Chiang Mai J. Sci. 2019; 46(5) : 867-879
http://epg.science.cmu.ac.th/ejournal/
Contributed Paper

Genome-scale Identification and Analysis of Genes Encoding Putative Light-harvesting Chlorophyll a/b-binding Proteins in Potato (*Solanum tuberosum* L.)

Phi Bang Cao* [a], Thi Thanh Huyen Tran [b], Van Dinh Nguyen [c], Viet Hong La [c] and Sahar Azar [d]

[a] Biotechnology Research Center, Faculty of Natural Sciences, Hung Vuong University, Vietnam.
[b] Department of Plant Physiology and Application, Faculty of Biology, Hanoi National University of Education, Vietnam.
[c] Department of Plant Physiology, Faculty of Biology-Agronomy, Hanoi Pedagogical University 2, Vietnam.
[d] Department of Life and Earth Sciences, Faculty of Sciences, Fanar Campus, Lebanese University, Lebanon.

*Author for correspondence; e-mail: phibang.cao@hvu.edu.vn

Received: 28 June 2018
Revised: 15 May 2019
Accepted: 16 May 2019

Abstract

Light-harvesting complex proteins (LHCs) are essential photosynthetic pigment-binding components within the thylakoid membrane. These proteins are encoded by one of the most complex gene families in higher plants. They transfer light energy to photosynthetic reaction center and play a major role in photoprotection and abiotic stress tolerance in many plants. Here, we identified a total of 46 putative LHC encoding genes in potato (*Solanum tuberosum*) genome by using *in silico* approaches. Most of the LHC deduced proteins (38 out of 46) exhibit the Chloroa_b-bind (PF00504) conserved domain. The potato *LHC* genes were classified into groups based on the phylogeny analysis, including PSI (9 genes), PSII (26 genes), LHC-related genes (5 genes), and light-inducible genes (6 genes). The PSI and PSI *LHC* genes were sorted into six subgroups and were designated as A1-A6 and B1-B6, respectively. Three PSI LHC subgroups contained two genes each, and three others are single-locus gene subgroups. Surprisingly, an expansion of B1 subgroup, resulting from recent gene duplication events, was observed in this crop genome. Generally, expression of most of the putative potato *LHC* genes was detected in leaves, except *StLIL1*. Moreover, the *LHC* genes were expressed more abundantly in aerial vegetative or reproductive tissues than underground tissues.

Keywords: potato, chlorophyll a/b-binding protein, light-harvesting complex proteins, gene identification, gene classification, gene expression.

1. Introduction

Plants use two types of light-harvesting complexes (LHC-I and LHC-II) to absorb light energy for photosynthesis. The light-harvesting proteins bind to photosynthetic pigmentsand transfer light energy to photosynthetic reaction center [1]. These light-harvesting proteins are

present in different taxa including cyanobacteria, purple bacteria, and green sulfur bacteria, showing a low level of sequence similarity although some structural similarity can be observed [2]. 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) [1].

The structures of LHC proteins from many different species such as algae and higher plants contain three transmembrane helices together with characteristic LHC motif (ExxxxRxAM) [3]. LHC proteins play an important role in light absorption and photoprotection [4]. The LHC proteins of PSII (LHCB proteins), involved in the stomatal response to abscisic acid, are important for drought tolerance in *Arabidopsis thaliana* [5]. The light-induced proteins were the most investigated among the LHC-related proteins. These proteins play a significant role in photoprotection and abiotic stress response in a large number of species, for instance *A. thaliana* [6], *Vitis vinifera* (grape vine) and *Pisum sativum* (pea) [7]. Because of the public availability of the genome sequences, the *LHC* gene family has been genome-wide identified and analyzed in some plant species, such as *A. thaliana*, *Oryza sativa* (rice) [8], *Coffea canephora* (coffee) [9] and *Zostera marina*, a seegrass species [10] .

Potatoes (*Solanum tuberosum* L.), a member of the Solanaceae, are the most important non-grain crop, the fourth most important food crop for human, and are central to global food security. Potato tubers, an important dietary source of starch, proteins, antioxidants, and vitamins, function as both a storage organ and a vegetative propagation system of plants [11].

The genomic sequence of potato (*S. tuberosum* L.) which has been completely sequenced and published since 2011, is a useful resource for the research on main traits such as breeding, quality, yield, protection against pests, and abiotic stress tolerance facing climatic changes [12]. The analysis of LHC encoding genes in potato may be particularly relevant since potato, a C3 species, is widely grown in many regions where relatively high intensities of light are present [13]. While Potato photosynthesis is limited by the light reaction at low light intensities (below 700 μmol photons/m^2/s), and by the dark reaction at higher light intensities [14]. This work aims at a genome-wide identification of the putative LHC genes in potato (*S. tuberosum* L.) genome using *in silico* methods. In addition, the classification and expression of these *LHC* genes were also analyzed.

2. MATERIAL AND METHODS

2.1 Identification of LHC from Potato Genomic Sequences

Based on the public available genomic data of potato (http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Stuberosum) [12], an extensive research was performed for identifying all members of the LHC family using the LHC sequences of *A. thaliana* [8] as queries to perform tblastn [15] against the potato genome database with an e-value of le-10. The selected potato genes were used as queries for BLASTP search on the potato genome for identifying the potato paralogs that had been excluded by their dissimilarity to the *Arabidopsis* orthologs. To identify the putative domains, all candidate sequences were submitted to the online Pfam software (http://www.sanger.ac.uk/software/pfam) [16].

2.2 Sequence Analysis and Construction of the Phylogenetic Tree

The molecular weight and theoretical pI of putative sequences were calculated by the PROTPARAM tool (http://web.expasy.org/protparam/) [17]. Subcellular localization analysis of the deduced amino acids was performed

using TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) [18] and ProtComp Ver. 9.0 (http://linux1.softberry.com). The transmembrane protein topology was predicted by the PSIPRED Server (http://bioinf.cs.ucl.ac.uk/psipred/) [19].

Phylogenetic analysis was conducted using MEGA X [20]. The complete potato LHC predicted proteins were aligned using the MAFFT server (http://mafft.cbrc.jp/alignment/server/) [21], and phylogenetic tree was constructed using the Maximum Likelihood method with 1000 bootstraps.

2.3 *In Silico* Gene Expression Analysis

The *LHC* gene expression (Fragments Per Kilobase of transcript per Million fragments mapped, FPKM) values were obtained from [12] and transformed to log10 of RPKM values.

3. Results and Discussion

3.1 Identification of the *LHC* Genes in the Potato Genome

As a result, a total of forty-six full-length genes encoding putative LHC proteins have been identified (Table 1). The Chloroa_b-bind (PF00504) conserved domain was found in 38 out of the 46 candidate genes when these predicted proteins were analyzed by the Pfam application. Other eigth sequences absent of Chloroa_b-bind domain were grouped with *A. thaliana* and rice orthologs on the phylogenetic tree (Figure 1). With 46 members, the LHC gene family in potato is larger than that in *A. thaliana* and rice which comprises 30 and 29 genes, respectively.

Table 1. Inventory and characteristics of the *LHC* genes identified in *S. tuberosum*.

Gene	Subgroub	Locus name	Genomic full lenght (bp)	Protein full lenght (aa)	MW (kDa)	pI	TM	Chromosome	Subcellular location	Introns number
StLHCB6.1	B6	PGSC0003DMG400012591	988	256	27,28	6,15	3	chr01	C	1
StLHCB6.2	B6	PGSC0003DMG400012590	867	256	27,35	6,15	3	chr01	C	1
StSEP1	SEP	PGSC0003DMG400025866	2036	141	14,64	9,89	2	chr01	C	2
StLHCB1.1	B1	PGSC0003DMG400042498	798	265	28,07	5,14	3	chr02	C	0
StLHCB1.2	B1	PGSC0003DMG400008309	798	265	28,12	5,15	3	chr02	C	0
StLHCB1.3	B1	PGSC0003DMG400008297	874	242	25,70	5,95	3	chr02	C	1
StLHCB1.4	B1	PGSC0003DMG400008298	798	265	28,06	5,14	3	chr02	C	0
StLHCB1.5	B1	PGSC0003DMG400008299	798	265	28,09	5,14	3	chr02	C	0
StLHCB1.6	B1	PGSC0003DMG400008300	783	260	27,65	5,02	3	chr02	C	0
StLHCB1.7	B1	PGSC0003DMG400008301	783	260	27,65	5,14	3	chr02	C	0
StLHCB1.8	B1	PGSC0003DMG400013460	805	252	26,71	6,32	3	chr03	C	1
StLHCB1.9	B1	PGSC0003DMG400013411	789	262	27,84	5,47	3	chr03	C	0
StLHCB1.10	B1	PGSC0003DMG400013412	789	262	27,98	5,69	3	chr03	C	0
StLHCB1.11	B1	PGSC0003DMG400013413	725	225	23,71	6,82	3	chr03	C	1
StLHCB1.12	B1	PGSC0003DMG400013414	804	267	28,28	5,47	3	chr03	C	0
StLHCB1.13	B1	PGSC0003DMG400013415	789	262	27,81	5,47	3	chr03	C	0
StLHCB1.14	B1	PGSC0003DMG400013416	789	262	27,83	5,47	3	chr03	C	0
StLHCB1.15	B1	PGSC0003DMG400013461	805	246	26,19	5,11	3	chr03	C	1
StLHCB1.16	B1	PGSC0003DMG400013417	804	267	28,39	5,47	3	chr03	C	0
StLHCB1.17	B1	PGSC0003DMG401013418	804	255	26,94	5,46	3	chr03	C	1
StLIL1	LIL	PGSC0003DMG400008343	1052	253	28,80	7,76	2	chr03	C	2
StLHCA4.1	A4	PGSC0003DMG400019508	961	251	27,77	6,43	3	chr03	C	2

Table 1. Inventory and characteristics of the *LHC* genes identified in *S. tuberosum*. (Continued)

Gene	Subgroub	Locus name	Genomic full lenght (bp)	Protein full lenght (aa)	MW (kDa)	pI	TM	Chromosome	Subcellular location	Introns number
StSEP2	SEP	PGSC0003DMG400016590	1327	192	20,89	4,82	2	chr04	C	1
StOHP2A	OHP	PGSC0003DMG400025068	2001	168	18,33	9,30	1	chr04	C	1
StChla/b-like1	CHLa/b-like	PGSC0003DMG401009929	3331	318	35,38	6,60	3	chr04	C	5
StLHCA1.1	A1	PGSC0003DMG400023344	2311	246	26,62	5,42	3	chr05	C	3
StLHCA1.2	A1	PGSC0003DMG400023461	1143	246	26,57	5,61	3	chr05	C	3
StPsbS	PsbS	PGSC0003DMG400017556	1566	276	29,30	8,67	3	chr06	C	3
StLHCB5	B5	PGSC0003DMG400026500	1656	285	30,41	5,99	3	chr06	C	5
StLHCA4.2	A4	PGSC0003DMG400033084	961	250	27,60	6,43	3	chr06	C	2
StLHCA5	A5	PGSC0003DMG400004458	1335	266	28,83	7,83	3	chr07	C	4
StLHCB2.1	B2	PGSC0003DMG400006149	903	265	28,72	5,33	3	chr07	C	1
StLHCB3.1	B3	PGSC0003DMG400019248	1128	262	28,39	5,24	3	chr07	C	2
StLIL2	LIL	PGSC0003DMG400020492	5146	256	28,64	5,47	2	chr08	C	2
StLHCB1.18	B1	PGSC0003DMG400007375	987	267	28,34	5,47	3	chr08	C	1
StLIL3	LIL	PGSC0003DMG400026245	1651	266	29,48	4,90	2	chr08	C	2
StLHCB4	B4	PGSC0003DMG400008488	1738	285	31,13	5,78	3	chr09	C	1
StOHP1	OHP	PGSC0003DMG400008901	1047	112	12,22	9,25	1	chr09	C	2
StELIP	ELIP	PGSC0003DMG400006442	926	197	20,83	9,00	2	chr09	C	2
StLHCA2	A2	PGSC0003DMG400014386	2555	187	20,76	6,06	3	chr10	C	2
StLHCA3.1	A3	PGSC0003DMG400021287	1817	273	29,34	8,65	3	chr10	C	2
StLHCB3.2	B3	PGSC0003DMG400008564	1203	265	28,63	5,09	3	chr12	C	2
StLHCA3.2	A3	PGSC0003DMG400007787	1322	273	29,23	8,61	3	chr12	C	2
StLHCA6	A6	PGSC0003DMG400002901	2316	260	28,86	5,57	3	chr12	C	3
StLHCB2.2	B2	PGSC0003DMG400004301	1506	265	28,66	5,06	3	chr12	C	1
StOHP2B	OHP	PGSC0003DMG400018793	1264	156	17,16	9,41	1	chr12	C	1

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

The full-length of potato putative *LHC* genomic sequences ranges from 725 to 5146 nucleotides. Major potato *LHC* genes (33 out of 46) exhibit intron (from one to five introns). Their deduced full-length protein sequences range from 112 to 318 amino acids. Among them, StOHP1 has the smallest size with the molecular mass of 12.22 kDa while StChla/b-like has the biggest size (35.38 kDa). Theoretical pI values of potato LHCs fluctuated in a wide range from 4.82 to 9.89, with 36 acidic and 10 basic proteins. The *LHC* encoding gene family has been genome-wide identified in *A. thaliana* and rice [8]. The characteristics of potato LHC are in agreement with orthologs of *A. thaliana* and rice. StOHP1 is the smallest potato LHC proteins with 112 amino acids, while AtOHP (locus At5G02120) and OsOHP1 (LOC_Os05g22730) are also the smallest Arabidopsis and rice LHC proteins with 110 amino acids (MW of 12.01 kDa) and 113 amino acids (MW of 12.11 kDa), respectively. Consistently, theoretical pI values of LHCs range from 4.61 to 11.51 (*A. thaliana*) and from 4.14 to 12.82 (rice) [8]. Using PSIPRED server [19], the transmembrane helix predictions showed that the majority (37 out of 46) of potato LHCs consist of three helices. Three sequences exhibited one helix (StOHP1, StOHP2A and StOHP2B) when the other six ones contained two helices. The prediction of subcellular localization of the deduced amino

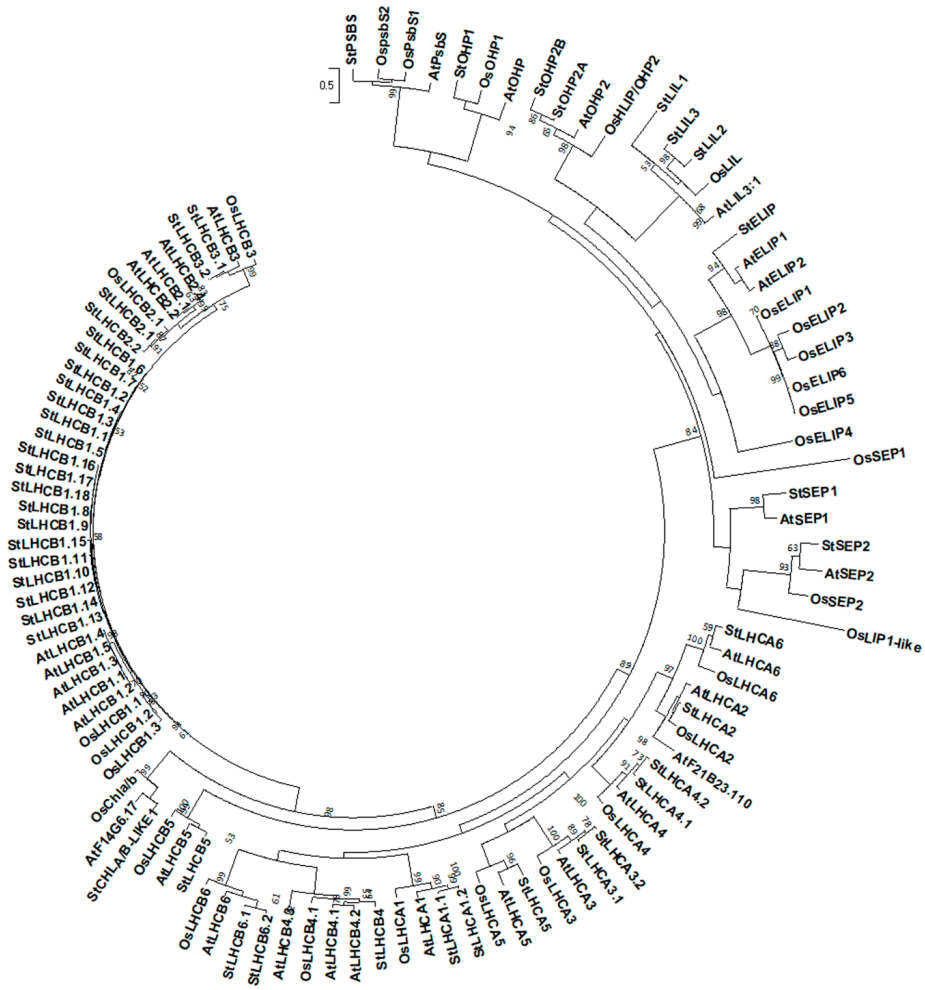

Figure 1. Phylogenetic tree of the LHC family from *A. thaliana* (At), rice (Os), and potato (St). The tree was generated using Mega X program by Maximum Likelihood method. Bootstrap values are indicated at each branch.

acids using TargetP 1.1 Server and ProtComp Version 9.0 suggested that all LHC proteins were present in chloroplast.

3.2 Classification of Potato *LHC* Genes in Comparison with Other Plant Species

The phylogenetic tree constructed from LHC proteins of three species including rice, *A. thaliana*, and potato (Figure 1) allows the classification of the potato LHCs. The results showed that the potato LHCs were classified into different groups, including the photosysstem I light-harvesting (PSI LHC), the photosystem II light-harvesting (PSII LHC), the LHC-related proteins and the light-inducible proteins. This classsification was in agreement to the situation in *A. thaliana* and rice. The PSI LHC group contained nine members which were classified into six subgroups (A1-A6). Three subgroups A1, A3, and A4 of potato comprised two genes each while three subgroups A2, A5 and A6 contained only one gene each, whilst each subgroup of *Arabidopsis* and rice contained only one gene. The potato PSII LHC group had 26 genes which were sorted into six subgroups (B1-B6). Two subgroups B4 and B5

had single-locus gene while others contained paralogous genes. The B1, B2, and B4 were multigenic subgroups in *Arabidopsis* whereas only B1 was multigenic subgroup in rice.

Particularly, B1 subgroup was very strongly over-represented in potato (Figure 2). This subgroup was composed of 18 genes corresponding to 51% of the *LHCA* and *LHCB* groups (18 out of 35 genes), which was the different from the other species (25% for *A. thaliana*, 26% for *Coffea canephora* and 21% for *Oryza sativa*). Additionally, this subgroup represented 39% of the whole LHC family in potato genome (18 out of 46 genes). However, the B1 subgroup only constituted 17%, 18% and 10% of the whole LHC family in genome of *A. thaliana*, coffee and rice, respectively.

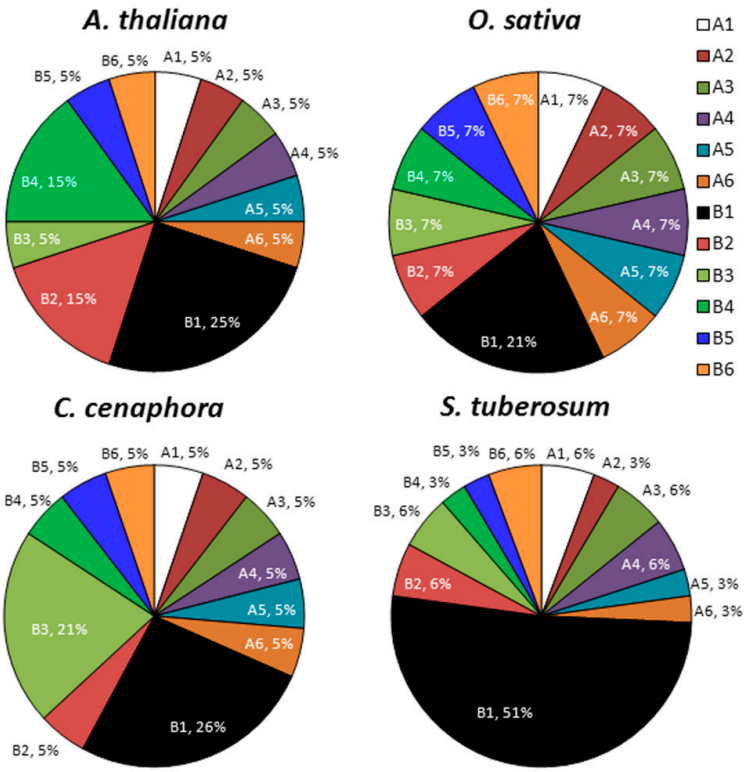

Figure 2. Distribution of genes within LHCA and LHCB subfamilies from *A. thaliana*, *O.sativa*, *C. cenaphora*,and *S.tuberosum*. Pie charts depicted as sectors the relative size of each subgroup A1-A6 and B1-B6 for the total of PSI and PSII LHC subfamilies.

Except for the *StLHCB1.18*, all other B1 genes exhibited a clustering distribution in the potato genome. These genes regrouped only two clusters in chromosome 2 (seven genes, *StLHCB.1-StLHCB1.7*) and in chromosome 3 (11 genes, StLHCB1.8-StLHCB1.17). The first cluster is located in the region about 197 Kb (from 51382817 to 51580106). Interestingly, the second one is located in the region about 32 kb (from 962814 to 995695) which does not contain any other genes (Figure 3). At the amino acid level, the homology is quite high between all B1 paralogs of potato. These amino acid sequences exhibited at least 76/80 % of the identity/similarity (Table 2).

In addition, it is suggested that the phylogenetic tree demonstrated a unique common ancestor of B1 subgroups before speciation

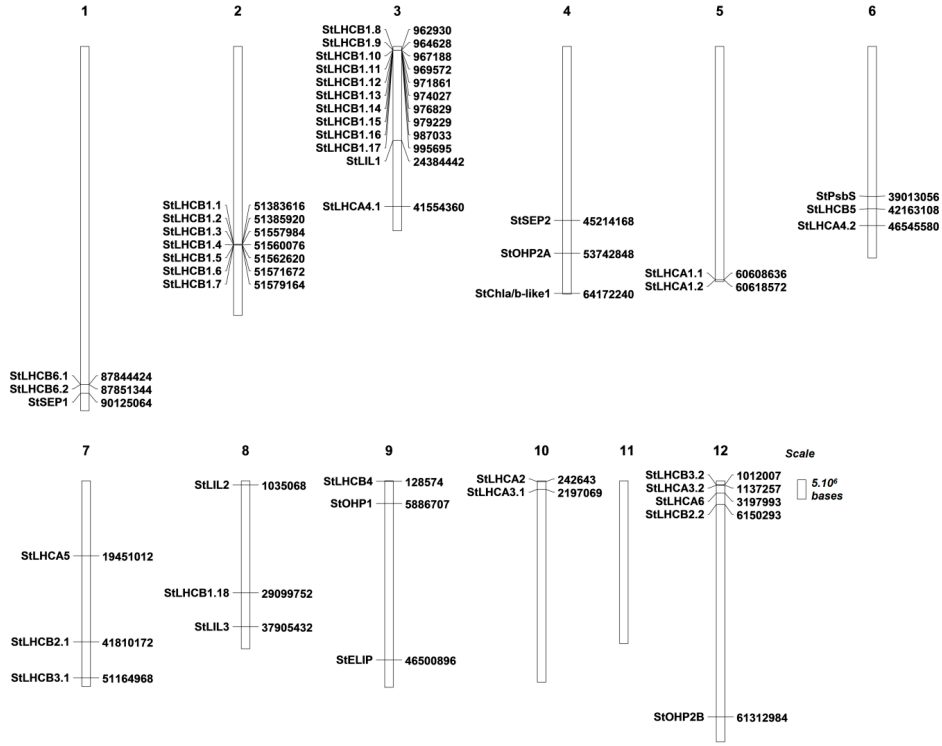

Figure 3. Genomic localization of *LHC* genes in potato chromosomes.

Table 2. Pairwise sequence comparison (% of identity/similarity) between potato LHCB1 proteins.

	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	1.12	1.13	1.14	1.15	1.16	1.17	1.18
1.1	100/100																	
1.2	98/99	100/100																
1.3	91/91	90/90	100/100															
1.4	99/100	98/99	91/91	100/100														
1.5	99/100	98/99	91/91	99/100	100/100													
1.6	98/98	98/99	89/90	98/98	98/98	100/100												
1.7	98/99	98/99	89/90	98/99	98/99	98/98	100/100											
1.8	90/92	89/92	93/95	90/92	90/92	89/92	89/92	100/100										
1.9	95/97	94/97	86/88	94/97	95/97	94/96	93/96	93/93	100/100									
1.10	94/97	93/96	85/88	94/97	94/97	93/96	92/96	92/92	97/98	100/100								
1.11	77/80	76/80	84/88	77/80	78/80	76/79	76/80	85/86	79/80	80/80	100/100							
1.12	93/96	93/96	84/88	93/96	94/96	94/96	93/96	91/92	91/92	97/98	81/81	100/100						
1.13	95/97	95/97	86/88	95/97	95/97	95/96	94/97	92/93	98/99	98/98	80/81	98/99	100/100					
1.14	94/96	94/96	85/88	94/96	94/96	94/96	93/96	91/93	97/98	97/98	79/80	98/99	99/99	100/100				
1.15	85/88	84/88	84/88	84/88	85/88	84/88	84/88	85/89	86/88	87/88	83/85	87/88	87/88	87/88	100/100			
1.16	94/97	94/97	85/88	94/97	94/97	94/96	93/96	91/92	96/97	96/97	78/79	96/97	97/98	97/98	85/87	100/100		
1.17	89/92	90/92	81/83	89/92	90/92	89/91	89/92	87/88	93/94	92/93	75/76	92/94	93/94	93/94	81/83	93/93	100/100	
1.18	95/96	94/96	86/88	94/96	94/96	94/95	94/96	92/92	98/98	97/97	79/80	97/98	98/98	97/98	86/88	97/97	93/94	100/100

between monocotyledons and dicotyledons. This subgroup expansion results from gene duplication events taking place in all of these species, *A. thaliana*, rice, and potato.

The B3 subgroup included two genes, *StLHCB3.1* and *StLHCB3.2*, locating in chromosome 7 and 12 in the potato genome while this subgroup contained only one member in Arabidopsis as well as in rice. On the contrary, the B4 subgroup of potato exhibited a similar size in comparison with that in rice (one gene) but this size was smaller than that in *Arabidopsis* (three genes). These data suggested a differential evolution of *PSII LHCs* between potato and Arabidopsis and rice.

Moreover, five *LHC-related* genes were identified in the potato genome, *StPsbs*, *StChla/b-like* and three *StLIL* genes. Their amino acid sequences are relatively conserved between three plants, *A. thaliana*, rice, and potato. Potato CcPsbS is similar to the orthologs of *Arabidopsis* PsbS (At1G44575) and rice PsbSs (LOC_Os01g64960 and LOC_Os04g59440). At the amino acid level, the homology is quite high between orthologs: the StPsbS sequence of 278 amino acids exhibits 73/80 % of identity/similarity with AtPsbS,73/82% and 80/86% with OsPsbS1 and OsPsbS2, respectively. These data indicated that StPsbs has the same function to AtSbsb. In *Arabidopsis,* PsbS protein, the subunit of photosystem II, plays a key role in non-photochemical quenching function in the regulation of photosynthetic light harvesting. This protein is needed for photoprotective thermal dissipation of excess absorbed light energy in plants [22].

The StChla/b-like deduced protein showed similarity level of 81/90% for 285 amino acids with F14G6.17 (At1G76570) of *A. thaliana* and of 76/87% for 281 amino acids with rice Chl a/b (LOC_Os09g12540), respectively. Three LIL genes were detected in the potato genome while only one sequence was found in *A. thaliana* (At4G17600) and rice (LoC_os02g03330) genome. However, no ortholog of *Arabidopsis* F21B23.110 (AT5G28450), another chlorophyll a/b-binding protein was found in the potato genome.

In the potato genome, six light-inducible genes have been identified in comparison to six and eleven orthologs in the genome of *A. thaliana* and rice, respectively [8]. Among them, three one-helix proteins (StOHP1, StOHP2A, and StOHP2B) are orthologs of a high light-inducible protein of these two other plants. Two two-helices (StSEP1 and StSEP2) are orthologs of stress-enhanced proteins. These five deduced proteins do not contain typical Chloroa_b-bind (PF00504) conserved domain. At the amino acid level, they are relatively identical to OHP and SEP of other plants. StOHP exhibits 73/82% of identity/positives for 79 amino acids with AtOHP (At5G02120) and 61/76% for 110 amino acids with OsOHP1 (LOC_os05g2273). StOHP2A shows a level of 72/75% (identities/positives) for 177 amino acids with AtOHP2 (At1G34000) and 58/68% for 160 amino acids with rice OsOHP2 (LOC_os01g40710). Whereas these identity/positives levels of StOHP2B are 68/73% for 143 amino acids with AtOHP2 and 86/95% for 80 amino acids with rice OsOHP2 (Figure 4). Similarly, StSEP1 shows a level of 57/71% for 149 amino acids with *Arabidopsis* SEP1 (At4G34190) when StSEP2 shares level of58/71% for 106 amino acids with *Arabidopsis* SEP2 (At2G21970) and 50/64% for 199 amino acids with rice OsSEP2 (LOC_Os04g54630). The full-length amino acid alignment of SEPs in *A. thaliana*, rice, and potato is presented in Figure 5. Surprisingly, only one ELIP, early light-inducible protein, was found in the potato genome while other plants included two or more genes. The full-length amino acid alignment of ELIP in rice, *Tortula ruralis*, *A. thaliana*, tea and potato is shown in Figure 6.

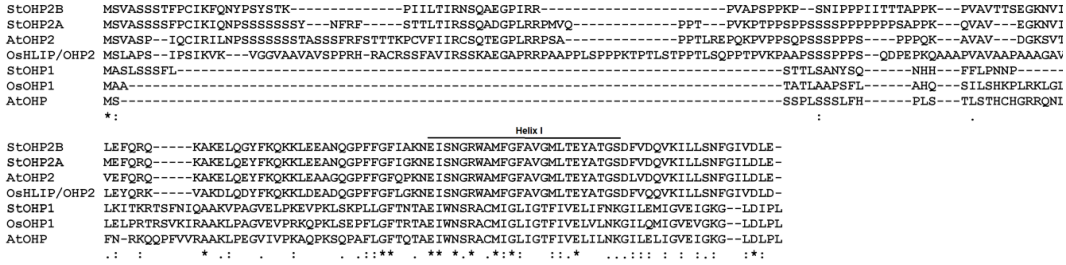

Figure 4. Full-length amino acid alignment of OHP proteins in rice (Os), *A. thaliana* (At) and potato (St) by using MAFFT. Asterisks and dots drawn on bottom of sequence indicate identical residues and conservative amino acid changes, respectively. Helix motif is noted by line on top.

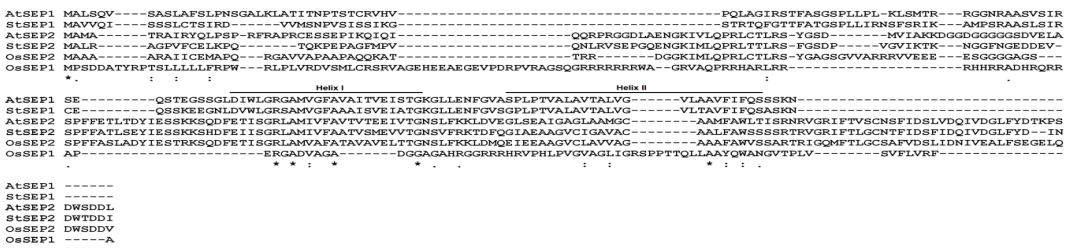

Figure 5. Full-length amino acid alignment of SEP proteins in rice (Os), *A.thaliana* (At) and Potato (St) by using MAFFT. Asterisks and dots drawn on bottom of sequence indicate identical residues and conservative amino acid changes, respectively. Helix motifs are noted by line on top.

Figure 6. Full-length amino acid alignment of ELIP proteins in Tortularuralis (Tr), rice (Os), *A. thaliana* (At), thea (Cs), and Potato (St) 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.

3.3 *In silico* Expression Analysis of Potato *LHC* Genes

The expression of the LHC genes was analysed via the *in silico* analysis from transcriptome (RNAseq) data of potato (*S. tuberosum*) tissues [12]. Expression analysis was performed on leaves, petiole, shoot apex, stem, stolon, young tuber, mature tuber, tuber pith,tuber peel, tuber sprout, tuber cortex, root, flower, and stamen. Heat map indicated that most of the potato *LHC* genes were expressed in leaf, except *StLIL1* which did not show in any tissues. In general, these genes were strongly expressed in aerial vegetative tissues (Figure 7).

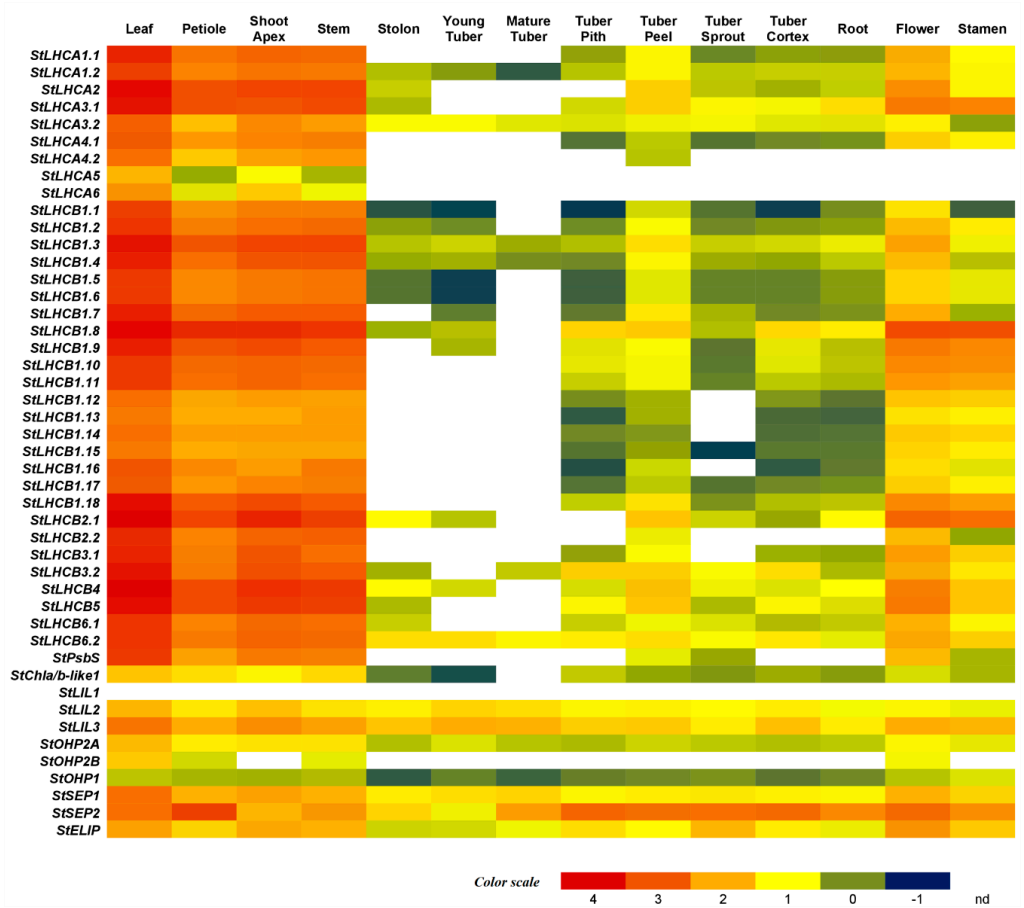

Figure 7. Heatmap showing expression level of potato *LHC* genes in 14 organs. Color scale represents log10 of RPKM counts. nd: nondetermined.

All the PSI *LHC* genes (A1-A6) were expressed in leaves, petioles, shoot apex and stems. Among the PSI *LHC* genes (A1-A6), *CcLHCA5* and *CcLHCA6* were the less expressed in all analysed tissues. The expression of two of these genes was detected only in four aerial vegetative tissues. The other genes were weakly or not expressed all together in stolon (six out of nine genes), young tuber (two out of nine genes), mature tuber (one out of nine genes), tuber pith (five out of nine), tuber peel (seven out of nine), tuber sprout, tuber cortex, and root (six out of nine). Except for *StLHCA5* and *LHCA6*, the expression of remaining genes was detected in flower. In the stamen, the expression of *LHCA4.2* was

not observed. *StLHCA3.2* was a unique gene which was expressed in all examined tissues but *StLHCA3.1* was the most strongly expressed gene in most tissues.

Generally, the expression profile of potato PSII *LHC* genes was of similarity to which of PSI *LHC* genes. All PSII *LHC* genes expressed in four aerial vegetative tissues. Leaves were the organ where the expression of these genes reached its highest. In addition, the expression of all of these genes was detected in flowers and stamens, except *StLHCB1.1* which did not express in stamens.*StLHCB1.3*, *StLHCB1.4*, and *StLHCB6.2* were the three genes which expressed in all vegetative and productive tissues. Compared to other genes, *StLHCB2.1* was the most highly expressed one in leaves, while *StLHCB1.8* in other aerial tissues, including flower and stamen. In the under aerial tissues, the expression of PSII *LHC* genes was weak or not detected.

The expression of PSI and PSII *LHC* genes has little been known in plants until now, especially under normal conditions. Accumulations of *Nicotiana sylvestris Lhcb1* transcripts were observed in leaves and stems but not in roots nor non-green cultured cells [23]. Generally, expression of the *LHCB* genes is regulated by multiple environmental and developmental factors [5]. Recently, the expression of *LHC* genes of coffee was *in silico* investigated in our previous work. These coffee genes differentially expressed in vegetative and productive tissues. Leaf and perisperm were two organs where the coffee *LHC* genes were most highly expressed [9].

Except for *StLIL1*, the expression of five *LHC*-related genes was detected in various tissues with the highest expression level in leaves. *StLIL2* and *StLIL3* expressed in all tissues while *StPsbS* and *StChla/b-like* in all aerial tissues and some under aerial tissues. In *A. thaliana, LIL* gene was described to play an important role in the chlorophyll and tocopherol biosynthesis [24] while *PsbS*, the subunit of PSII complex, played a key role in nonphotochemical quenching function in the regulation of photosynthetic light harvesting. PsbS protein was important for photoprotective thermal dissipation of excessive absorbed light energy in plants [22]. Recently, accumulation of *PsbS* transcripts under various cadmium concentrations in *Sedum alfredii* ecotypes was reported [25].

The expression of four out of six light-inducible genes was detected in all tissues. In comparison to other genes, the transcript of *StSEP2* was the most abundant in 12 out of 14 examined tissues. *StELIP* gene, coding for a three-helix early light-inducible protein, inparticular, expressed more strongly than three *OHP* genes but more weakly than *SEP* genes. This gene was most abundantly expressed in flowers. In other plants, light-stress induced the expression of *OHP*, *SEP* and *ELIP* genes in many plants, such as *A. thaliana* [26-28]. The expression of *ELIP* genes was upregulated by many abiotic stress including cold, drought, high temperature and salinity [7, 29-31]. In addition, expression of *ELIPs* was affected during development stages in pea [28] and leaf senescence in *Nicotiana tabacum* [32]. Recently, the expression of *LHC* genes in coffee under normal conditions had been reported [9]. These data contributed to a suggestion that *LHC* genes constitutively function in both vegetative and reproductive plant tissues. As important components of the light-harvesting complex, it is suggested the PSI and PSII LHC proteins are perhaps the most abundant membrane proteins in nature. Additionally, enhanced thylakoid photoprotection is reported to be able to increase yield and canopy radiation use efficiency [33]. Moreover, photosystem II Subunit S overexpression increases the efficiency of water use in a field-grown crop [34]. However, it is found that the function of LHC genes in potato remains unknown. For that reason, the

expression of the potato *LHC* genes should be the subject for further in-depth, such as comparative gene expressions, investigation under infuence of multiple developmental cues and environmental stresses. These findings are supposed to facilite the understanding of the role of LHCB proteins in the development and stress tolerance of potato, one of the most important tuber crops in the world.

4. Conclusions

In this work, a total of 46 putative LHC encoding genes were identified in potato (*Solanum tuberosum*) genome by using in silico approaches. Thirty-eight out of 46 LHC deduced proteins containthe *Chloroa_b-bind* (PF00504) conserved domain. Phylogenyanalysis suggested that these potato *LHC* genes were divided into many groups, including *PSI LHC* (9 genes), *PSII LHC* (26 genes), *LHC-related* genes (5 genes) and light-inducible genes (6 genes). The *PSI LHC* genes were classified into six subgroups (A1-A6) similar to the classification of *PSII LHC* genes (B1-B6). Three *PSI LHC* subgroups (A1, A2 and A4) contained two genes each. Interestingly, an over-presentation of B1 subgroup was observed in the potato genome. This situation resulted from recent gene duplication events. In general, transcripts of most of potato putative *LHC* genes were highly detected in leaves, except for *StLIL1*. Moreover, most of *LHC* genes wer abundantly expressed in aerial vegetative tissues or reproductive tissues. In contrast, transcripts of these genes were weakly or not detected in roots.

References

[1] Jansson S., *Biochim. Biophys. Acta,* 1994; **1184(1)**: 1-19.

[2] Green B.R., *Proc. Natl. Acad. Sci. USA.,* 2001; **98**: 2119-21.

[3] Green B.R. and Kuhlbrandt W., *Photosynth. Res.*, 1995; **44(1-2)**: 139-48.

[4] Neilson J.A. and Durnford D.G., *Photosynth. Res.*, 2010; **106(1-2)**: 57-71.

[5] Xu Y.H., Liu R., Yan L., Liu Z.Q., Jiang S.C., Shen Y.Y., Wang X.F. and Zhang D.P., *J. Exp. Bot.,* 2012; **63(3)**: 1095-106. DOI: 10.1093/jxb/err315.

[6] Hutin C., Nussaume L., Moise N., Moya I., Kloppstech K. and Havaux M., *Proc. Natl. Acad. Sci. USA.*, 2003; **100(8)**: 4921-6. DOI:10.1073/pnas.0736939100.

[7] Berti M. and Pinto M., *Chilean J. Agric. Res.*, 2012; **72**: 371-8. http://dx.doi.org/10.4067/S0718-58392012000300011

[8] Umate P., *Plant Signal. Behav.*, 2010; 5 (12): 1537-42. DOI: 10.4161/psb.5.12.13410

[9] Cao P.B. and Tran T.T.H., *J. Agric. Technol.,* 2015; **11(8)**:2547-61.

[10] Kong F., Zhou Y., Sun P., Cao M., Li H. and Mao Y., *J. Ocean Univ. China.*, 2016; **15(1)**: 152-162. doi: 10.1007/s11802-016-2688-3.

[11] Burlingame B., Mouillé B. and Charrondière R., *J. Food Compos. Anal.*, 2009; **22(6)**: 494-502. https://doi.org/10.1016/j.jfca.2009.09.001

[12] Xu X., Pan S., Cheng S., Zhang B., Mu D., Ni P., Zhang G., Yang S., Li R., Wang .J, Orjeda G., Guzman F., Torres M., Lozano R., Ponce O., Martinez D., De la Cruz G., Chakrabarti S.K., Patil V.U., Skryabin K.G., Kuznetsov B.B., Ravin N.V., Kolganova T.V., Beletsky A.V., Mardanov A.V., Di Genova A., Bolser D.M., Martin D.M., Li G., Yang Y., Kuang H., Hu Q., Xiong X., Bishop G.J., Sagredo B., Mejía N., Zagorski W., Gromadka R., Gawor J., Szczesny P., Huang S., Zhang Z., Liang C., He J., Li Y., He Y., Xu J., Zhang Y., Xie B., Du Y., Qu D., Bonierbale M., Ghislain M., Herrera Mdel R., Giuliano G., Pietrella M., Perrotta G., Facella P., O'Brien K., Feingold S.E., Barreiro L.E., Massa G.A., Diambra L., Whitty B.R., Vaillancourt B., Lin H., Massa A.N., Geoffroy M., Lundback S., DellaPenna D., Buell C.R., Sharma S.K., Marshall D.F., Waugh R., Bryan G.J., Destefanis M., Nagy I., Milbourne D., Thomson S.J., Fiers M., Jacobs J.M., Nielsen K.L., Sønderkær M., Iovene M., Torres G.A., Jiang J., Veilleux R.E., Bachem C.W., de Boer J., Borm T., Kloosterman B., van

Eck H., Datema E., Hekkert B., Goverse A., van Ham R.C. and Visser R.G., *Nature*, 2011; **475(7355)**: 189-95. DOI: 10.1038/nature10158

[13] Manrique L.A., *J. Plant Nutr.,* 1992; **15(12)**: 2679-2726. doi: 10.1080/01904169209364504.

[14] Laisk A., Eichelmann H., Oja V., Talts E. and Scheibe R., 2007. *Plant Cell Physiol.* **48(11)**: 1575-88. doi: 10.1093/pcp/pcm129.

[15] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J., *Nucleic Acids Res.,* 1997; **25**: 3389-402.

[16] Finn R.D., Bateman A., Clements J., Coggill P., Eberhardt R.Y., Eddy S.R., Heger A., Hetherington K., Holm L., Mistry J., Sonnhammer E.L., Tate J. and Punta M., *Nucleic Acids Res.*, 2014; **42 (Database issue)**: D222-30. DOI: 10.1093/nar/gkt1223

[17] Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D. and Bairoch A., *Nucleic Acids Res.*, 2003; **31**: 3784-8.

[18] Emanuelsson O., Brunak S., von Heijne G. and Nielsen H., *Nat. Protoc.*, 2007; **2(4)**: 953-71. DOI: 10.1038/nprot.2007.131

[19] Buchan D.W., Minneci F., Nugent T.C., Bryson K. and Jones D.T., *Nucleic Acids Res.*, 2013; **41 (Web Server issue)**: W349-57. DOI: 10.1093/nar/gkt381

[20] Kumar S., Stecher G., Li M., Knyaz C., and Tamura K.,. *Mol. Biol. Evol.,* 2018; **35(6)**: 1547-1549. doi: 10.1093/molbev/msy096

[21] Katoh K. and Standley D.M., *Mol. Biol. Evol.,* 2013; **30(4)**: 772-80. DOI: 10.1093/molbev/mst010

[22] Niyogi K.K., Li X.P., Rosenberg V. and Jung H.S., *J. Exp. Bot.*, 2005; **56 (411)**: 375-82. DOI: 10.1093/jxb/eri056

[23] Hasegawa K., Yukawa Y., Sugita M. and Sugiura M., *Gene*, 2002; **289(1-2)**: 161-8.

[24] Tanaka R., Rothbart M., Oka S., Takabayashi A., Takahashi K., Shibata M., Myouga F., Motohashi R., Shinozaki K., Grimm B. and Tanaka A., *Proc. Natl. Acad. Sci. USA.*, 2010; **107(38)**: 16721-5. DOI: 10.1073/pnas.1004699107

[25] Zhang M. and Yang X.E., *Russ. J. Plant Physiol.*, 2014; **61(4)**: 427-33.

[26] Andersson U., Heddad M. and Adamska I., *Plant Physiol.*, 2003; **132(2)**: 811-20. DOI: 10.1104/pp.102.019281

[27] Heddad M. and Adamska I., *Proc. Natl. Acad. Sci. USA.*, 2000; **97(7)**:3741-6.

[28] Heddad M., Noren H., Reiser V., Dunaeva M., Andersson B. and Adamska I., *Plant Physiol.*, 2006; **142(1)**: 75-87. DOI: 10.1104/pp.106.081489

[29] Adamska I. and Kloppstech K., *J. Biol. Chem.*, 1994; **269(48)**: 30221-6.

[30] Montane M.H. and Kloppstech K., *Gene*, 2000; **258(1-2)**: 1-8.

[31] NorÉN H., Svensson P., Stegmark R., Funk C., Adamska I. and Andersson B., *Plant, Cell Environ.*, 2003; **26(2)**: 245-53. https://doi.org/10.1046/j.1365-3040.2003.00954.x

[32] Binyamin L., Falah M., Portnoy V., Soudry E. and Gepstein S., *Planta*, 2001; **212(4)**: 591-7.

[33] Hubbart S., Smillie I.R.A., Heatley M., Swarup R., Foo C. C., Zhao L. and Murchie E.H.. 2018. *Commun. Biol.*, 2018; **1(1)**: 22. doi: 10.1038/s42003-018-0026-6.

[34] Głowacka K., Kromdijk J., Kucera K., Xie J., Cavanagh A.P., Leonelli L., Leakey A.D.B., Ort D. R., Niyogi K.K. and Long S.P. 2018. *Nat. Commun.*, 2018; **9(1)**: 868. doi: 10.1038/s41467-018-03231-x.