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 Bang¹*, Tran Thi Thanh Huyen²

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


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 canephora) 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 B1 and B3 subgroup which contain multiple genomic loci (five members). The B2 and B4 are single locus subgroups in C. canephora but they are multiple genomic loci in both A. thaliana and rice. In contrast, B3 subgroup contains four genes in C. canephora 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@hv.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 and analyzed in some plant species such as brown alga (Dittamiet 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 (Denœud 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 (Denœud 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 (Denœud et al., 2014). The analysis of LHC encoding genes in coffee is particularly relevant since coffee 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. canephora genome by using in silico methods. In addition, the
classification and expression of these \textit{LHC} genes were analyzed.

\textbf{Material and methods}

\textbf{Identification of LHC from coffee genomic sequences}

Based on the coffee genome (http://coffee-genome.org/) (Dereeper \textit{et al}., 2015), an extensive research was performed for identifying all members of the LHC family. Firstly, the LHC sequences of \textit{A. thaliana} (Umate, 2010) were used as queries to perform blastp (Altschul \textit{et al}., 1997) against coffee genome database with an \textit{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 \textit{Arabidopsis} orthologs. Finally all candidate sequences were submitted to domain research by using the Pfam software (http://www.sanger.ac.uk/software/pfam) (Finn \textit{et al}., 2014).

\textbf{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 \textit{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 \textit{et al}., 2007). The transmembrane protein topology was predicted by using the PSIPRED Server (http://bioinf.cs.ucl.ac.uk/psipred/) (Buchan \textit{et al}., 2013).

Phylogenetic analyses were conducted using MEGA version 5 (Tamura \textit{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.

\textbf{Gene expression analyses}

The relative \textit{LHC} gene expression log10(RPKM) (Reads Per Kilobase of transcript per Million fragments mapped) values were obtained from (Denoeud \textit{et al}., 2014).
Results and Discussions

Identification of the HLC genes in coffee genome

The LHC protein sequences from *C. cenchora* 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 exhibit 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*

<table>
<thead>
<tr>
<th>Gen</th>
<th>Subgroup</th>
<th>Locus name</th>
<th>Genomic full length (bp)</th>
<th>Protien full length (aa)</th>
<th>MW (kDa)</th>
<th>pI</th>
<th>TM</th>
<th>Intron number</th>
<th>Subcellular location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcLHCB3.1</td>
<td>LHC3</td>
<td>Cc00_g127 90</td>
<td>510</td>
<td>169</td>
<td>18.5</td>
<td>4.5</td>
<td>8</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>CcLHCB3.2</td>
<td>LHC3</td>
<td>Cc00_g348 70</td>
<td>510</td>
<td>169</td>
<td>18.5</td>
<td>4.5</td>
<td>8</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>CcSEP1</td>
<td>SEP1</td>
<td>Cc02_g191 80</td>
<td>2242</td>
<td>143</td>
<td>14.5</td>
<td>9.7</td>
<td>8</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>CcLHCB6</td>
<td>LHC6</td>
<td>Cc02_g217 20</td>
<td>999</td>
<td>262</td>
<td>27.7</td>
<td>8.7</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CcLHCB1.1</td>
<td>LHC1</td>
<td>Cc02_g335 60</td>
<td>1841</td>
<td>229</td>
<td>25.8</td>
<td>9.9</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CcELIP</td>
<td>ELIP</td>
<td>Cc03_g043 00</td>
<td>806</td>
<td>191</td>
<td>20.2</td>
<td>8.9</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCA4</td>
<td>LHCA4</td>
<td>Cc04_g164 10</td>
<td>1002</td>
<td>252</td>
<td>27.7</td>
<td>6.1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCB1.2</td>
<td>LHC1</td>
<td>Cc05_g096 50</td>
<td>802</td>
<td>242</td>
<td>25.7</td>
<td>5.3</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CcLHCA3</td>
<td>LHCA3</td>
<td>Cc05_g099 30</td>
<td>1257</td>
<td>273</td>
<td>29.3</td>
<td>6.4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCB3.3</td>
<td>LHC3</td>
<td>Cc05_g127 20</td>
<td>1012</td>
<td>263</td>
<td>28.4</td>
<td>5.0</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCB4</td>
<td>LHC4</td>
<td>Cc06_g014 60</td>
<td>1087</td>
<td>286</td>
<td>31.1</td>
<td>5.5</td>
<td>9</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>CcOHP</td>
<td>OHP</td>
<td>Cc06_g023 40</td>
<td>596</td>
<td>122</td>
<td>13.3</td>
<td>9.8</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCA6</td>
<td>LHCA6</td>
<td>Cc06_g124 80</td>
<td>1761</td>
<td>260</td>
<td>28.6</td>
<td>5.5</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>CcLHCB3.4</td>
<td>LHC3</td>
<td>Cc07_g002 60</td>
<td>510</td>
<td>169</td>
<td>18.5</td>
<td>4.5</td>
<td>3</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>CcSEP2</td>
<td>SEP2</td>
<td>Cc07_g161 20</td>
<td>1471</td>
<td>209</td>
<td>22.6</td>
<td>5.2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>CcLIL</td>
<td>LIL</td>
<td>Cc08_g113 60</td>
<td>2526</td>
<td>261</td>
<td>28.9</td>
<td>5.0</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CcLHCB1.3</td>
<td>LHC1</td>
<td>Cc09_g090 30</td>
<td>795</td>
<td>264</td>
<td>28.1</td>
<td>5.4</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CcLHCB2</td>
<td>LHC2</td>
<td>Cc09_g095 00</td>
<td>1851</td>
<td>265</td>
<td>28.5</td>
<td>5.9</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>CcLHCB1.4</td>
<td>LHC1</td>
<td>Cc09_g090 10</td>
<td>795</td>
<td>264</td>
<td>28.2</td>
<td>5.6</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CcLHCB1.5</td>
<td>LHC1</td>
<td>Cc09_g090 20</td>
<td>795</td>
<td>264</td>
<td>28.2</td>
<td>5.2</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CcCHLAI</td>
<td>LHCA1</td>
<td>Cc09_g020 10</td>
<td>1334</td>
<td>244</td>
<td>26.2</td>
<td>5.8</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 1 continued

<table>
<thead>
<tr>
<th>Gen</th>
<th>Subgroup</th>
<th>Locus name</th>
<th>Genomic full length (bp)</th>
<th>Protein full length (aa)</th>
<th>MW (kDa)</th>
<th>pI</th>
<th>TM</th>
<th>Intron number</th>
<th>Subcellular location</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcChlα/b-like</td>
<td>CHLα/b-like</td>
<td>Cc10_g00140</td>
<td>4057</td>
<td>330</td>
<td>36.3</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>CcLHCA5</td>
<td>LHCA5</td>
<td>Cc10_g04190</td>
<td>1639</td>
<td>268</td>
<td>28.9</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>CcPsbS</td>
<td>PsbS</td>
<td>Cc10_g11890</td>
<td>2312</td>
<td>271</td>
<td>28.4</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>CcOHP2</td>
<td>OHP2</td>
<td>Cc10_g15180</td>
<td>3125</td>
<td>188</td>
<td>20.0</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>C</td>
</tr>
<tr>
<td>CcLHCB5</td>
<td>LHCB5</td>
<td>Cc10_g16210</td>
<td>1632</td>
<td>289</td>
<td>31.0</td>
<td>0</td>
<td>6</td>
<td>3</td>
<td>C</td>
</tr>
<tr>
<td>CcLHCA2</td>
<td>LHCA2</td>
<td>Cc11_g16910</td>
<td>1340</td>
<td>287</td>
<td>31.4</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>Other</td>
</tr>
</tbody>
</table>

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 subgroup 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/positive 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 51/66% for 166 amino acids with rice OsSEP2 (LOC_Os04g54630). The full-length amino acid alignment 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 alignment of ELIP in rice, A.thaliana 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.

Figure 3. 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 root compared 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 observed 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 transcripts 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 tissue. 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 cenanphora*) genome. Twenty-three out of twenty-eight putative LHC deduced proteins exhibit the Chloroa 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 multiple
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 examined 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**


