of 1:1)

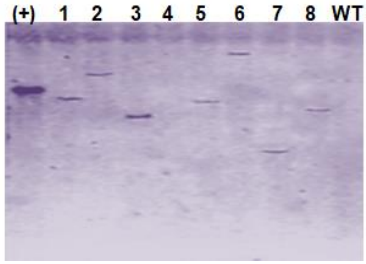


Figure 3.23. Southern blot analysis results of *GmCHI1A* transgenic soybean plants with *nptII* transducer marked with biotin. (+):*pCB301-GmCHI1A* vector; 1-8: Transgenic soybean lines positive for PCR (1: T0-1; 2: T0-3; 3: T0-4; 4: T0-5; 5: T0-21; 6: T0-22 ; 7: T0-24; 8: T0-25); WT: Non-transgenic soybean plants.

Results of Southern blot analysis showed that the *GmCHI1A* transgene was incorporated into the soybean genome. Transgenic lines T0-1, T0-3, T0-4, T0-21, T0-22, T0-24, T0-25 producing Southern hybrid results continue to be evaluated for growth, development and their seeds were collected to serve the analysis of transgenic plants in T1 gene generation.

3.3.3. Analyse the expression of recombinant CHI1A protein by Western blot and ELISA

The seeds of 7 T0 transgenic plants (T0-1, T0-3, T0-4, T0-21, T0-22, T0-24, T0-25) were sown in each experimental plot for 7 T1transgenic lines, labelled as T1-1, T1-3, T1-4, T1-21, T1-22, T1-24, T1-25. Simultaneously, T1 transgenic soybean lines were used to analyze recombinant CHI1A protein expression (symbolized as rCHI1A) by Western blot and ELISA.

Results of analysing by Western blot the protein in 7 transgenic soybean lines and non-transgenic control lines showed that of the 7 lines of *GmCHI1A* transgenic soybean in T1 generation, there were 4 lines producing Western blot results (Figure 3.24). Thus, at the time of analysing recombinant rCHI1A protein expression, the gene transfer efficiency at the time of Southern blot analysis was 4/390 = 1.03%.

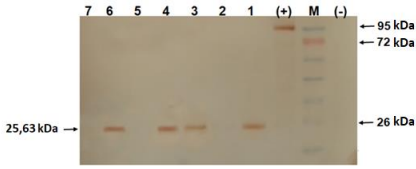


Figure 3.24. Results of analysing by Western blot the protein of T1 transgenic soybean plants and non-transgenic soybean plants. M: standard protein ladder; (+) The positive control is cmyc-tagged HA protein; (-) The negative control is a protein sample obtained from non-transgenic plants; 1-7 (T1-1, T1-3, T1-4, T1-21, T1-22, T1-24, T1-25): Protein collected from transgenic soybean plants positive for Southern blot hybridization

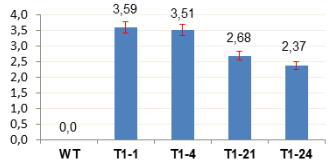


Figure 3.25. Results of ELISA analysis to determine recombinant protein content of transgenic rCHI1A soybean lines T1-1, T1-4, T1-21, T1-24 and non-transgenic control plants (WT)

Figure 3.25 shows that recombinant rCHI1A protein content of 4 transgenic soybean lines T1-1, T1-4, T1-21 and T1-24 ranged from 2.37-3.59 µg/mg. The T1-1 line had the highest recombinant rCHI1A protein content (3.59 µg/g), followed by the T1-4 line (3.51 µg/g) and T1-21 line (2.68 µg/g) and the lowest was T1-24 (2.37 µg/g) (Figure 3.25).

Thus, it can be remarked that the *GmCHI1A* transgene was genetically transmitted through sexual reproduction from the T0 to T1 generation and was active for transcribing and decoding protein synthesis in transgenic soybean plants in T1 generation.

3.3.4. Analyse the daidzein and genistein content of transgenic soybean lines

The seed germs of 4 transgenic lines in T2 generation (T2-1, T2-4, T2-21, T2-24) were used to analyze daidzein and genistein content (Table 3.8).

Table 3.8. Changes in daidzein and genistein content at seed germination stage of transgenic soybean lines compared to non-transgenic plants

WT plants and transgenic lines	Daidzein and genistein content				
	Daidzein (µg/g dry weight)	Increase compared to WT (%)	Genistein (µg/g dry weight)	Increase compared to WT (%)	Total daidzein and genistein (µg/g dry weight)
WT	253,05 ± 3,60	100,00	$113,11^{A}$ ±1,78	100,00	366,16
T2-1	$473,79^{c}$ ± 9,63	187,23	$524,64^{D}$ ±4,27	463,93	998,43
T2-4	457,07 ±18,14	180,62	$467,66^{C}$ ±17,97	413,46	870,53
T2-21	447,92 ±14,87	177,01	$499,72^{CD}$ ±15,95	441,80	947,64
T2-24	421,22 ± 8,91	166,46	373,00 ± 9,82	329,77	750,99

The analysis results showed that, compared with non-transgenic soybean lines (WT), the contents of daidzein and genistein in the seeds of transgenic soybean lines T2-1, T2-4, T2-21, T2-24 are all high, and the daidzein content of transgenic lines increased from 139.17% to 186.86%; the content of genistein increased from 329.80% to 463.93%. The difference in isoflavone content between transgenic lines and WR plants was analyzed using Duncan test with $p < 0.05$.

3.4. DISCUSSION ON THE RESEARCH RESULTS

In higher plants, CHI is an important enzyme that catalyzes the reactions in the phenylpropanoid pathway to produce isoflavonoid, flavonoid, anthocyanin products (Oliver et al., 2005). The *GmCHI1A* genes of soybean plants encoding the CHI1A enzyme protein were isolated from

Vietnamese soybean cultivars which have been registered on GenBank, with the codes LT594994.1, LT594995.1, LT594993.1, and LT594996.1. The *GmCHI1A* gene encoding segment has 657 nucleotides, encoding 218 amino acids isolated from four soybean cultivars DT26, DT51, DT84, DT2008 agrees with the publications of Ralston et al. (2005), Chiu et al. (2001).), Chung and Nam (2007), Dastmalchi and Dhaubhadel (2019).

The research direction of *CHI* gene expression in other species was recorded in the results of transferring SmCHI from *S. medusa* peony into tobacco plants and expressing *CHI* gene of petunia in tomato plants. The results show that transgenic plants have flavonoid increasing many times higher than that of non-transgenic plants.

Ralston et al. (2005) analyzed the expression of both *GmCHI* type I and type II in yeast (*Saccharomyces cerevisiae*), and claimed that CHI type II enzymes coexisting with an isoflavone synthase and other chalcones added to the culture medium has been converted into isoflavanone and isoflavone. The study of Vu Thi Nhu Trang et al analyzed expression of type II *GmCHI* gene of soybean in Talinum paniculatum and the result was that the transgenic lines had total flavonoid content increasing 4,8-7.4 times higher than non-transgenic plants (Vu et al., 2018). Many valuable medicinal plants do not contain CHI type II enzyme, and there is no branch of isoflavone synthesis in the phenylpropanoid pathway; therefore, our research results are the basis for the supplement of CHI type II enzyme from legumes into medicinal plants by gene transfer technique to obtain isoflavones.

In the approach of increasing isoflavone content in soybean seed germs, in this study we have successfully expressed the *GmCHI1A* gene in soybean plants under the control of the CAMV35S promoter. Transgenic soybean lines in the T1 generation (T1-1, T1-4, T1-21 and T1-24) have recombinant rCHI1A protein content ranging from 2.37 to 3.59 µg/mg. Four T2 transgenic lines have been created (T2-1, T2-4, T2-21, T2-24) and T2 soybean germs were analyzed for daidzein and genistein content. The results of HPLC analysis showed that the overexpression of *GmCHI1A* gene increased the GmCHI1A enzyme content and the germs of T2 transgenic soybean lines had daidzein content increasing from 166.46% to 187.23%, and genistein content increasing from 329.80% to 463.93% compared those of non-transgenic plants. Remarkably, the transgenic lines had high genistein content which increased from 3.3 to 4.64 times compared with the non-transgenic control. Thus, it can be said that the overexpression of *GmCHI1A* gene in soybean increased the activity of CHI

enzyme and increased the accumulation of liquiritigenin and naringenin. Involvement of the IFS enzyme also increased the content of isoflavones and other flavonoids in transgenic plants.

In our study, a remarkable sign was that the genistein content in germs of the transgenic lines was higher than that of the non-transgenic germs and higher than the daidzein content. Isoflavones are important natural products that can be used for human health care and protection (Jiang et al., 2015). Genistein in soybeans is a rare pharmaceutical substance that has antioxidant properties, produces significant amounts of collagen, prevents breast cancer in women, prostate cancer in men and some other cancers such as colon cancer, lung cancer, skin cancer and blood cancer (Perabo et al., 2008). Therefore, creating a genistein-rich soybean line has practical implications for community health care.

For the first time, the *GmCHI1A* gene was successfully transformed and expressed in soybean plants. The overexpression of the *GmCHI1A* gene resulted in an increase in the CHI1A enzyme content. At the same time, the daidzein and genistein content in the germs of T2 transgenic soybean lines was higher than that of the non-transgenic control.

CONCLUSION AND RECOMMENDATIONS

1. Conclusion

1.1. Daidzein and genistein content in the seed germs of the five soybean cultivars DT26, DT51, DT90, DT84 and DT2008 were investigated. In particular, DT26 soybean had the highest content of daidzein and genistein (64.27 mg/100g in each germ), and the lowest content was found in DT2008. The encoding region of the *GmCHI1A* gene isolated from the mRNA of soybean plants was 657 nucleotides in size and encodes 218 amino acids.

1.2. The *pCB301_GmCHI1A* gene transfer vector containing the *GmCHI1A* gene and the CaMV35S promoter was successfully designed and transformed into tobacco plants. The *GmCHI1A* transgene was verified to be incorporated into tobacco genome through Southern blot analysis, and was expressed to produce recombinant rGmCHI1A protein in transgenic tobacco plants.

1.3. The *GmCHI1A* transgene was successfully transformed into DT2008 soybean using recombinant *A.tumefaciens* infection through the axillary cotyledons, which produced 26 T0 transgenic soybean plants. Among them, the transgene was incorporated into the DT2008 soybean genome

and four T1 transgenic soybean plants generated from T0 expressed recombinant protein products with a transgenic efficiency of 1.03%.

The expression of *GmCHI1A* transgene increased the CHI1A enzyme content in the germs and increased daidzein content from 166.46% to 187.23%, genistein content from 329.80% -463.93% compared with non-transgenic plants. Four transgenic soybean lines T2-1, T2-4, T2-21, T2-24 are materials for further selection and evaluation in the next generations.

2. Recommendations

2.1. The nucleotide sequence and amino acid sequence of the *GmCHI1A* gene among the soybean cultivars have differences. Especially, DT26 soybean has a relatively high content of isoflavones. Because it is not within the scope of the thesis research, it is not known whether these differences are related to the isoflavone content. Therefore, further research is needed to shed light on the possible association between the differences in the sequence of this gene in DT26 and the high isoflavone content that it has produced.

2.2. Transgenic soybean lines T2-1, T2-4, T2-21, T2-24 can be used as materials for selecting soybean cultivars, so it should be further analyzed in the next generation to select and create a soybean line with high and stable isoflavone content.

2.3. The results of the thesis have demonstrated that the overexpression of the gene encoding the key enzyme in soybean isoflavone biosynthesis has increased the isoflavone content in transgenic soybean seed germs. This is a scientific basis to be applied to legumes and other plants in the direction of improving isoflavone content to research functional foods for community health care.