Journal of Agricultural Technology 2015 Vol. 11(8): 2547-2561
Available online http://www.ijat-aatsea.com
ISSN 1686-9141

In Silico Identification, Classification And Expression Analysis Of Genes Encoding Putative Light-Harvesting Chlorophyll A/B-Binding Proteins In Coffee (*Coffea Canephora* L.)

Cao Phi Bang[1*], Tran Thi Thanh Huyen[2]

[1]Biotechnology Research Center, Faculty of Natural Sciences, Hung Vuong University
[2]Department of Plant Physiology and Application, Faculty of Biology, Hanoi National University of Education

Cao P.B. and Tran T.T.H. (2015). In Silico identification, classification and expression analysis of genes encoding putative light-harvesting chlorophyll a/b-binding proteins in coffee (Coffea canephora L.). Journal of Agricultural Technology 11(8): 2547-2561.

The light-harvesting proteins bind to chlorophyll a/b and transfer light energy to photosynthetic reaction centre. These proteins also play a major role in photoprotection and abiotic stress tolerance in many plants. By using in silico methods, a total of 28 genes encoding putative light harvesting complex (LHC) have been identified in coffee (Coffea cenaphora) genome. Most of putative LHC deduced proteins possess the Chloroa_b-bind (PF00504) conserved domain. Based on phylogeny analysis, these coffee LHC genes have been classified into many groups, including photosystem (PSI, PSII), LHC-related and light-inducible genes. Both PSI and PSII groups were divided into six subgroups, respectively (A1-A6) and (B1-B6) All the subgroups contain one member each, except B1and B3 subgroup which contain multiple genomic loci (five members). The B2 and B4 are single locus subgroups in C. cenaphora but they are multiple genomic loci in both A. thaliana and rice. In contrast, B3 subgroup contains four genes in C. cenaphora but only one member in A. thaliana. In general, the transcripts of most of coffee putative LHC genes are abundant in leaves and perisperm but weakly or not detectable in roots. In addition, most of the genes are expressed in pistil. The coffee early light-inducible protein encoding genes are strongly expressed in all the investigated tissues.

Key words: coffee , chlorophyll a/b binding protein, light-harvesting complex proteins, gene identification, gene classification, gene expression, in silico analysis.

Introduction

Plants absorb light energy for photosynthesis by using two types of light-harvesting complexes (LHC-I and LHCII). The light harvesting proteins bind to chlorophyll a/b and transfer light energy to photosynthetic reaction centre (Jansson, 1994). Besides plants, the light-harvesting proteins, present in different taxa including cyanobacteria, purple bacteria and green sulphur bacteria, exhibit low sequence similarity although some structural similarity

[*]**Corresponding Author:** Cao Phi Bang, **Email:** phibang.cao@hvu.edu.vn

can be observed (Green, 2001). In higher plants, the LHC proteins constitute a large family of proteins which consists of chlorophyll a/b-binding proteins (CABs), high light-induced proteins (HLIPs), early light-induced proteins (ELIPs), the psbS subunit of photosystem II (psbS), and stress-enhanced proteins (SEPs) (Jansson, 1994).

The structure of LHC proteins from many different species such as algae and higher plants contain three transmembrane helices together with characteristic LHC motif (ExxxxRxAM) (Green and Kuhlbrandt, 1995). LHC proteins play a major role in light absorption and photoprotection (reviewed in (Neilson and Durnford, 2010). The LHC proteins of PSII (LHCB proteins), involved in the stomatal response to abscisic acid, are important for drought tolerance of *A.thaliana* (Liu *et al.,* 2013; Xu *et al.,* 2012). Among the LHC-related proteins, the early light-induced proteins were the most studied. These proteins play a key role in photoprotection and abiotic stress response in a large number of species such as *A. thaliana* (Hutin *et al.,* 2003), *Rhododendron catawbiense* (Peng *et al.,* 2008), grape vine and pea (Berti and Pinto, 2012) and tea (Li *et al.,* 2013). Recently, thanks to public availability of the genome sequences, the LHC gene family has been genome-wide identified an analyzed in some plant species such as brown alga (Dittami*et al.,* 2010), *A. thaliana* and rice (Umate, 2010).

Coffee is one of the most important trade beverages with over 2.25 billion cups consumed a day (Denoeud *et al.,* 2014). The coffee seeds contain many classes of biochemical compounds such as flavonoids, phenolics, alkaloids, terpenoids and rich in stimulant caffeine (Leroy *et al.,* 2006).This plant is an economically important crop in more than 60 countries in South and Central America, Asia, and Africa with over 11 million ha of plantation (Denoeud *et al.,* 2014). Currently, *Coffea arabica* and *C. canephora* are two major varieties of the coffee cultivated worldwide. *C. canephora*, which represents approximately 38% of coffee production (Vieira *et al.,* 2006) is the diploid species (2n = 2x = 22). This species is one of the parents of the allotetraploid *C. arabica* (2n = 4x = 44 chromosomes) which was derived from hybridization between *C. canephora* and *C. eugenioides* (Lashermes *et al.,* 1999). The genomic sequence of *C. canephora* which has been completely sequenced in 2014, is a powerful resource for helping the research on main traits such as quality, yield, protection against pests, and abiotic stress tolerance to face climatic changes (Denoeud *et al.,* 2014).The analysis of LHC encoding genes in coffee is particularly relevant since coffea is mainly grown in tropical and subtropical regions under relatively high intensities of light.

Objectives: This work aimed to genome-wide identify the putative LHC genes in *C. cenaphora* genome by using *in silico* methods. In addition, the

classification and expression of these *LHC* genes were analyzed.

Material and methods

Identification of LHC from coffee genomic sequences

Based on the coffee genome (http://coffee-genome.org/) (Dereeper *et al.*, 2015), an extensive research was performed for identifying all members of the LHC family. Firstly, the LHC sequences of *A. thaliana* (Umate, 2010) were used as queries to perform blastp (Altschul *et al.*, 1997) against coffee genome database with an e-value of le-10. Secondly, the selected coffee genes were used as queries for BLAST search on the coffee genome for identifying the coffee paralogs that had been excluded by their dissimilarity to the *Arabidopsis* orthologs. Finally all candidate sequences were submitted to domain research by using the Pfam software (http://www.sanger.ac.uk/software/pfam) (Finn *et al.*, 2014).

Sequence analysis and construction of the phylogenetic tree

The molecular weight, theoretical pI and GRAVY (grand average of hydropathy) of putatives sequences were calculated by the PROTPARAM tool (http://web.expasy.org/protparam/) (Gasteiger *et al.*, 2003). Subcellular localization analysis of the deduced amino acids was performed using TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson *et al.*, 2007). The transmembrane protein topology was predicted by using the PSIPRED Server (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan *et al.*, 2013).

Phylogenetic analyses were conducted using MEGA version 5 (Tamura *et al.*, 2011). Complete coffee LHC predicted proteins were aligned using the MAFFT server (http://mafft.cbrc.jp/alignment/server/) (Katoh and Standley, 2013) then phylogenetic trees was constructed by using the Maximum Likelihood method with 1000 bootstraps.

Gene expression analyses

The relative *LHC* gene expression log10(RPKM) (Reads Per Kilobase of transcript per Million fragments mapped) values were obtained from (Denoeud *et al.*, 2014).

Results and Discussions

Identification of the HLC genes in coffee genome

The LHC protein sequences from *C. cenaphora* have been identified by using the BLASTP programme with queries from *A. thaliana* (Umate, 2010). Then, the second blastp was performed against coffee genome using selected coffee genes as queries. A total of twenty-eight full-length genes encoding putative LHC proteins have been identified. When these predicted proteins were analyzed by using Pfam, only 24 of the 28 candidate sequences exhibit Chloroa_b-bind (PF00504) conserved domain. The four remaining sequences (CcSEP1, CcSEP2, CcHOP and CcHOP2) were grouped with *A.thaliana* and rice orthologs on phylogenetic tree (Figure 1). The LHC gene family is smaller in coffee compared to *A. thaliana* and rice families which have 30 and 29 genes respectively.

The coffee putative *LHC* genes have a size ranging from 510 to 4057 nucleotides in genomic full-length. Except of six, all of the genes exibit intron (from 1 to four introns). Their deduced full-length protein sequences range from 122 to 330 amino acids. Among them, CcOHP has a smallest size with molecular mass of 13.39kDa while CcChla/b-like has a biggest size with molecular mass of 36.35kDa. Theoretical pI values of Coffee LHCs fluctuate in a wide range from 4.53 to 9.91, with 21 acidic and seven basic LHCs. The *LHC* encoding gene family has been genome-wide indentified in *A.thaliana* and rice (Umate, 2010). The characteristics of coffee LHC are in agreement with orthologs of *A. thaliana* and rice. AtOHP (locus At5G02120) and OsOHP1 (LOC_Os05g22730) are also the smallest LHC protein with 110 amino acids (MW of 12,01 kDa) and 113 amino acids (MW of 12,11 kDa), respectively. Theoretical pI values of coffee LHCs is consistent with pI range of *A. thaliana* (4.61~11.51) and rice (4.14~12.82) (Umate, 2010). Transmembrane helix predictions using PSIPRED server (Buchan *et al.*, 2013) show that major (17/28) of coffee LHCs have three helices. Two sequences exhibit one helix (CcOHP and CcOHP2) when the remaining ones (including CcSEP1 and CcSEP2) contain two helices. Subcellular localization analysis of the deduced amino acids using TargetP 1.1 Server (Emanuelsson *et al.*, 2007) suggests that majority of LHC proteins are present in chloroplast but one member (CcLHCB1.1) is probably targeted to mitochondria.The predicted localization of the four additional ones is ambiguous.

Table 1. Inventory and characteristics of the *LHC* genes identified in *C. canephora*

Gen	Subgroup	Locus name	Genomic full lenght (bp)	Protein full lenght (aa)	MW (kDa)	pI	TM	Introns number	Subcellular location
CcLHCB3.1	LHCB3	Cc00_g12790	510	169	18.51	4.58	2	0	Other
CcLHCB3.2	LHCB3	Cc00_g34870	510	169	18.51	4.58	2	0	Other
CcSEP1	SEP1	Cc02_g19180	2242	143	14.56	9.78	2	3	C
CcLHCB6	LHCB6	Cc02_g21720	999	262	27.79	8.78	3	1	C
CcLHCB1.1	LHCB1	Cc02_g33560	1841	229	25.85	9.91	3	2	M
CcELIP	ELIP	Cc03_g04300	806	191	20.23	8.93	2	2	C
CcLHCA4	LHCA4	Cc04_g16410	1002	252	27.71	6.11	3	2	C
CcLHCB1.2	LHCB1	Cc05_g09650	802	242	25.70	5.32	2	1	C
CcLHCA3	LHCA3	Cc05_g09930	1257	273	29.38	6.43	3	2	C
CcLHCB3.3	LHCB3	Cc05_g12720	1012	263	28.40	5.03	3	2	C
CcLHCB4	LHCB4	Cc06_g01460	1087	286	31.11	5.59	3	1	C
CcOHP	OHP	Cc06_g02340	596	122	13.39	9.80	1	2	C
CcLHCA6	LHCA6	Cc06_g12480	1761	260	28.62	5.52	3	4	C
CcLHCB3.4	LHCB3	Cc07_g00260	510	169	18.57	4.53	2	0	Other
CcSEP2	SEP2	Cc07_g16120	1471	209	22.68	5.25	2	1	C
CcLIL	LIL	Cc08_g11360	2526	261	28.97	5.02	2	2	C
CcLHCB1.3	LHCB1	Cc09_g09030	795	264	28.17	5.44	3	0	C
CcLHCB2	LHCB2	Cc09_g09500	1851	265	28.57	5.96	3	3	C
CcLHCB1.4	LHCB1	Cc09_g09010	795	264	28.28	5.67	3	0	C
CcLHCB1.5	LHCB1	Cc09_g09020	795	264	28.22	5.27	3	0	C
CcCHLA1	LHCA1	Cc09_g02010	1334	244	26.24	5.84	3	1	C

Table 1 continued

Gen	Subgroup	Locus name	Genomic full lenght (bp)	Protein full lenght (aa)	MW (kDa)	pI	TM	Introns number	Subcellular location
CcChla/b-like	CHLa/b-like	Cc10_g00140	4057	330	36.35	6.32	3	5	C
CcLHCA5	LHCA5	Cc10_g04190	1639	268	28.93	6.6	3	5	C
CcPsbS	PsbS	Cc10_g11890	2312	271	28.40	6.76	3	3	C
CcOHP2	OHP2	Cc10_g15180	3125	188	20.06	9.33	1	1	C
CcLHCB5	LHCB5	Cc10_g16210	1632	289	31.00	5.34	3	5	C
CcLHCA2	LHCA2	Cc11_g16910	1340	287	31.47	6.21	3	2	Other

MW : molecular weight, TM : transmembrane helix, pI : isoelectrical point, C: chloroplast, M: Mitochondria

Classification of coffee LHC genes

The coffee LHCs were classified based on phylogenetic tree which was constructed from LHC proteins of three species including rice, A.thaliana and Coffee (Figure 1). The results of phylogeny analysis show that the coffee LHCs are divided into many groups, the chlorophyll a/b-binding proteins of PSI light-harvesting, the chlorophyll a/b-binding proteins of PSII light-harvesting, the LHC-related proteins and light-inducible proteins. The first group contains six members classified into six subgroups (A1-A6), with one gene in each subgroup like in *A. thaliana* and rice. These predicted protein sequences are not so different in size (ranging from 244 to 287 amino acids) and in theoretical pI (ranging from 5,52 to 6.60). The second group includes 14 genes divided into six subgroups (B1-B6). Similarly to *A. thaliana* and rice, the B5 and B6 subgroups have single locus gene while B1 subgroup has multiple genomic loci. The size of coffee B1 subgroup is similar (five members) to *Arabidopsis* but larger than in rice which contains only three genes. The B3 subgoup includes four genes in coffee while this subgroup contains only one member in *Arabidopsis* as well as in rice. At the opposite, two subgroups (B2 and B4) contain only one member each in coffee when they include three genes in *Arabidopsis*. These data suggest different evolution of PSII LHCs between coffee and *Arabidopsis*. In addition, the phylogenetic tree suggests a common ancestor of multiple genomic loci each PSII subgroups before speciation

between monocotyledons and dicotyledons. This subgroup expansion results from gene duplication events taking place in each species, *A. thaliana*, rice and coffee.

Furthermore, three additional LHC related genes were identified in coffee genome, likely in *A.thaliana* and rice. Their amino acid sequences are relatively conserved between three plants, *A. thaliana*, rice and coffee. Coffee CcPsbS is ortholog of *Arabidopsis* PsbS (At1G44575) and two rice PsbSs (LOC_Os01g64960 and LOC_Os04g59440). In *Arabidopsis*, PsbS protein, subunit of photosystem II, plays a key role in nonphotochemical quenching function in the regulation of photosynthetic light harvesting. This protein is needed for photoprotective thermal dissipation of excess absorbed light energy in plants (Niyogi *et al.*, 2005). At the amino acid level, the homology is quite high between orthologs: the CcPsbS sequence of 274 amino acids exhibits 72 /79 % of identity/similarity with AtPsbS,75/82% and 73/79% with OsPsbS1 and OsPsbS2, respectively.

The CcChla/b-like deduced protein shows similarity level of 75/90% for 285 amino acids with F14G6.17 (At1G76570) of *A. thaliana* and of 75/87% for 281 amino acids with rice Chl a/b (LOC_Os09g12540), respectively. While CcLIL is ortholog of *Arabidopsis* LIL3:1 (At4G17600, homology level at %63/74 for 218 amino acids) and rice LIL (LoC_os02g03330, homology level at %79/89 for 158 amino acids). However, any ortholog of *Arabidopsis* F21B23.110 (AT5G28450), another chlorophyll a/b-binding protein was found in coffee genome.

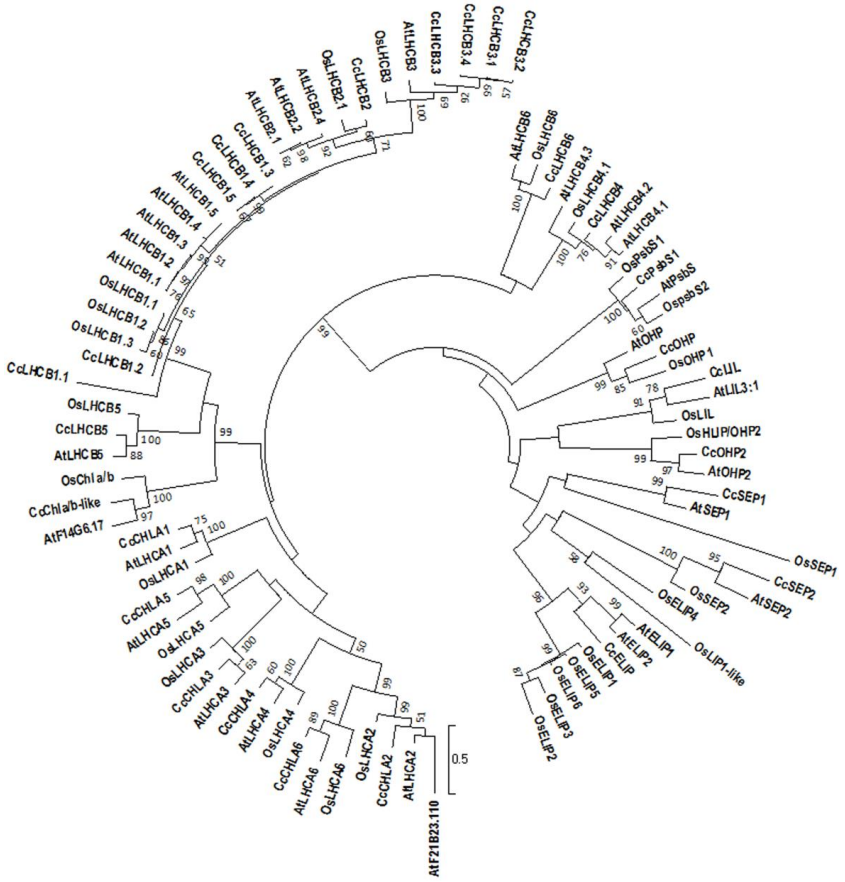

Figure 1. Phylogenetic tree of the LHC family from *A. thaliana* (At), rice (Os) and coffee (Cc). The tree was generated using Mega 5 program by Maximum Likelihood method. Bootstrap values are indicated at each branch.

In coffee genome, five light-inducible genes have been isolated in contrast to the six and eleven orthologs reported in genome of *A. thaliana* and rice, respectively (Umate, 2010). Among them, two one-helix proteins (CcOHP and CcOHP2) are orthologs of high light-inducible protein. Two two-helices (CcSEP1 and CcSEP2) are orthologs of stress-enhanced proteins. These four proteins do not contain typical Chloroa_b-bind (PF00504) conserved domain. But they are relatively identical to OHP and SEP of other plants at the amino acid level. CcOHP exhibits 74/86% of identity/positif for 81 amino acids with AtOHP (At5G02120) and 81/91% for 74 amino acids with OsOHP1 (LOC_os05g2273). While CcOHP2 shows similarity level of 67/73% (identities/positives) for 153 amino acids with AtOHP2 (At1G34000) and 56/65% for 170 amino acids with rice OsOHP2 (LOC_os01g40710). The protein alignment of OHPs in rice, *Arabidopsis* and coffee is presented in figure

2. Similarly, CcSEP1 shares similarity level of 55/71% for 144 amino acids with *Arabidopsis* SEP1 (At4G34190) when CcSEP2 shows similarity level of 60/75% for 162 amino acids with *Arabidopsis* SEP2 (At2G21970) and 51/68% for 166 amino acids with rice OsSEP2 (LOC_Os04g54630). The full-length amino acid aligment of SEPs in rice, A.thaliana and coffee is shown in figure 3. The remaining, ortholog of early light-inducible proteins of *A.thaliana* and rice, was named as CcELIP. At amino acid level, CcELIP shares homology of 60/73% 191 amino acid with tea ELIP2 (genbank accession FE942102) and %51/66% with *Arabidopsis* ELIP2 (At4G14690). The full-length amino acid aligment of ELIP in rice, *Tortula ruralis*, *A.thaliana*, tea and coffee is shown in figure 4.

Figure 2. Full-length amino acid alignment of OHP proteins in rice (Os), *Arabidopsis thaliana* (At) and Coffee (Cc) by using MAFFT. Asterisks and dots drawn on bottom of sequence indicate identical residues and conservative amino acid changes, respectively. Helix motif is noted by line on top.

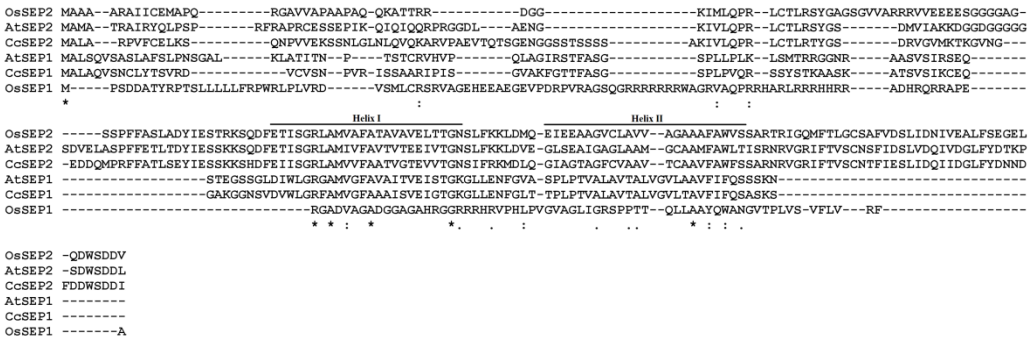

Figure3. Full-length amino acid alignment of SEP proteins in rice (Os), *Arabidopsis* thaliana (At) and Coffee (Cc) by using MAFFT. Asterisks and dots drawn on bottom of sequence indicate identical residues and conservative amino acid changes, respectively. Helix motifs are noted by line on top.

Figure 4. Full-length amino acid alignment of ELIP proteins in Tortula ruralis (Tr), rice (Os), *Arabidopsis* thaliana (At), thea (Cs), and Coffee (Cc) by using MAFFT. Asterisks and dots drawn on bottom of sequence indicate identical residues and conservative amino acid changes, respectively. Helix motifs are noted by line on top.

Expression of coffee LHC genes

The expression of the *LHC* genes was analyzed via the *in silico* analyses from transcriptome (RNAseq) data of coffee (*C. canephora*) tissues (Denoeud *et al.*, 2014). Expression analysis was performed on endosperm, perisperm, leaf, pistil, stamen, and root. These data indicate that most of the coffee *LHC* genes are strongly expressed in leaf and perisperm tissues while their expression level is weakest in rootcompared to all investigated tissues. Among the PSI *LHC* genes (A1-A6), *CcLHCA5* is the less expressed in all analyzed tissues. The other genes are highly expressed in leaves and perisperm, while they are more weakly expressed in endosperm, roots, pistil and stamen. Highest expression level was observed for *CcLHCA1* gene in leaf and stamen, for *CcLHCA2* and *CcLHCA4* in endosperm and perisperm. In addition, *CcLHCA4* is the most strongly expressed gene in pistil while CcLHCA3 is the most strongly expressed in root.

Among the PSII *LHC* genes, single locus genes are expressed in all examined tissues, except in roots where the expression of *CcLHCB2* gene is very weak. These genes are the most expressed in leaves and perisperm and the

less expressed in roots. High expression levels are observed in pistil for these four genes. The expression of multiple genomic loci genes varies. Only three members of B1 subgroup (*CcLHCB1.3*-*CcLHC1.5*) are expressed in most of investigated tissues while the expression of two remaining genes is very weak or not detected. Similarly, *CcLHCB4.3* was unique gene of B4 subgroup expressed in all studied tissues. To date, the expression of PSI and LHC genes is little known in plants, especially in normal conditions. *Nicotiana sylvestris Lhcb1* transcripts are accumulated in leaves and stems but not in roots and non-green cultured cells (Hasegawa*et al.,* 2002). In general, expression of the *LHCB* genes is regulated by multiple environmental and developmental factors (for review, see Xu *et al.,* 2012).

The expression of three *LHC*-related genes is detected in various tissues with highest expression in leaf and perisperm. *CcChla/b-like* and *CcLIL* genes expressed in all tissues. *LIL* gene was known to play important role in the chlorophyll and tocopherol biosynthesis in *A.thaliana* (Tanaka *et al.,* 2010). The expression of *CcPsbS*, subunit of PSII complex, was oberved in leaf, perisperm, pistil and stamen. *Arabidopsis* orthologs play a key role in nonphotochemical quenching function in the regulation of photosynthetic light harvesting. This protein is important for photoprotective thermal dissipation of excessive absorbed light energy in plants (Niyogi *et al.,* 2005). Expression of *PsbS* gene was responsive to high light in *Arabidopsis* (Li *et al.,* 2000). While psbS transcription seems to be influenced only by phytochrome and not by the blue light low fluence system in spinach (Adamska *et al.,* 1996). Recently, accumulation of PsbS transcrits under various cadmium concentrations was reported in *Sedum alfredii* ecotypes (Zhang and Yang, 2014).

The expression of light-inducible genes have been detected in all tissues. In particular, *CcELIP* gene strongly expressed in leaves, endosperm, perisperm, stamen and pistil. Interestingly, this gene is highly expressed in pistil, a reproductive tissu. The expression of light-inducible genes is most studied among *LHC* and *LHC*-related genes in other plants. Light-stress induced the expression of *OHP*, *SEP* and *ELIP* genes in many plants, such as *A. thaliana* OHP (Andersson *et al.,* 2003; Heddad and Adamska, 2000; Heddad *et al.,* 2006). The expression of *ELIP* genes is induced by many abiotic stress including cold, drought, high temperature and salinity in other plants (Adamska and Kloppstech, 1994; Berti and Pinto, 2012; Montane and Kloppstech, 2000; Peng *et al.,* 2008; Wang *et al.,* 2014). In addition, expression of ELIPs were influenced by developmental stage of pea (NorÉN *et al.,* 2003) and leaf senescence of Nicotiana tabacum (Binyamin *et al.,* 2001). In this work, the expression of *LHC* genes in coffee under normal condition was reported, suggesting that these genes play a constitutive role in both vegetative and

reproductive tissues of this plant.

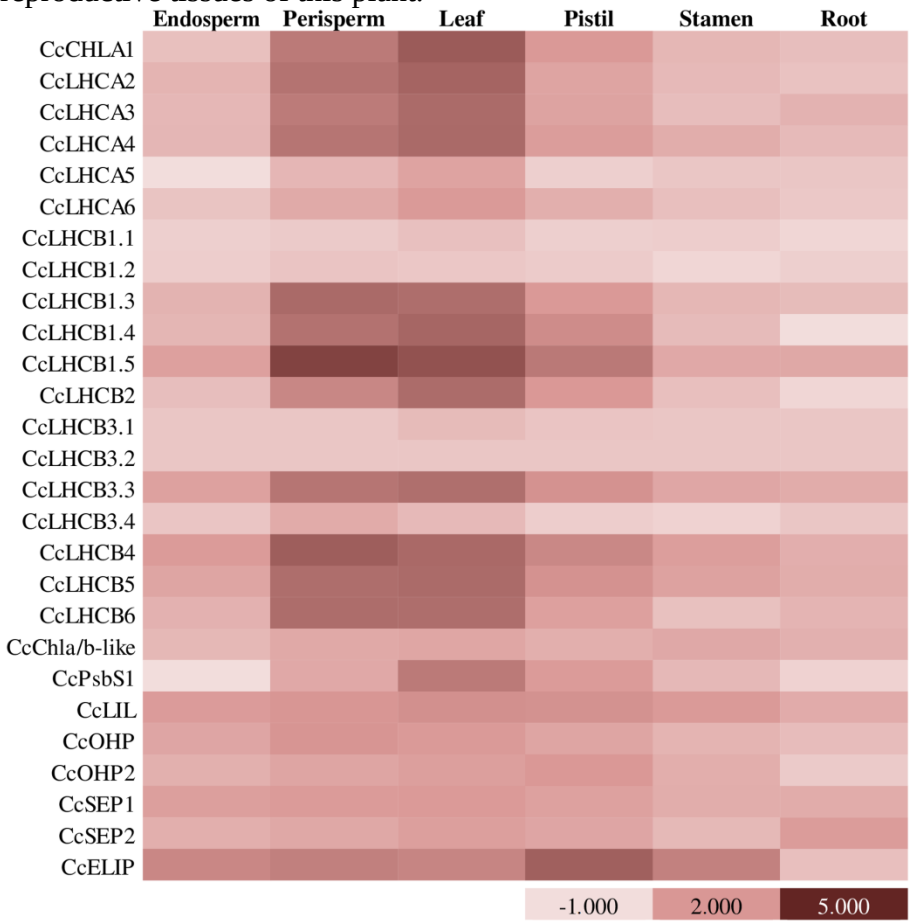

Figure 5. Heatmap showing expression level of coffee *LHC* genes in six organs. Color scale represents RPKM normalized log10 transformed counts. Light red indicates low expression and red indicates high expression.

Conclusion

By using *in silico* methods, a total of 28 putative *LHC* encoding genes were found in coffee (*Coffea cenaphora*) genome. Twenty-three out of twenty-eight putative LHC deduced proteins exhibit the *Chloroa_b-bind* (PF00504) conserved domain. Based on phylogeny analysis, these coffee *LHC* genes were classified into many groups, including PSI (six genes), PSII (14 genes), LHC-related genes (three genes) and light-inducible genes (five genes). The PSI *LHC* genes were divided into six subgroups (A1-A6) similarly to the PSII *LHC* genes (B1-B6). In agreement to the situation in rice and *A.thaliana*, the *B5* and *B6* subgroup include one gene each while *B1* subgroup contains muliple

genomic loci (five members). In adverse, *B2* and *B4* are single locus subgroup in coffee but they are multiple genomic loci in both *A. thaliana* and rice. However, *B3* subgroup contains four genes in coffee but this subgroup has only one member in *A. thaliana*. In general, the transcript of most of coffee putative *LHC* genes were strongly detected in leaves and perisperm but weakly or no detected in roots. In addition, most of these genes are expressed in pistil. The coffee early light-inducible protein encoding gene strongly expressed in all examinaed tissues.

List of abbreviations

CAB: chlorophyll a/b-binding protein; ELIP: early light-induced protein; EST: expressed sequence tag; HLIP: high light-induced protein; LHC: light-harvesting complex; SEP: stress-enhanced proteins;

Acknowledgments

This work was funded by fundamental research programme of Hung Vuong University. We acknowledge Pr C. Teulières and Dr C. Marques from Paul Sabatier University (France) for critical reviewing of the manuscript and English editing.

References

Adamska, I., Funk, C., Renger, G., and Andersson, B. (1996). Developmental regulation of the *PsbS* gene expression in spinach seedlings: the role of phytochrome. *Plant Mol Biol, 31*(4), 793-802.

Adamska, I., and Kloppstech, K. (1994). Low temperature increases the abundance of early light-inducible transcript under light stress conditions. *J Biol Chem, 269*(48), 30221-30226.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res, 25*, 3389-3402.

Andersson, U., Heddad, M., and Adamska, I. (2003). Light Stress-induced one-helix protein of the Chlorophyll a/b-binding family associated with photosystem I. *Plant Physiology, 132*(2), 811-820.

Berti, M., and Pinto, M. (2012). Expression of early light induced protein in Grapevine and Pea, under different conditions and its relation with photoinhibition. *Chilean journal of agricultural research, 72*, 371-378.

Binyamin, L., Falah, M., Portnoy, V., Soudry, E., and Gepstein, S. (2001). The early light-induced protein is also produced during leaf senescence of Nicotiana tabacum. *Planta, 212*(4), 591-597.

Buchan, D.W., Minneci, F., Nugent, T.C., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED protein analysis Workbench. *Nucleic Acids Res, 41*(Web Server issue), W349-357.

Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., . . . Lashermes, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science, 345*(6201), 1181-1184.

Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., . . . Droc, G. (2015). The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Res, 43*(Database issue), D1028-1035.

Dittami, S.M., Michel, G., Collen, J., Boyen, C., and Tonon, T. (2010). Chlorophyll-binding proteins revisited--a multigenic family of light-harvesting and stress proteins from a brown algal perspective. *BMC Evol Biol, 10*, 365. doi:10.1186/1471-2148-10-365

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc, 2*(4), 953-971. doi:10.1038/nprot.2007.131

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., . . . Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res, 42*(Database issue), D222-230. doi:10.1093/nar/gkt1223

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R..D., and Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res, 31*, 3784-3788.

Green, B. R. (2001). Was "molecular opportunism" a factor in the evolution of different photosynthetic light-harvesting pigment systems? *Proc Natl Acad Sci U S A* (Vol. 98, pp. 2119-2121). United States.

Green, B.R., and Kuhlbrandt, W.(1995). Sequence conservation of light-harvesting and stress-response proteins in relation to the three-dimensional molecular structure of LHCII. *Photosynth Res, 44*(1-2), 139-148.

Hasegawa, K., Yukawa, Y., Sugita, M., and Sugiura, M. (2002). Organization and transcription of the gene family encoding chlorophyll a/b-binding proteins in Nicotiana sylvestris. *Gene, 289*(1-2), 161-168.

Heddad, M., and Adamska, I. (2000). Light stress-regulated two-helix proteins in *Arabidopsis* thaliana related to the chlorophyll a/b-binding gene family. *Proc Natl Acad Sci U S A, 97*(7), 3741-3746.

Heddad, M., Noren, H., Reiser, V., Dunaeva, M., Andersson, B., and Adamska, I. (2006). Differential expression and localization of early light-induced proteins in *Arabidopsis*. *Plant Physiol, 142*(1), 75-87.

Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K., and Havaux, M. (2003). Early light-induced proteins protect *Arabidopsis* from photooxidative stress. *Proc Natl Acad Sci U S A, 100*(8), 4921-4926. doi:10.1073/pnas.0736939100

Jansson, S. (1994). The light-harvesting chlorophyll a/b-binding proteins. *Biochim Biophys Acta, 1184*(1), 1-19.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol, 30*(4), 772-780. doi:10.1093/molbev/mst010

Lashermes, P., Combes, M.C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., and Charrier, A. (1999). Molecular characterisation and origin of the Coffea arabica L. genome. *Molecular and General Genetics MGG, 261*(2), 259-266. doi:10.1007/s004380050965

Leroy, T., Ribeyre, F., Bertrand, B., Charmetant, P., Dufour, M., Montagnon, C., . . . Pot, D. (2006). Genetics of coffee quality. *Brazilian Journal of Plant Physiology, 18*, 229-242.

Li, X.P., Bjorkman, O., Shih, C., Grossman, A.R., Rosenquist, M., Jansson, S., and Niyogi, K.K. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature, 403*(6768), 391-395. doi:10.1038/35000131

Li, X.W., Liu, H.J., Xie, S.X., and Yuan, H.Y. (2013). Isolation and characterization of two genes of the early light-induced proteins of Camellia sinensis. *Photosynthetica, 51*(2), 305-311.

Liu, R., Xu, Y.H., Jiang, S.C., Lu, K., Lu, Y.F., Feng, X.J., . . . Zhang, D.P. (2013). Light-harvesting chlorophyll a/b-binding proteins, positively involved in abscisic acid signalling, require a transcription repressor, WRKY40, to balance their function. *J Exp Bot, 64*(18), 5443-5456.

Montane, M.H., and Kloppstech, K. (2000). The family of light-harvesting-related proteins (LHCs, ELIPs, HLIPs): was the harvesting of light their primary function? *Gene, 258*(1-2), 1-8.

Neilson, J.A., and Durnford, D.G. (2010). Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res, 106*(1-2), 57-71.

Niyogi, K.K., Li, X.P., Rosenberg, V., and Jung, H.S. (2005). Is PsbS the site of non-photochemical quenching in photosynthesis? *J Exp Bot, 56*(411), 375-382. doi:10.1093/jxb/eri056

NorÉN, H., Svensson, P., Stegmark, R., Funk, C., Adamska, I., and Andersson, B. (2003). Expression of the early light-induced protein but not the PsbS protein is influenced by low temperature and depends on the developmental stage of the plant in field-grown pea cultivars. *Plant, Cell and Environment, 26*(2), 245-253. doi:10.1046/j.1365-3040.2003.00954.x

Peng, Y., Lin, W., Wei, H., Krebs, S.L., and Arora, R. (2008). Phylogenetic analysis and seasonal cold acclimation-associated expression of early light-induced protein genes of Rhododendron catawbiense. *Physiol Plant, 132*(1), 44-52. doi:10.1111/j.1399-3054.2007.00988.x

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S..(2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol, 28*(10), 2731-2739. doi:10.1093/molbev/msr121

Tanaka, R., Rothbart, M., Oka, S., Takabayashi, A., Takahashi, K., Shibata, M., . . . Tanaka, A. (2010). LIL3, a light-harvesting-like protein, plays an essential role in chlorophyll and tocopherol biosynthesis. *Proc Natl Acad Sci U S A, 107*(38), 16721-16725. doi:10.1073/pnas.1004699107

Umate, P. (2010). Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in *Arabidopsis* and rice. *Plant Signal Behav, 5*(12), 1537-1542. doi:10.4161/psb.5.12.13410

Vieira, L.G.E., Andrade, A.C., Colombo, C.A., Moraes, A.H.d.A., Metha, Â., Oliveira, A.C.d., . . . Pereira, G.A.G. (2006). Brazilian coffee genome project: an EST-based genomic resource. *Brazilian Journal of Plant Physiology, 18*, 95-108.

Wang, H., Cao, F., Li, G., Yu, W., and Aitken, S. (2014). The transcript profiles of a putative early light-induced protein (ELIP) encoding gene in Ginkgo biloba L. under various stress conditions. *Acta Physiologiae Plantarum, 37*(1), 1-12. doi:10.1007/s11738-014-1720-8

Xu, Y.H., Liu, R., Yan, L., Liu, Z.Q., Jiang, S.C., Shen, Y.Y., . . . Zhang, D.P. (2012). Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in *Arabidopsis*. *J Exp Bot, 63*(3), 1095-1106. doi:10.1093/jxb/err315

Zhang, M., and Yang, X.E. (2014). PsbS expression analysis in two ecotypes of Sedum alfredii and the role of SaPsbS in Cd tolerance. *Russian Journal of Plant Physiology, 61*(4), 427-433. doi:10.1134/S1021443714040219