Structural and functional differentiation of the lightharvesting protein Lhcb4 during land plant diversification

Structural and functional differentiation of the light-harvesting protein Lhcb4 during land plant diversification

Pascal Albanese[a], Marcello Manfredi[b,c], Emilio Marengo[c], Guido Saracco[a], Cristina Pagliano[a,*]

[a]Applied Science and Technology Department–BioSolar Lab, Politecnico di Torino, Environment Park, Via Livorno 60, 10144 Torino, Italy

[b]ISALIT–Department of Science and Technological Innovation, University of Eastern Piedmont, Viale T. Michel 11, 15121 Alessandria, Italy

[c]Department of Science and Technological Innovation, University of Eastern Piedmont, Viale T. Michel 11, 15121 Alessandria, Italy

Correspondence

*corresponding author,

*e-mail: cristina.pagliano@polito.it

About 475 million years ago plants originated from an ancestral green alga and evolved first as non-vascular and later as vascular plants, becoming the primary producers of biomass on lands. During that time, the light harvesting complex II (LHCII), responsible for sunlight absorption and excitation energy transfer to the Photosystem II (PSII) core, underwent extensive differentiation. Lhcb4 is an ancestral LHCII that, in flowering plants, differentiated into up to three isoforms, Lhcb4.1, Lhcb4.2 and Lhcb4.3. The pivotal position of Lhcb4 in the PSII-LHCII supercomplex (PSII-LHCIIsc) allows functioning as linker for either S- or M-trimers of LHCII to the PSII core. The increased accumulation of Lhcb4.3 observed in PSII-LHCIIsc of plants acclimated to moderate and high light intensities induced us to investigate, whether this isoform has a preferential localization in a specific PSII-LHCIIsc conformation that might explain its light-dependent accumulation. In this work, by combining an improved method for separation of different forms of PSII-LHCIIsc from thylakoids of *Pisum sativum* L. grown at increasing irradiances with quantitative proteomics, we assessed that Lhcb4.3 is abundant in PSII-LHCIIsc of type C_2S_2, and, interestingly, similar results were found for

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/ppl.12964

the PsbR subunit. Phylogenetic comparative analysis on different taxa of the Viridiplantae lineage and structural modelling further pointed out to an effect of the evolution of different Lhcb4 isoforms on the light-dependent modulation of the PSII-LHCIIsc organization. This information provides new insight on the properties of the Lhcb4 and its isoforms and their role on the structure, function and regulation of PSII.

Abbreviations – ±-DDM, *n*-dodecyl-±-D-maltoside; Chl, chlorophyll; LHC, light harvesting complex; MS, mass spectrometry; RC, reaction centre; SWATH, sequential window acquisition of all theoretical fragment ion spectra

Introduction

Solar energy is the most abundant renewable energy source on Earth that oxygenic photosynthetic organisms are able to convert into chemical energy through the photosynthetic process. During this process, O_2 is released into the atmosphere and fixation of CO_2 into organic molecules occurs, ultimately supplying the oxygen that we breathe and the biomass sustaining essentially all life on our planet.

About 475 million years ago, plants originated from an ancestral green alga and evolved first as non-vascular and later as vascular plants, becoming the primary producers of biomass on lands. Plants succeeded in colonizing extremely different environments in terms of light quality and quantity, by adapting their photosynthetic apparatus. This has been made possible by coupling the highly conserved catalytic core of photosystems (PS) I and II with a differentiated antenna system, variable in number and type of proteins, to harvest sunlight and transport excitation energy towards the reaction centers (RCs). The transition of oxygenic photosynthesis from water (i.e. in cyanobacteria and green algae) to land (i.e. in plants), led to an important structural remodelling of the thylakoid membranes, which harbor the protein complexes responsible for the photosynthetic light reactions. The appearance and affirmation of the integral light harvesting complex LHCII, the most abundant membrane protein on Earth, favoured the differentiation of tightly appressed thylakoid membrane discs, called grana, interconnected by single membranes, the stroma lamellae, which are a peculiar feature of land plants. All the LHCIIs are membrane proteins with three transmembrane ±-helices, which bind a high number of pigments (i.e. chlorophylls (Chl) and carotenoids) as cofactors to harvest sunlight (Pan et al. 2013). In the grana, LHCIIs associate to the PSII dimeric core forming the PSII-LHCII supercomplex (PSII-LHCIIsc), a crucial component for solar energy conversion, as it catalyzes the photolysis of water

molecules initiating the photosynthetic electron transport. LHCIIs are nuclear encoded by a multigene family and diversified for function and organization in land plants, thus counterbalancing the evolutionarily high conservation of the catalytic PSII core, allowing its adaptation to varying environmental conditions. The most abundant LHCII occurs in trimers that can bind the PSII dimeric core (C_2) either strongly (S-trimer), if formed by Lhcb1-2 subunits, or moderately (M-trimer), if containing also the Lhcb3 subunit. The attachment of LHCII trimers to the PSII core relies on the presence of monomeric LHCIIs acting as linkers, with Lhcb5 and Lhcb6 serving exclusively for the S- and the M-trimer, respectively, and Lhcb4 for both. In higher plants, the main forms of PSII-LHCIIsc are the $C_2S_2M_2$, C_2S_2M, C_2S_2 and C_2S, with the first three representing the most complete forms isolated in paired conformation from stacked *Pisum sativum* (pea) thylakoids (Albanese et al. 2017, Su et al. 2017).

LHCII trimers and monomers seem to have followed evolutionarily different trajectories. The Lhcbs that form trimers are widely differentiated in several isoforms (five isoforms for Lhcb1, three for Lhcb2 and one for Lhcb3 are known in *Arabidopsis thaliana*) and share high sequence similarity, function and structural arrangement (Crepin and Caffarri 2018). Monomeric Lhcbs, contrarily, show several peculiar features. Lhcb5 is the only monomeric subunit capable of forming trimeric LHCII in plants depleted of Lhcb1-2-3 (Ruban et al. 2003), as it contains the trimerization motif WYXXXR (Hobe et al. 1995), and its expression seems not affected by light cues. Lhcb6 is one of the latest outcomes of evolution, being present only in land plants (Alboresi et al. 2008), where it is important to maintain the macro-organization of the grana membranes (Kovács et al. 2006), but eventually evolutionarily lost in some gymnosperms in concomitance with Lhcb3 (KouYl et al. 2016). Its expression is affected by light intensity, being Lhcb6 less abundant in plants grown under high irradiances (Albanese et al. 2016a, Albanese et al. 2018). Lhcb4 is ubiquitous in all classes of green plants (Koziol et al. 2007) and its presence is mandatory for maintaining the thylakoid macrostructure and the proper assembly of PSII-LHCIIsc (de Bianchi et al. 2011). In *A. thaliana* Lhcb4 is present in three different isoforms, Lhcb4.1, Lhcb4.2 and Lhcb4.3 (Jansson 1999). Differently from Lhcb4.1 and Lhcb4.2, Lhcb4.3 lacks a relatively large part of the C-terminal domain, thus was renamed as a distinct Lhcb (Lhcb8, Klimmek et al. 2006), and is accumulated within PSII-LHCIIsc of plants grown at high irradiances (Albanese et al. 2016a).

Despite the recent considerable efforts to solve the near-atomic structures of PSII-LHCIIsc of type C_2S_2 (Wei et al. 2016) and $C_2S_2M_2$ (Su et al. 2017, van Bezouwen et al. 2017), and understand the PSII-LHCIIsc reorganization during light acclimation (Albanese et al. 2016a, Bielczynski et al. 2016),

limited information is available on the possible assignment of some Lhcb isoforms to specific forms of PSII-LHCIIsc, as well as their positioning in the corresponding high resolution structures. Recently, we isolated and determined at 14 Å resolution of the three-dimensional (3D) structure of paired C_2S_2M supercomplexes from pea plants, interacting on their stromal sides through a specific overlap between apposing LHCII trimers and via physical connections that span the stromal gap. One of these stromal connections, called "knot", is likely formed by interactions between the N-terminal loops of two Lhcb4 subunits (corresponding to the Lhcb4.1/4.2 isoforms in *A. thaliana*) the other, termed "hinge", was tentatively assigned to the extrinsic PsbR subunit (Albanese et al. 2017). These supercomplexes were isolated from stacked thylakoids of plants grown at moderate light intensity, where the C_2S_2M is the predominant form of PSII-LHCIIsc (Albanese et al. 2016a). Long-term acclimation to high irradiances in this plant led to a substantial reorganization of the PSII-LHCIIsc due to a remarkable increment of the Lhcb4.3 isoform paralleled by a reduction of the Lhcb3 and Lhcb6 subunits, which favoured the major accumulation of supercomplexes of type C_2S_2 (Albanese et al. 2016a). Hence, elucidating the structural and functional role(s) of specific Lhcb4 isoforms in the PSII-LHCIIsc organization is determinant to understand the regulation of light absorption capacity in plants acclimated to increasing irradiances. Our preparations of paired PSII-LHCIIsc, isolated stacked thylakoid membranes solubilized with *n*-dodecyl-±-D-maltoside (±-DDM), retain the PsbR subunit (Barera et al. 2012, Pagliano et al. 2014, Albanese et al. 2017), conversely to classical unpaired PSII-LHCIIsc where this subunit is absent (Wei et al. 2016, Su et al. 2017, van Bezouwen et al. 2017). Therefore, these preparations are suitable starting material to also investigate possible light-dependent accumulations of PsbR within specific forms of PSII-LHCIIsc. In this work, by combining an improved method for isolation and separation of the different forms of paired PSII-LHCIIsc from solubilised thylakoid membranes with quantitative mass spectrometry (MS)-based proteomics, we assessed the protein composition of the different types of PSII-LHCIIsc present in pea plants grown at different irradiances. Phylogenetic comparative analysis on different taxa of the Viridiplantae lineage and structural modelling further pointed out to an effect of the evolution of different Lhcb4 isoforms on the light-dependent modulation of the PSII-LHCIIsc organization. This information provides new insight on the properties of the Lhcb4 and its isoforms and their role on the structure, function and regulation of PSII.

Materials and methods

Plant growth conditions and isolation of PSII-LHCII supercomplexes

Pisum sativum L. (pea; Bavicchi, Italy) plants were grown for three weeks inside a growth chamber (Sanyo MLR-351H) at 20°C, 60% humidity and a 8 h light/16 h dark photoperiod with white light at three different intensities, 30 (low, L), 150 (moderate, M) and 750 (high, H) µmol m^{-2} s^{-1} photons. The L and M conditions were provided with 3 and 15 fluorescent lamps FL40SS W/37 (Sanyo), of 40 W each, respectively; the H light condition was supplied by four LEDs LXR7-SW50 (Agonlight), of 35 W each, mounted inside the growth chamber. Both types of light sources have similar spectral power distribution curves (Albanese et al. 2018).

Stacked thylakoid membranes were isolated from three-week-old leaves at the end of the daily dark phase in the presence of divalent cation concentrations (i.e. 5 mM Mg^{2+}) that resemble the native chloroplast ionic environment (Schröppel-Meier and Kaiser 1988), preserving the stacked morphology of the grana membranes, following a procedure described in Albanese et al. (2017). Briefly, pea leaves were disrupted by grinding with a blender in 50 mM HEPES pH 7.5, 300 mM sucrose and 5 mM $MgCl_2$. The suspension was filtered through four cotton cloth layers, and the filtrate was centrifuged at 1500 *g* for 10 min. The pellet was washed once by centrifugation in the same buffer and then homogenized in 5 mM $MgCl_2$ and diluted 1:1 with 50 mM MES pH 6.0, 400 mM sucrose, 15 mM NaCl and 5 mM $MgCl_2$, followed by 10 min centrifugation at 3000 *g*. The resulting pellet of stacked thylakoid membranes was washed once by centrifugation in 25 mM MES pH 6.0, 10 mM NaCl and 5 mM $MgCl_2$, and then suspended and stored in 25 mM MES pH 6.0, 10 mM NaCl, 5 mM $MgCl_2$ and 2 M glycine betaine. All the steps of isolation were performed at 4°C in dim green light. When necessary, stacked thylakoids were flash frozen in liquid nitrogen and stored at -80°C.

Subsequently, thylakoid membranes were solubilized with ±-DDM (Anatrace) as in Barera et al. (2012) and separated by sucrose gradient centrifugation. Solubilized membranes (350 µl) were loaded onto a linear gradient containing 0.65 M sucrose, 25 mM MES pH 5.7, 10 mM NaCl, 5 mM $MgCl_2$ and 0.03% (w/v) ±-DDM (SG buffer) and centrifugation was carried out at 100 000 *g* for 12 h at 4°C (TH-641 rotor, Thermo Scientific). All visible pigment-containing bands of the sucrose gradient were carefully harvested with a syringe and concentrated by membrane filtration via Amicon Ultra 10- or 100-kDa cut-off devices (Millipore), and then flash-frozen and stored at -80°C.

Spectroscopic analyses

Absorption spectra of all sucrose gradient bands were recorded in native conditions at 12°C using a Lambda25 spectrophotometer (Perkin Elmer). The Chl (*a* and *b*) concentration of each band was

determined spectrophotometrically after extraction in 80% (v/v) acetone (Porra et al. 1989), and subsequently the corresponding Chl *a*/*b* ratio was deduced.

Gel electrophoresis and western blotting

Large pore blue native polyacrylamide gel electrophoresis (lpBN-PAGE) for the optimal separation of thylakoid protein complexes and PSII-LHCIIsc was performed according to Järvi et al. (2011) with minor modifications described in Albanese et al. (2016b). Mono-dimensional SDS-PAGEs were performed according to the Laemmli's system (Laemmli 1970) on a 12.5% (w/v) polyacrylamide gel containing 5 M urea. For the quantification of PsbR and D2, 1, 2 and 3 µg of Chl of three independent samples were loaded on mono-dimensional SDS-PAGE. The separated proteins were transferred onto nitrocellulose membrane and immuno-detected with specific antisera against D2 (home-made polyclonal antibody, for details see Barbato et al. 1992) and PsbR (Agrisera code, AS05 059) subunits, by using the alkaline phosphatase conjugate method, with 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium as chromogenic substrates (Sigma-Aldrich). To avoid any deviation between different immunoblots, every PSII-LHCIIsc containing bands of the sucrose gradient were loaded onto the same gel and antibody dilution was checked in the linearity range. Densitometry measurements of PsbR and D2 band signals were performed with the GelDoc XR^+ imager (Bio-Rad), by using Quantity One software version 4.6.9 (Bio-Rad). Quantitative comparison was performed on the basis of the slope value of the line interpolating the three incremental values of intensity for each condition tested (Albanese et al. 2016a).

Protein digestion, mass spectrometry analysis and data processing

For each light condition tested, 15 µg of Chl of each sucrose gradient band containing PSII-LHCIIsc were diluted in SG buffer to a final Chl concentration of 125 µg ml^{-1}. Proteins were precipitated overnight at -20°C in four volumes of cold acetone to remove the adhered pigments. After centrifuging at 20 000 *g* for 20 min at 4°C, proteins were resuspended in a buffer made of 50 mM Tris-HCl pH 8.0, 8 M urea and 2 M thiourea (denaturing buffer). Insoluble material was removed by centrifuging at 15 000 *g* for 10 min. Protein concentration in the supernatant was determined using the Bradford assay with bovine serum albumin as standard (Bradford 1976). Denatured proteins (10 µg for each replicate) from the same biological condition were pooled together. Proteins at 0.5 mg ml^{-1} in the denaturing buffer were reduced with 10 mM DTT (at 37°C, 30 min) and alkylated with 20 mM iodoacetamide (at 22°C, 30 min in the dark). To preserve trypsin activity, the urea concentration was diluted to 1 M by

adding 50 mM Tris-HCl pH 8.0. The protein in-solution digestion was conducted by adding Trypsin (Sigma-Aldrich, code T6567) to a final protein:protease ratio of 25:1 (w/w), followed by overnight incubation at 37°C. Trifluoroacetic acid was added to a final concentration of 0.5% (v/v) to terminate the tryptic digestion process. Insoluble material was removed by centrifuging at 15 000 *g* for 10 min. Subsequently, peptides desalting was conducted by solid phase extraction as in Villén and Gygi (2008) using 30 mg HyperSep C18 cartridges (Thermo Fisher Scientific). The resulting elutes were mixed with approximately 1500 femtomoles of a synthetic heavy peptide used as internal standard (Cellmano Biotech) and lyophilized. Dried peptides were dissolved in 30 µl of LC-MS/MS mobile phase A (water containing 0.1% (v/v) formic acid).

LC-MS/MS analyses, for either protein identification or quantification by SWATH (Sequential Window Acquisition of all THeoretical fragment ion spectra), were performed as described in Albanese et al. (2016a). Mass spectral data sets were analyzed and searched using the database search engine ProteinPilot™ v.5.0.1.0, 4895, using the Paragon algorithm 5.0.1.0, 4874 (AB Sciex). The following sample parameters were used: trypsin digestion, cysteine alkylation set to carbamidomethylation and no special factors. Processing parameters were set to "Biological modification". In total twelve Data Dependent Acquisition (DDA) MS raw files, two for each analysed sucrose gradient band, were searched for protein and transcript identification, thorough ID search effort, using a UniProtKB/TrEMBL database containing Viridiplantae proteins (version 2018.04.20, with a total of 5677289 sequences) merged with a de novo proteome containing 34740 unique protein sequences derived from the transcriptome of *P. sativum* (p.sativum_csfl_reftransV1 downloaded from https://www.coolseasonfoodlegume.org/organism/Pisum/sativum/reftrans/v1) using the online Galaxy platform (Afgan et al. 2016). The resulting protein database was concatenated with a reversed "decoy" version of the "forward" database. After searching, we accepted protein IDs that had a ProteinPilot Unused Score of at least 1.3 (equivalent to a 95% confidence interval) as a cutoff threshold and an estimated local false discovery rate not higher than 1% (Rardin et al. 2015).

Quantitative analysis based on Data Independent Acquisition (DIA) MS2 (i.e. fragment ion masses) chromatogram was carried out in Skyline (v4.1.0.11714), an open source software project (http://proteome.gs.washington.edu/software; MacLean et al. 2010). Spectral libraries were generated in Skyline from database searches of the raw data files (.group) performed with ProteinPilot. All raw files acquired in DIA were directly imported into Skyline and MS1 precursor ions and MS2 MS/MS fragment ions were extracted for all peptides present in the MS/MS spectral libraries. Quantitative analysis was based on extracted ion chromatograms (XICs) of four DIA runs for each biological

replicate, considering up to five MS/MS highest ranked fragment ions, typically y- and b-ions. Quantitative SWATH MS2 analysis was based on XICs matching to specific peptides present in the spectral libraries and unique to the proteins unambiguously identified. Proteins were considered eligible for quantification with a minimum of two tryptic peptides with a length between 6 and 25 amino acids. The 25 N-terminal residues of each peptide and peptides post-translationally modified or containing missed cleavages or cysteine and/or methionine residues were excluded. Quantitative differences of proteins and peptides between samples were statistically assessed using MSstats (v3.8.3), an open-source R-based package (Choi et al. 2014). Significant variation thresholds in protein amounts were defined according to Clough et al. (2012) and only fold changes e 1.35 or d 0.74 with adjusted p-value d 0.05 were further discussed.

Proteins were further annotated only if eligible for quantification, as described in Albanese et al. (2018). The MS/MS-derived peptide sequence data were used to retrieve the translated transcripts (p.sativum_csfl_reftransV1) for database searching combined with BLAST analysis if the sequences were not present in the Viridiplantae database. The highest ranked hit to a homologous protein with reviewed accession number, if available, was chosen for identification.

Phylogenetic analysis and structural model predictions

Phylogenetic analysis of Lhcb4 isoforms was performed using 20 sequences from distantly related photosynthetic organisms, representative of different taxa of the Viridiplantae lineage. Representatives of green algae (*Mesostigma viride*, *Micromonas commoda, Chlamydomonas reinhardtii*) and Charophytes (*Klebsormidium nitens*) for aquatic environments, in addition to several species for land plants, ranging from Embryophytes considered "basal land plants", among which two bryophytes (*Marchantia polymorpha, Physcomitrella patens*) and one lycophyta (*Selaginella moellendorffii*), to Tracheophytes including three gymnosperms (*Gingko biloba, Araucaria cunninghamii*, *Picea sitchensis*) and several angiosperms, either dicots (*Populus trichocarpa, Spinacia oleracea, Arabidopsis thaliana, Pisum sativum*) or monocots (*Oryza sativa subsp. Japonica, Zea mays*), were selected. As outgroup, all Lhcb5 sequences from the organisms mentioned above, retrieved from UniprotKB database, were used. The only exception regards Lhcb5 of *Gingko biloba*, whose sequence has been retrieved by tBLASTn search on PlantGDB (www.plantgbd.org).

Sequence alignment was performed in MUSCLE (Edgar 2004) using the default settings. The alignment was then curated with Gblocks, selecting the conserved positions and flanking regions for further analysis (33% of the total; Castresana 2000). The phylogenetic tree was built with the

Maximum Likelihood method and the Poisson correction model in PhyML (Guindon et al. 2010). Bootstrap (BS) values resulted from 100 replicates, branches were considered significant with BS > 0.7.

Predicted Lhcb4 structural models were performed with PHYRE2 (Kelley et al. 2015) using Lhcb4 sequences from *C. reinhardtii* (Lhcb4, Q93WD2), *P. sitchensis* (Lhcb4, A9NYR8) and *P. sativum* (Lhcb4.3, p.sativum_csfl_reftransV1_0068262).

Results

Improved fractionation of solubilized thylakoid membranes allowed to obtain two distinct preparations of paired PSII-LHCII supercomplexes of different size, depending on the plant growth irradiance

To improve the homogeneity of the preparation of paired PSII-LHCIIsc, we optimized the conditions for fractionation of the stacked thylakoid membranes solubilized with ±-DDM by modifying our previous protocol (Barera et al. 2012, Albanese et al. 2016a, Albanese et al. 2017) with the introduction of a separation step performed on a smaller ultracentrifuge tube (TH-641 instead of Surespin 630 rotor, Thermo Scientific) which improved the linearity of the sucrose gradient. Using this procedure, we were able to obtain up to nine pigment-binding bands (numbered 1-9, according to their position on the sucrose gradient) and separate two distinct bands in the region containing paired PSII-LHCIIsc for each light condition (Fig. 1A).

The main visible bands (Fig. 1A) were further characterized by absorption spectroscopy (Fig. S1) and SDS-PAGE (Fig. 1B-D). By combining these two analyses, it was evident that starting from stacked thylakoids, this separation system allows a good separation of the paired PSII-LHCIIsc in the lower part of the tube (bands 6-8) at the expense of an overall low-level separation of the protein complexes in the upper part of the tube (bands 3-5) especially due to overlap of ATP synthase (ATPase)/PSII/PSI, more evident in some light conditions. For this reason, for each green band of the sucrose gradient only the main pigment-binding protein complexes were indicated in Fig. 1A, leaving to the description below the indication of the presence of ATPase or other pigment-binding protein complexes occurring in minor amounts or arising from cross contamination of neighboring bands. Band 1 contained either monomeric LHCII or free proteins and pigments and its amount increased as growth light intensity raised (H1>M1>L1), the contrary was found for band 2 (H2<M2<L2) which contains trimeric LHCII. Band 3 contained mostly monomeric PSII, despite a partial contamination likely occurred during its

harvesting by PSI-LHCI and ATPase present as main components of band 4, more evident in M and H lights, where band 4 is more abundant. Band 5 contained both PSI-LHCI and PSII-LHCII subunits, in addition to ATPase contaminants, and was remarkably more abundant at higher growth irradiances (H5>M5>L5). In this band, the presence of PSI-LHCI particles (indicated with * in Fig. 1A) with higher molecular mass than those of band 4, might suggest the occurrence of either PSI-LHCI-LHCII particles or oligomeric PSI-LHCI complexes (recently found in stacked thylakoid membranes of spinach plants in Wood et al. 2018). The simultaneous higher amount of PSII core subunits and, to a lower extent, of LHCII antenna proteins in band 5 suggests the presence of dimeric PSII particles retaining some LHCII antennae (i.e. likely PSII-LHCIIsc of type C_2S, indicated with * in Fig. 1A). The increased resolving power of the sucrose gradient adopted allowed the separation of two bands for each light condition, localized from position 6 to 8 (i.e. H6 and H7, M7 and M8, L7 and L8), containing paired PSII-LHCIIsc with increasing molecular weight. The native spectra of these bands (Fig. 2A) showed, at increasing molecular weights, a higher content of Chl *b* bound to Lhcb (peaking around 470 and 650 nm) paralleled by decreased values of the corresponding Chl *a/b* ratios (Fig. 2B). This evidence attests the separation of two distinct fractions of PSII-LHCIIsc in each light condition, with antenna size significantly different in M and H, but not in L light. Finally, band 9, visible only in L and M lights, displayed an absorption spectrum identical to that of band 8 (for this reason not shown in Fig. S1) as well as a similar protein profile (Fig. 1B-D), thus it contained PSII-LHCII megacomplexes (i.e. corresponding to those of band ±4 in Albanese et al. 2016a).

Lhcb4.3 and PsbR are abundant in paired PSII-LHCII supercomplexes of type C_2S_2

In order to determine the organization and protein composition of the main PSII-LHCIIsc present in the two distinct bands of the sucrose gradient visible in each light condition (Fig. 1A), sucrose gradient bands L7 and L8, M7 and M8, H6 and H7 were subjected to either native electrophoresis (Fig. 2C) or to liquid-trypsin digestion followed by MS-based proteomic quantification (Fig. 2D, Table S1, Table S2). The MS proteomic profiling of these preparations was greatly improved by exploiting transcriptomic data available for *P. sativum* to retrieve the protein sequences, as described in Albanese et al. (2018). Relative quantification of proteins was performed by pairwise comparison within the same light condition in order to assess the role(s) of specific protein subunits in determining the overall PSII-LHCIIsc organization in plants grown at a certain light regime.

The $C_2S_2M_2$ and C_2S_2M resulted as the dominant PSII-LHCIIsc in L plants, showing only a slight variation in their ratio both in bands L7 and L8 (Fig. 2C). Indeed, when L7 was compared to L8, no

significant variation in the amount of the antennae subunits related to M-trimers bound to the PSII core (i.e. Lhcb3 and Lhcb6) was detected, noteworthy the quantity of Lhcb4.3 and PsbR was roughly doubled (Fig. 2D, Table S2). The C_2S_2 apparently was the PSII-LHCIIsc that changed more in abundance between bands L7 and L8 (Fig. 2C), suggesting a relation between this conformation of supercomplex and the presence of Lhcb4.3 and PsbR. Similar results for Lhcb4.3 and, to a lesser extent, for PsbR were found by comparing PSII-LHCIIsc isolated from either M (M7 vs M8) or H (H6 vs H7) plants (Fig. 2D, Table S2). In M and H conditions, this relation was even more evident, since the increased amount of Lhcb4.3 and PsbR observed, correlates well with the remarkably reduced antenna cross-section displayed by the PSII-LHCIIsc present in the upper band with respect to those in the lower band. This is pointed out by (1), native electrophoresis (Fig. 2C) that showed that the C_2S_2 was clearly more abundant in bands M7 than M8 and H6 than H7 and (2), by MS-based quantitation (Fig. 2D, Table S2), that highlighted an halved quantity of Lhcb3 and Lhcb6, indicating a strong reduction of the M-trimers present in these preparations.

Due to the availability of the PsbR antibody, we performed an additional experiment of western blotting to further check the relationship between PsbR and the C_2S_2 conformation, testing for the quantity of PsbR and two PSII-LHCIIsc preparations, which represent the extremes for abundance of the C_2S_2 form in our experimental conditions: band H6 for the highest amount (i.e. occurring in H plants) and band L8 for the lowest amount (i.e. occurring in L plants). Western blot analysis, based on identical Chl loading (3 μg), revealed a higher amount of PsbR in band H6 with respect to L8 (Fig. 3A). To estimate the different abundance of this subunit in H6 and L8 bands, we measured the relative PsbR content (and that of the D2 protein of the PSII RC for reference) by immunoblot titration (example reported in Fig. S2A). To ascertain the non-saturation, signal linearity through the different dilutions was checked in both samples, as exemplified in Fig. S2B-C. The results of the quantification, shown in Fig. 3B, demonstrate that the amount of PsbR (either on a Chl base or normalized to the D2 protein content) is roughly 3-fold higher in band H6 with respect to L8, confirming that this subunit is more abundant in supercomplexes of type C_2S_2 with respect to forms with bigger antenna size (i.e. with M trimers bound), in accordance to mass spectrometry quantifications (Fig. 2D, Table S2).

Differentiation of the Lhcb4 isoforms at the C-terminus played a major role in land plants diversification

Lhcb4 and Lhcb5 are ancestral LHCs that developed as PSII antenna systems prior to green algal diversification (Koziol et al. 2007, Neilson and Durnford 2010). Following the emergence of different

green algal groups, and during land plant diversification, Lhcb4 likely underwent major evolutionary changes in different lineages driven by gene duplication, similarly to most of the peripheral LHCIIs (Koziol et al. 2007), finally differentiating in flowering plants in up to three isoforms, Lhcb4.1, Lhcb4.2 and Lhcb4.3, as known in *A. thaliana*. By using sequence analysis, we investigated the diversification of the Lhcb4 sequence over the green lineage, from green algae to flowering plants, in relation with its possible role(s) in determining specific PSII-LHCIIsc organizations. The sequences of the Lhcb4.1/4.2/4.3 isoforms from *A. thaliana* and those distinguishable from *P. trichocarpa* and *P. sativum* were compared to the Lhcb4 sequences of other organisms from different taxa of the Viridiplantae lineage, using the Lhcb5 sequences as outgroup (Fig. 4A, Fig. S3A). The organisms have been selected to cover a broad range of taxa and environmental conditions and are listed in Table S3. Noteworthy, in our search for Lhcb4.3 sequences to perform the phylogenetic analysis, we did not find any representative belonging to monocots. This evidence seems to indicate that the *LHCB4.3* gene is confined to dicotyledonous plants (Klimmek et al. 2006). The complete alignment of the Lhcb4 sequences (Fig. S3A) in flowering plants evidenced a high sequence similarity of the Lhcb4.1/4.2 isoforms (i.e. 87.2% identity in *A. thaliana*), responsible for their relatively low diversification in terms of tree configuration (Fig. 4A), and a higher variability at the C-terminus, which is truncated in the Lhcb4.3 isoform (Fig. 4A, Fig. S3A). Noteworthy, the Lhcb4.3 sequences clustered with the Lhcb4 sequences of the gymnosperm clade, forming a well-supported "Lhcb4.3-like" cluster (Fig. 4A). The sequences in this cluster shared a truncated conformation, lacking roughly the last 13-15 amino acids if compared to the Lhcb4.1/4.2 isoforms from angiosperms and bryophytes (Fig. 4A). In particular, the C-terminus of Lhcb4.1/4.2 isoforms contains a histidine (i.e. His230 on PDB 5XLN: chain R) involved in the coordination of a Chl *b* molecule (*b*614, according to the nomenclature of Pan et al. 2013), which stands at the binding interface with the Lhcb6 (PDB 5XLN: chain 4; Fig. 4B). The Chl *b*614, despite apparently quite excitonically isolated from the other Chls in the protein (Jassas et al. 2018), represents the crossway in the energy transfer pathways within the $C_2S_2M_2$ supercomplex, being the closest pigment to either the Chl *b*606 of Lhcb6 or the Chl *a*614 of Lhcb3 (Su et al. 2017). This histidine residue is well conserved in Lhcb4 sequences of angiosperms and bryophytes that clustered with the Lhcb4.1/4.2 isoforms of Arabidopsis (Fig. 4A), forming a "Lhcb4.1/4.2-like" cluster. All these plants contain also the *LHCB6* coding gene and/or have evidences of its expression at a protein level. Interestingly, this histidine is also present in the Charophyta *K. nitens* (Fig. 4A), which lives in aquatic environments but shows some features required for land adaptation (Hori et al. 2014).

Up to now, high-resolution structures are available only for an undistinguishable Lhcb4.1/4.2 isoform isolated either as single protein (Pan et al. 2011) or within PSII-LHCIIsc (Wei et al. 2016, Su et al. 2017, van Bezouwen et al. 2017) from *Spinacia oleracea* (spinach), pea and Arabidopsis plants. Here, by using structure prediction tools and template matching, we showed the predicted C-terminus conformation for the Lhcb4.3 isoform of *P. sativum* (Fig. 4D). Based on this prediction and taking into account the localization of the Lhcb4 subunit within the PSII-LHCIIsc, we propose a specific organization of type C_2S_2 for supercomplexes containing the Lhcb4.3 isoform (Fig. 4D). The structural superimposition of the Lhcb4.1/4.2 of *P. sativum* (PDB 5XNL: chain R) with the predicted structures of either Lhcb4.3 of *P. sativum* or Lhcb4 of *P. sitchensis* for gymnosperms has revealed common structural features (Fig. 4D-E) that reflect their grouping as members of the same "Lhcb4.3-like" cluster (Fig. 4A). Among those features, the truncated C-terminus, putative interface of interaction with Lhcb6, and the lack of the His230 residue, coordinating the Chl *b*614 molecule, might therefore be determinant for a lower binding affinity for and/or a different interface of interaction of the M-trimer in the PSII-LHCIIsc. Noteworthy, the PSII-LHCIIsc structure in the Norway spruce *P. abies*, whose Lhcb4 sequence is identical to that from *P. sitchensis* used for the phylogenetic analysis shown in Fig. 4A, shows a rotational offset of ~ 52° of the M-trimer bound to the PSII core (Fig. 4E, Kouřil et al. 2016), with respect to that one present in the $C_2S_2M_2$ of angiosperms (Fig. 4C, Su et al. 2017). In this gymnosperm, the absence of the Lhcb3 and Lhcb6 proteins is responsible for the different PSII-LHCIIsc organization, which in part resembles its counterpart in the alga *C. reinhardtii*, for the rotational offset of the M-trimer (Fig. 4F, Drop et al. 2014). In *C. reinhardtii*, which lacks the *LHCB3* and *LHCB6* genes, the PSII-LHCIIsc contains an additional N-trimer associated with the PSII core in the position occupied by Lhcb6 in plants (Drop et al. 2014), whose interaction likely depends on specific features of the C-terminus of the Lhcb4 therein (Fig. 4A).

Discussion

In order to colonize lands in any ecological niche, plants have evolved a wide range of strategies to harvest efficiently the sunlight by differentiating their antenna systems to dynamically modulate the PSII-LHCIIsc organization. When considering the composition of an early PSII antenna system, the Lhcb4 subunit appeared ancestral (Koziol et al. 2007, Neilson and Durnford 2010), and its presence in most Chl *a/b* containing organisms suggests its structural role in the PSII antenna system. In some flowering plants, as the model organism *A. thaliana*, this subunit differentiated in up to three isoforms,

Lhcb4.1, Lhcb4.2 and Lhcb4.3. Contrarily to *LHCB4.1* and *LHCB4.2*, the *LHCB4.3* gene is strongly up-regulated in moderate and high light, as attested at level of transcription (Alboresi et al. 2011), translation (Floris et al. 2013) and protein expression (Albanese et al. 2016a, Miller et al. 2017, Albanese et al. 2018). The increased accumulation of the Lhcb4.3 subunit observed in PSII-LHCIIsc of plants acclimated to increasing intensity of light (Albanese et al. 2016a), induced us to investigate whether this isoform has a preferential localization in a specific PSII-LHCIIsc conformation that might explain a light-regulated function for this protein.

By optimizing our previous protocol for separation of solubilized thylakoid membranes (Barera et al. 2012, Albanese et al. 2016a, Albanese et al. 2017), we were able to obtain two distinct fractions enriched in paired PSII-LHCIIsc with different antenna cross-section for each light condition (Fig. 1A). Irrespective of the light intensity, from proteomic analyses (Fig. 2D, Table S2), we found higher amounts of Lhcb4.3 in the fraction enriched with paired PSII-LHCIIsc with smaller antenna, likely attributable to the C_2S_2 form (Fig. 2C). Lhcb4.3, with respect to the other two isoforms, lacks ~ 10 amino acids at the C-terminus domain (Fig. 4A), which are the site for binding the Lhcb6 and consequently, the M-trimer to the PSII-LHCIIsc (Su et al. 2017). This suggests that the presence of the Lhcb4.3 within the C_2S_2 might have a determinant structural function to avoid the binding of additional M-trimers to the PSII core at high irradiances, by significantly modifying the interaction interface for its specific linker, the Lhcb6 subunit. Interestingly, Lhcb4 has been proposed as a major interactor of PsbS (Sacharz et al. 2017), a special "Lhc-like protein" appeared upon land colonization that plays a crucial role in plant adaptation by harmlessly quenching excessive irradiation as heat in a process called non-photochemical quenching (NPQ; Li et al. 2000). Noteworthy, the pattern of expression of the *LHSB4.3* gene was found more similar to that of *PSBS,* with respect to that of other *LHC* genes (Klimmek et al. 2006), suggesting a cooperative function of the corresponding proteins under conditions where elevated capacity for non-radiative energy dissipation is of physiological relevance. Since the amount of both, Lhcb4.3 and PsbS, significantly increases in thylakoids of plants exposed to high irradiances (Albanese et al. 2018), we hypothesize that PsbS might interact with Lhcb4.3 to regulate the PSII-LHCIIsc organization in high light, ultimately controlling their macro-organisation in the grana membranes (Kereïche et al. 2010).

Irrespective of the light intensity, and similarly to what found for Lhcb4.3, proteomic analyses revealed a higher amount of PsbR in the sucrose gradient band containing the fraction enriched with paired PSII-LHCIIsc with smaller antenna, likely attributable to the C_2S_2 form (Fig. 2D, Table S2). PsbR is an extrinsic 10 kDa subunit present in higher plants and green algae, whose localization is still

unknown, since it is not present in the high-resolution structures of PSII-LHCIIsc available for plants (Wei et al. 2016, Su et al. 2017, van Bezouwen et al. 2017). Due to its capacity to crosslink with PsbP, a luminal position for PsbR was hypothesized (Ido et al. 2014). In vascular plants PsbR, together with PsbP and PsbQ, contributes to the stability of PSII-LHCIIsc, and its depletion causes a lower quantum yield of PSII (Allahverdiyeva et al. 2013), due to a slowing of electron transfer from Q_A to Q_B, as a consequence of a modification proposed to occur either at both the acceptor and the donor side (i.e. lumenal, Allahverdiyeva et al. 2007) or principally at the acceptor side of PSII (i.e. stromal, Stockhaus et al. 1990, Liu et al. 2009). In the 3D structure of the paired C_2S_2M supercomplexes isolated from pea plants grown in moderate light, we tentatively ascribed to PsbR the stromal "hinge" density connecting two facing PSII dimeric cores on the S-trimer side of the supercomplex, which provides physical connections that can link adjacent thylakoids, ultimately mediating the stacking of grana membranes (Albanese et al. 2017). Since here we found that PsbR is principally bound to paired C_2S_2 supercomplexes (Fig. 3B), which are the most abundant PSII-LHCIIsc of plants acclimated to high light (Albanese et al. 2016a), and considering the proposed stromal localization for PsbR (Albanese et al. 2017, despite further experiments are necessary to locate it with confidence), we hypothesize for this protein a functional/structural role(s) in mediating the interaction between adjacent thylakoid membranes, preponderant in plants grown in high light. Conversely, the higher abundance of M-trimers bound to the paired PSII-LHCIIsc in plants grown in low light (Albanese et al. 2016a) might already fulfill this task through the so called "Velcro effect" (Jia et al. 2012), thus counterbalancing the reduced amount of PsbR (Fig. 3B). Since the amount of either PsbR or PSII core proteins does not significantly change in the stacked thylakoid membranes of pea plants grown in low, moderate and high lights (Albanese et al. 2018) used as starting material for the isolation of the paired supercomplexes of this work, our findings suggest that in thylakoid membranes of high light plants, a higher proportion of PsbR is bound to PSII-LHCIIsc, mostly in the paired C_2S_2 form. However, in low light, plants might have a higher proportion of PsbR present in PSII complexes different from fully assembled paired PSII-LHCIIsc of type $C_2S_2M_2$.

Sequence analysis revealed that the Lhcb4.3 sequences clustered with the Lhcb4 sequences of the gymnosperm clade, forming a well-supported "Lhcb4.3-like" cluster (Fig. 4A, Fig. S3B), are composed of sequences sharing a truncated conformation at the C-terminus. The differentiation of a truncated C-terminus of the Lhcb4.3 might have been beneficial for angiosperms to thrive in high irradiances, limiting the income of solar energy in excess by reducing the M-trimers bound to the PSII core. Indeed, this structural trait is absent in bryophytes and lycophytes, considered "basal" to the

Tracheophytes lineage, which are largely segregated in environments with limited light availability. The Lhcb4 of gymnosperms (including some Pinaceae lacking the *LHCB3* and *LHCB6* genes, Kouřil et al. 2016), a clade appeared just after the greatest extinction event that probably exposed plants to harsh illumination conditions by opening the canopy (McElwain and Punyasena, 2007), has a "Lhcb4.3-like" C-terminus. This truncated C-terminus might have been advantageous to some gymnosperms experiencing high light, by providing an alternative modulation of the energy transfer from the LHCII antenna system to the PSII core. In conclusion, this work highlights the evolutionary role of the Lhcb4 during land plant colonization, suggesting that the differentiation of C-truncated Lhcb4 forms introduced structural changes in the organization of the antenna systems of the PSII-LHCIIsc, thus representing an additional strategy at the level of the Lhcb subunits, adopted by different plants to adapt to changing light conditions and environments.

Author contributions

C.P. and P.A. designed the experiments. P.A. prepared samples for MS analysis, M.M. collected MS data, P.A. analysed MS data. P.A. performed phylogenetic analyses. C.P. supervised the experiments. C.P. and P.A. wrote the manuscript. G.S. and E.M. contributed analysis tools. All authors approved the final version of the manuscript.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Vizcaíno et al. 2016) partner repository with the dataset identifier PXD011213.

Acknowledgements

This work was supported by the Italian Ministry of Education, University and Research, "Futuro in Ricerca 2013" program RBFR1334SB to C.P. The authors kindly thank Prof Roberto Barbato (University of Eastern Piedmont, Italy) for supplying the antibody against D2.

References

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, ech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J (2016) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res

44: W3–W10

Albanese P, Manfredi M, Meneghesso A, Marengo E, Saracco G, Barber J, Morosinotto T, Pagliano C (2016a) Dynamic reorganization of photosystem II supercomplexes in response to variations in light intensities. Biochim Biophys Acta 1857: 1651–1660

Albanese P, Nield J, Tabares JA, Chiodoni A, Manfredi M, Gosetti F, Marengo E, Saracco G, Barber J, Pagliano C (2016b) Isolation of novel PSII-LHCII megacomplexes from pea plants characterized by a combination of proteomics and electron microscopy. Photosynth Res 130: 19–31

Albanese P, Melero R, Engel BD, Grinzato A, Berto P, Manfredi M, Chiodoni A, Vargas J, Sorzano CÓS, Marengo E, Saracco G, Zanotti G, Carazo JM, Pagliano C (2017) Pea PSII-LHCII supercomplexes form pairs by making connections across the stromal gap. Sci Rep 7: 10067

Albanese P, Manfredi M, Re A, Marengo E, Saracco G, Pagliano C (2018) Thylakoid proteome modulation in pea plants grown at different irradiances: quantitative proteomic profiling in a non-model organism aided by transcriptomic data integration. Plant J 96: 786–800

Alboresi A, Caffarri S, Nogue F, Bassi R, Morosinotto T (2008) In silico and biochemical analysis of *Physcomitrella patens* photosynthetic antenna: identification of subunits which evolved upon land adaptation. PLoS ONE 3: e2033

Alboresi A, Dall'Osto L, Aprile A, Carillo P, Roncaglia E, Cattivelli L, Bassi R (2011) Reactive oxygen species and transcript analysis upon excess light treatment in wild-type *Arabidopsis thaliana* vs a photosensitive mutant lacking zeaxanthin and lutein. BMC Plant Biol 11: 62

Allahverdiyeva Y, Mamedov F, Suorsa M, Styring S, Vass I, Aro EM (2007) Insights into the function of PsbR protein in *Arabidopsis thaliana*. Biochim Biophys Acta, 1767: 677–685

Allahverdiyeva Y, Suorsa M, Rossi F, Pavesi A, Kater MM, Antonacci A, Tadini L, Pribil M, Schneider A, Wanner G, Leister D, Aro EM, Barbato R, Pesaresi P (2013) Arabidopsis plants lacking PsbQ and PsbR subunits of the oxygen-evolving complex show altered PSII super-complex organization and short-term adaptive mechanisms. Plant J 75: 671–684

Barbato R, Friso G, de Laureto PP, Frizzo A, Rigoni F, Giacometti GM (1992) Light-induced degradation of D2 protein in isolated photosystem II reaction center complex. FEBS Lett 311: 33–36

Barera S, Pagliano C, Pape T, Saracco G, Barber J (2012) Characterization of PSII-LHCII supercomplexes isolated from pea thylakoid membrane by one-step treatment with ±- and ²-dodecyl-D-maltoside. Philos Trans R Soc Lond B Biol Sci 367: 3389–3399

Bielczynski LW, Schansker G, Croce R (2016) Effect of light acclimation on the organization of Photosystem II super- and sub-complexes in *Arabidopsis thaliana.* Front Plant Sci 7: 105

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72: 248–54

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552

Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, Vitek O (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics 30: 2524–2526.

Clough T, Thaminy S, Ragg S, Aebersold R, Vitek O (2012) Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. BMC Bioinformatics 13 Suppl 1: S6

Crepin A, Caffarri S (2018) Functions and evolution of Lhcb isoforms composing LHCII, the major light harvesting complex of Photosystem II of green eukaryotic organisms. Curr Protein Pept Sci 19: 1–15

de Bianchi S, Betterle N, KouYl R, Cazzaniga S, Boekema E, Bassi R, Dall'Osto L (2011) Arabidopsis mutants deleted in the light-harvesting protein Lhcb4 have a disrupted Photosystem II macrostructure and are defective in photoprotection. Plant Cell 23: 2659–2679

Drop B, Webber-Birungi M, Yadav SKN, Filipowicz-Szymanska A, Fusetti F, Boekema EJ, Croce R (2014) Light-harvesting complex II (LHCII) and its supramolecular organization in *Chlamydomonas reinhardtii*. Biochim Biophys Acta 1837: 63–72

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797

Floris M, Bassi R, Robaglia C, Alboresi A, Lanet E (2013) Post-transcriptional control of light-harvesting genes expression under light stress. Plant Mol Biol 82: 147–154

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New Algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307–321

Hobe S, Forster R, Klingler J, Paulsen H (1995) N-proximal sequence motif in light-harvesting chlorophyll a/b-binding protein is essential for the trimerization of light-harvesting chlorophyll *a/b* complex. Biochemistry 34: 10224–10228

Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H,

Tajima N, Moriyama T, Ikeuchi M, Watanabe M, Wada H, Kobayashi K, Saito M, Masuda T, Sasaki-Sekimoto Y, Mashiguchi K, Awai K, Shimojima M, Masuda S, Iwai M, Nobusawa T, Narise T, Kondo S, Saito H, Sato R, Murakawa M, Ihara Y, Oshima-Yamada Y, Ohtaka K, Satoh M, Sonobe K, Ishii M, Ohtani R, Kanamori-Sato M, Honoki R, Miyazaki D, Mochizuki H, Umetsu J, Higashi K, Shibata D, Kamiya Y, Sato N, Nakamura Y, Tabata S, Ida S, Kurokawa K, Ohta H (2014) *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. Nat Commun 5: 3978

Ido K, Nield J, Fukao Y, Nishimura T, Sato F, Ifuku K (2014) Cross-linking evidence for multiple interactions of the PsbP and PsbQ proteins in a higher plant photosystem II supercomplex. J Biol Chem 289: 20150–20157

Jansson S (1999) A guide to the Lhc genes and their relatives in Arabidopsis. Trends Plant Sci 4: 236–240

Järvi S, Suorsa M, Paakkarinen V, Aro EM (2011) Optimized native gel systems for separation of thylakoid protein complexes: novel super- and mega-complexes. Biochem J 439: 207–214

Jassas M, Chen J, Khmelnitskiy A, Casazza AP, Santabarbara S, Jankowiak R (2018) Structure-based exciton hamiltonian and dynamics for the reconstituted wild-type CP29 protein antenna complex of the Photosystem II. J Phys Chem B 122: 4611–4624

Jia H, Liggins JR, Chow WS (2012) Acclimation of leaves to low light produces large grana: the origin of the predominant attractive force at work. Philos Trans R Soc Lond B Biol Sci 367:3494–502

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10: 845–858

Kereïche S, Kiss AZ, KouYl R, Boekema EJ, Horton P (2010) The PsbS protein controls the macro-organisation of photosystem II complexes in the grana membranes of higher plant chloroplasts. FEBS Lett 584: 759–764

Klimmek F, Sjodin A, Christos N, Dario L, Stefan J (2006) Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. Plant Physiol 140: 793–804

KouYl R, Nosek L, Bartoš J, Boekema EJ, Ilík P (2016) Evolutionary loss of light-harvesting proteins Lhcb6 and Lhcb3 in major land plant groups - break-up of current dogma. New Phytol 210: 808–814

Kovács L, Damkjaer J, Kereiche S, Ilioaia C, Ruban AV, Boekema EJ, Jansson S, Horton P (2006) Lack of the light-harvesting complex CP24 affects the structure and function of the grana

membranes of higher plant chloroplasts. Plant Cell 18: 3106–3120

Koziol AG, Borza T, Ishida KI, Keeling P, Lee RW, Durnford DG (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant Physiol 143: 1802–1816

Laemmli U (1970) Cleavage of the structural protein during the assembly of the head of bacteriophage T4. Nature 227: 680–685

Li XP, Björkman O, Shih C, Grossman AR, Rosenquist M, Jansson S, Niyogi KK (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. Nature 403: 391–395

MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, Kern R, Tabb DL, Liebler DC, MacCoss MJ (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26: 966–968

Liu H, Frankel LK, Bricker TM (2009) Characterization and complementation of a psbR mutant in *Arabidopsis thaliana*. Arch Biochem Biophys 489: 34–40

McElwain JC, Punyasena SW (2007) Mass extinction events and the plant fossil record. Trends Ecol Evol 22: 548–557

Miller MAE, O'Cualain R, Selley J, Knight D, Karim MF, Hubbard SJ, Johnson GN (2017) Dynamic acclimation to high light in *Arabidopsis thaliana* involves widespread reengineering of the leaf proteome. Front Plant Sci 8: 1–15

Neilson JADD, Durnford DG (2010) Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. Photosynth Res 106: 57–71

Pagliano C, Nield J, Marsano F, Pape T, Barera S, Saracco G, Barber J (2014) Proteomic characterization and three-dimensional electron microscopy study of PSII-LHCII supercomplexes from higher plants. Biochim Biophys Acta 1837: 1454–1462

Pan X, Li M, Wan T, Wang L, Jia C, Hou Z, Zhao X, Zhang J, Chang W (2011) Structural insights into energy regulation of light-harvesting complex CP29 from spinach. Nat Struct Mol Biol 18: 309–315

Pan X, Liu Z, Li M, Chang W (2013) Architecture and functin of plant light harvesting complex II. Curr Opin Struct Biol 23: 515–525

Porra RJ, ThompsonWA, Kriedemann PE (1989) Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents: verification of the concentration of chlorophyll standards by atomic absorption spectrometry. Biochim Biophys Acta 975: 384–394

Rardin MJ, Schilling B, Cheng LY, MacLean BX, Sorenson DJ, Sahu AK, MacCoss MJ, Vitek O,

Gibson BW (2015) MS1 peptide ion intensity chromatograms in MS2 (SWATH) data independent acquisitions. Improving post acquisition analysis of proteomic experiments. Mol Cell Proteomics 14: 2405–2419

Ruban AV, Wentworth M, Yakushevska AE, Andersson J, Lee PJ, Keegstra W, Dekker JP, Boekema EJ, Jansson S, Horton P (2003) Plants lacking the main light-harvesting complex retain photosystem II macro-organization. Nature 421: 648–652

Sacharz J, Giovagnetti V, Ungerer P, Mastroianni G, Ruban AV (2017) The xanthophyll cycle affects reversible interactions between PsbS and light-harvesting complex II to control non-photochemical quenching. Nat Plants 3: 16225

Schröppel-Meier G, Kaiser WM (1988) Ion homeostasis in chloroplasts under salinity and mineral deficiency: II. Solute distribution between chloroplasts and extrachloroplastic space under excess or deficiency of sulfate, phosphate, or magnesium. Plant Physiol 87: 828–832

Stockhaus J, Höfer M, Renger G, Westhoff P, Wydrzynski T, Willmitzer L (1990) Anti-sense RNA efficiently inhibits formation of the 10 kd polypeptide of photosystem II in transgenic potato plants: analysis of the role of the 10 kd protein. EMBO J 9: 3013–3021

Su X, Ma J, Wei X, Cao P, Zhu D, Chang W, Liu Z, Zhang X, Li M (2017) Structure and assembly mechanism of plant $C_2S_2M_2$-type PSII-LHCII supercomplex. Science 357: 815–820

van Bezouwen LS, Caffarri S, Kale RS, Kouřil R, Thunnissen AMWH, Oostergetel GT, Boekema EJ (2017) Subunit and chlorophyll organization of the plant photosystem II supercomplex. Nat Plants 3: 17080

Villén J, Gygi SP (2008) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc 3: 1630–1638

Vizcaíno JA, Csordas A, del-Toro N, Dianes JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger F, Ternent T, Xu QW, Wang R, Hermjakob H (2016) 2016 update of the PRIDE database and related tools. Nucleic Acids Res 44(D1): D447–D456

Wei X, Su X, Cao P, Liu X, Chang W, Li M, Zhang X, Liu Z (2016) Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. Nature 534: 69–74

Wood WHJ, MacGregor-Chatwin C, Barnett SFH, Mayneord GE, Huang X, Hobbs JK, Hunter CN, Johnson MP (2018) Dynamic thylakoid stacking regulates the balance between linear and cyclic photosynthetic electron transfer. Nat Plants 4: 116–127

Supporting information

Fig. S1. Absorption spectra of sucrose gradient bands obtained from thylakoids L, M, H solubilized with ±-DDM.

Fig. S2. Western blotting and quantitation of PsbR by immunoblot densitometry.

Fig. S3. Comparison of Lhcb4 and Lhcb5 amino acid sequences from different taxa of the Viridiplantae lineage and phylogeny reconstruction.

Table S1. List of proteins/peptides identified using the custom background proteome.

Table S2. Relative quantification data.

Table S3. List of sequences used for phylogenetic and structural prediction analyses.

Figure legends

Fig. 1. Isolation and characterization of PSII-LHCII supercomplexes obtained from plants grown at low (L), moderate (M) and high (H) light intensity. (A) Sucrose gradient profile of solubilized stacked thylakoid membranes with the main pigment-binding protein complexes of each band indicated on the right and named based on the corresponding protein profiles shown in panels (B-D) and absorption spectra shown in Fig. S1. On each lane of the SDS-PAGEs 3 µg Chl were loaded for sucrose gradient bands and corresponding thylakoids (T). On the left, pre-stained protein markers (lane Mk, Bio-Rad precision plus) with indicated apparent molecular weight (kDa). Abbreviations: m, monomer; t, trimer; sc, supercomplex; mc, megacomplex.

Fig. 2. Spectroscopic and proteomic characterization of PSII-LHCII supercomplexes isolated from plants grown at low (L), moderate (M) and high (H) light intensity. Native absorption spectra of sucrose gradients bands containing PSII-LHCII supercomplexes, normalized to their maximum in the red region (A), and corresponding chlorophyll *a/b* ratios (B). (C) lpBN-PAGE of solubilized thylakoids (T) and sucrose gradient bands containing PSII-LHCII supercomplexes named as in Fig. 1A. Labels on the right indicate the main protein complexes assigned according to Albanese et al. (2016a). Lane Mk, mixture of native high molecular weight marker (GE Healthcare) and blue dextran (Sigma-Aldrich). (D) Graphical heat-map of the relative quantification of the protein content changes of main components of the isolated PSII-LHCII supercomplexes. Scale bar with maximum and

minimum Log_2 fold change (Log_2 FC) is shown, subunits with no significant variation are coloured white. Relative quantification of proteins was assessed by pairwise comparison, within each illumination condition, of the band containing PSII-LHCII supercomplexes with smaller antenna size with respect to the band containing PSII-LHCII supercomplexes with larger antenna size (i.e. L7 *vs* L8; M7 *vs* M8; H6 *vs* H7). Complete quantitative mass-spectrometry data are reported in Table S2. Abbreviations: m, monomer; d, dimer; t, trimer; sc, supercomplex; mc, megacomplex.

Fig. 3. Detection and quantification of PsbR in PSII-LHCII supercomplexes of sucrose gradient bands L8 and H6, isolated from plants grown at low (L) and high (H) light intensity, respectively. (A) Immunoblot with the antibodies against PsbR and D2. The same amount of Chl (3 µg) was loaded on each lane. (B) PsbR amount in the two samples as quantified by western blotting. PsbR content relative to the amount of Chl loaded (black bar) and relative to D2 content (white bar, quantified by the same method on the same gel). Data were normalized to the PsbR content in the H6 sample.

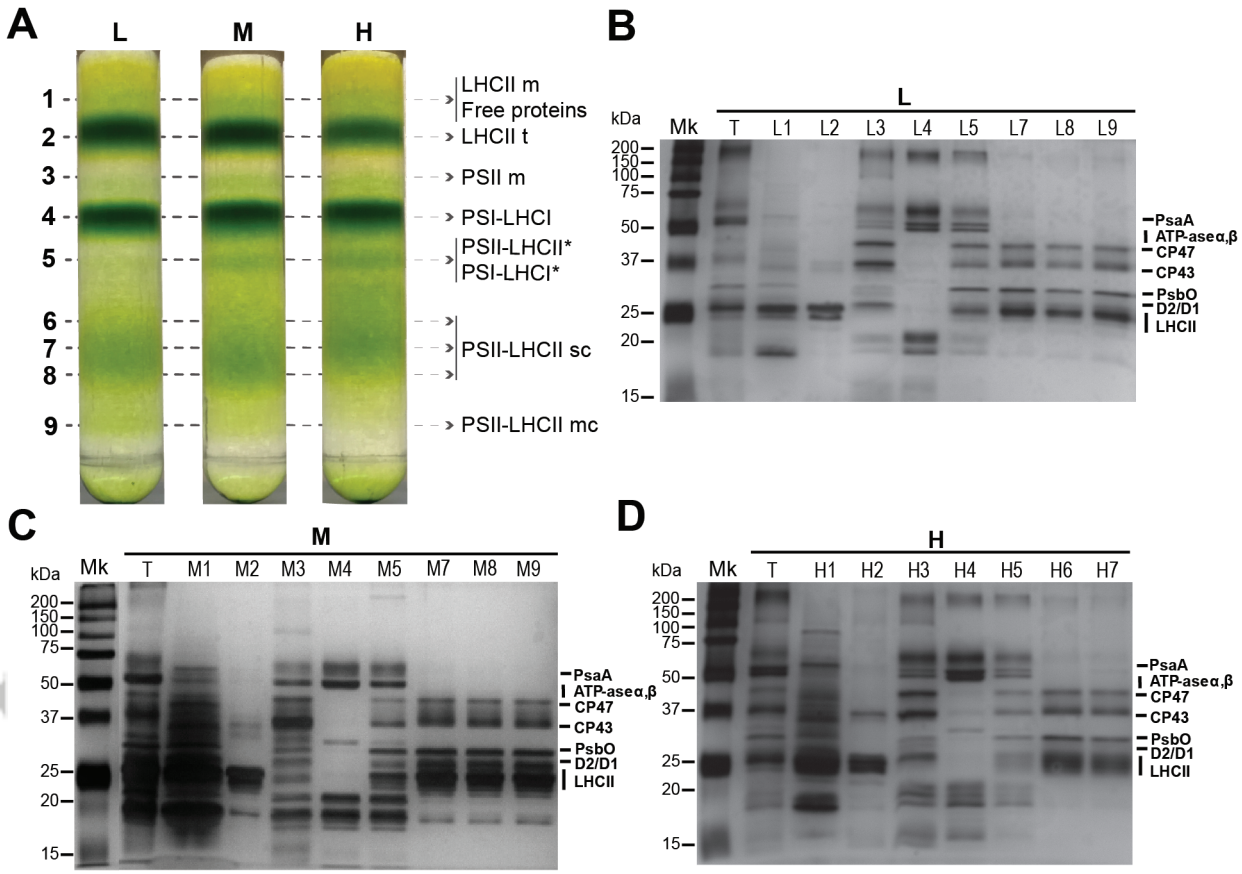
Fig. 4. Phylogenetic and structural comparison of Lhcb4 sequences from different taxa of the Viridiplantae lineage. (A) Phylogenetic tree of 20 Lhcb4 sequences retrieved from phylogenetically divergent photosynthetic organisms, including Lhcb4.3 isoforms when available, and corresponding alignment of the C-terminal domain. Branch lengths are scaled to the number of substitutions per site (scale bar shown). Branches supported by Bootstrap values > 0.7 (> 70% confidence) are marked with a red dot. A condensed branch includes all Lhcb5 sequences of the same organisms used as outgroup, its expanded version is shown in Fig. S3B. Extended species names and accession numbers of the sequences are listed in Table S3. Complete alignment, including the corresponding Lhcb5 sequences, is shown in Fig. S3A. (B) Interaction interface between Lhcb4 (PDB 5XNL: chain R; in red) and Lhcb6 (PDB 5XNL: chain 4; in yellow) and schematic illustration (C) of their positioning in the PSII-LHCII supercomplex of *P. sativum* (as in Su et al. 2017). The Lhcb4 C-terminus and the Chl *b*614 are highlighted. (D-F) Structural superimposition of Lhcb4 from *P. sativum* (PDB 5XNL: chain R; in red) with Lhcb4 structural models (in grey) predicted with PHYRE2 using sequences representative of Lhcb4.3 of angiosperms (*P. sativum* Lhcb4.3, p.sativum_csfl_reftransV1_0068262), Lhcb4 of gymnosperms (*P. sitchensis* Lhcb4, A9NYR8) and Lhcb4 of green algae (*C. reinhardtii* Lhcb4, Q93WD2) shown in panel A. Schematic illustration of the corresponding PSII-LHCII supercomplex organization: for *P. sativum* containing the Lhcb4.3, the proposed organization is based on that of *P. sativum* as in Su et al. (2017) (D), for gymnosperms is based on that of *P. abies* as in KouYl et al. (2016) (E) and for green algae is that of *C. reinhardtii* as in Drop et al. (2014) (F). S, M and N indicate

S-, M- and N-trimers of LHCII, respectively. The degree of rotation of the M-trimer is shown as reported in KouYl et al. 2016 and Drop et al. 2014.

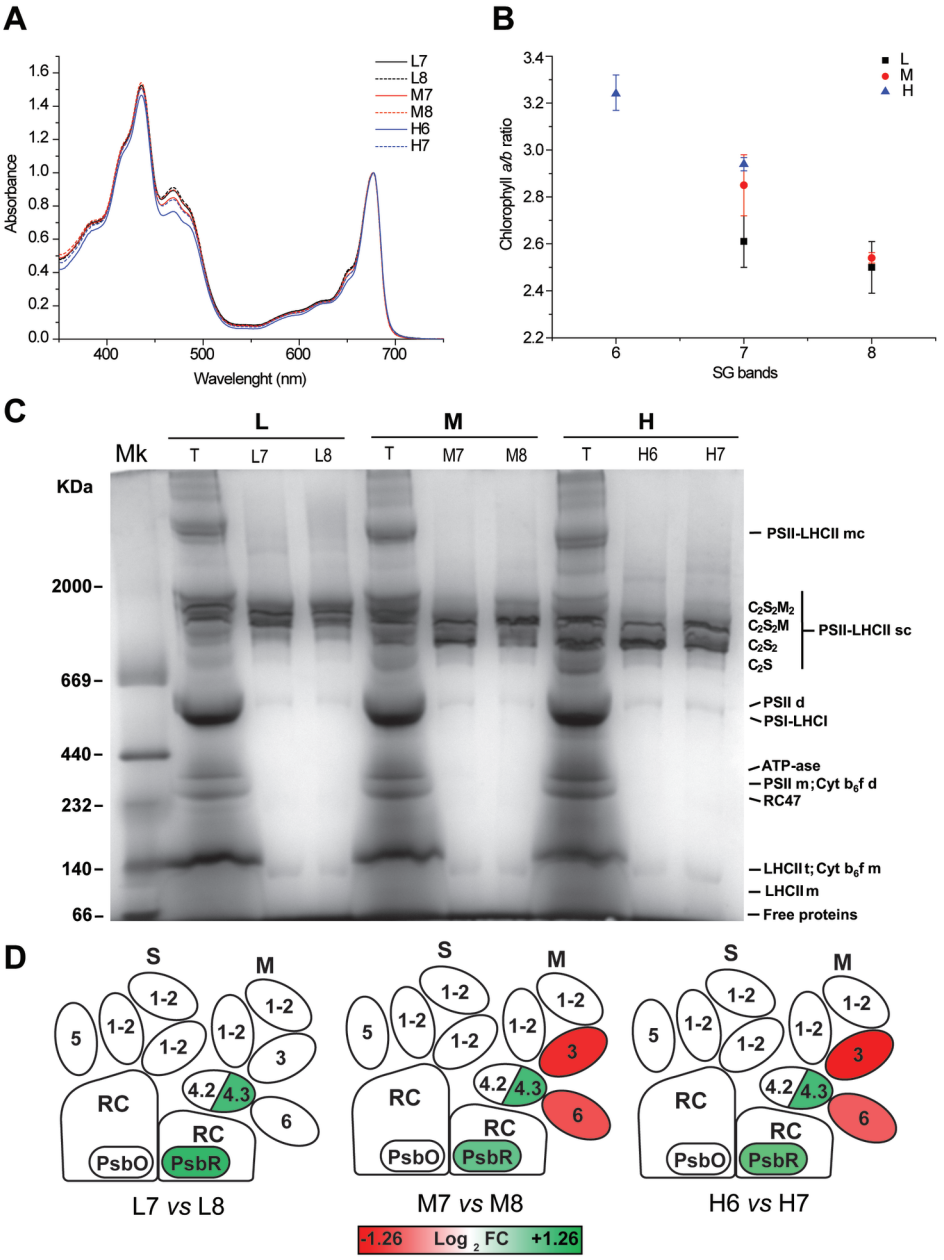
Fig. S1. Absorption spectra of sucrose gradient bands obtained from thylakoids L, M, H solubilized with ±-DDM.

Fig. S2. Western blotting and quantitation of PsbR by immunoblot densitometry. (A) Example of Ponceau staining of a nitrocellulose membrane with loaded increasing amount of PSII-LHCII supercomplexes of sucrose gradient bands L8 and H6 and corresponding antibody signal for detection of D2 and PrbR. 1, 2, 3 µg Chl were loaded for each band. Examples of verification of signal linearity used for the quantification of D2 (B) and PsbR (C) content.

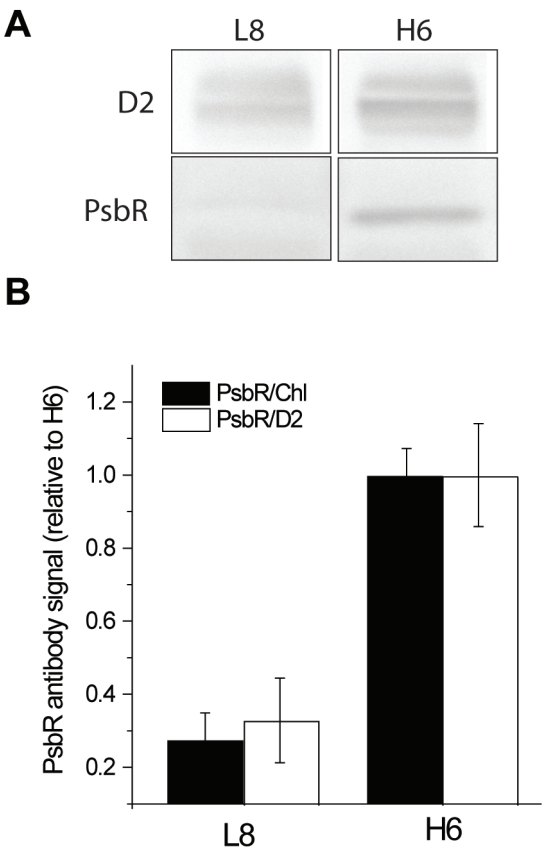
Fig. S3. Phylogenetic analysis of amino acid sequences of Lhcb4 and Lhcb5 from different taxa of the Viridiplantae lineage. (A) Multiple alignment performed with MUSCLE using the default settings, showing the conservation of each position and the consensus sequence. Extended species names and accession numbers of the sequences are listed in Table S3. (B) Phylogenetic tree of the retrieved Lhcb4 and Lhcb5 sequences by Maximum Likelihood method based on the Poisson correction model. The tree with the highest log likelihood (-3440.67) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches and results from 100 Bootstrap replicates. Initial tree(s) was obtained by applying neighbour-joining and bio-neighbour-joining algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (scale bar is shown).



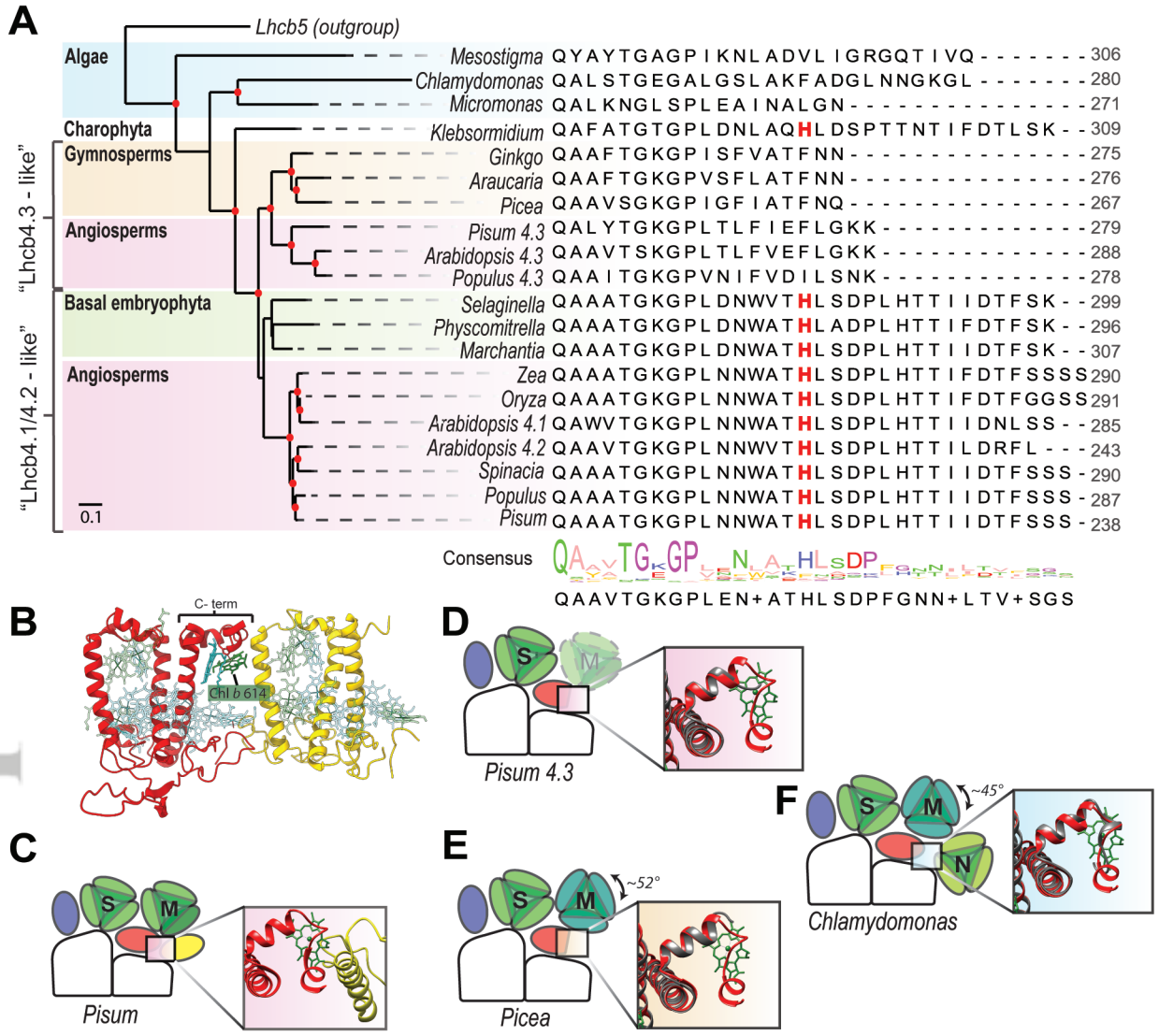
PPL_12964_Fig.1.tif



PPL_12964_Fig.2.tif



PPL_12964_Fig.3.tif



PPL_12964_Fig.4.tif