Journal of Experimental Botany
doi:10.1093/jxb/erz127 Advance Access Publication 2 April 2019
This paper is available online free of all access charges (see https://academic.oup.com/jxb/pages/openaccess for further details)

RESEARCH PAPER

The unique photosynthetic apparatus of Pinaceae: analysis of photosynthetic complexes in *Picea abies*

Steffen Grebe[1], Andrea Trotta[1], Azfar A. Bajwa[1], Marjaana Suorsa[1], Peter J. Gollan[1], Stefan Jansson[2], Mikko Tikkanen[1] and Eva-Mari Aro[1,*]

[1] Molecular Plant Biology, Department of Biochemistry, University of Turku, FI-20014 Turku, Finland
[2] Umeå University, Faculty of Science and Technology, Department of Plant Physiology, Umeå Plant Science Centre (UPSC), SE-90187 Umeå, Sweden

* Correspondence: evaaro@utu.fi

Received 3 December 2018; Editorial decision 11 March 2019; Accepted 13 March 2019

Editor: Christine Raines, University of Essex, UK

Abstract

Pinaceae are the predominant photosynthetic species in boreal forests, but so far no detailed description of the protein components of the photosynthetic apparatus of these gymnosperms has been available. In this study we report a detailed characterization of the thylakoid photosynthetic machinery of Norway spruce (*Picea abies* (L.) Karst). We first customized a spruce thylakoid protein database from translated transcript sequences combined with existing protein sequences derived from gene models, which enabled reliable tandem mass spectrometry identification of *P. abies* thylakoid proteins from two-dimensional large pore blue-native/SDS-PAGE. This allowed a direct comparison of the two-dimensional protein map of thylakoid protein complexes from *P. abies* with the model angiosperm *Arabidopsis thaliana*. Although the subunit composition of *P. abies* core PSI and PSII complexes is largely similar to that of Arabidopsis, there was a high abundance of a smaller PSI subcomplex, closely resembling the assembly intermediate PSI* complex. In addition, the evolutionary distribution of light-harvesting complex (LHC) family members of Pinaceae was compared *in silico* with other land plants, revealing that *P. abies* and other Pinaceae (also Gnetaceae and Welwitschiaceae) have lost LHCB4, but retained LHCB8 (formerly called LHCB4.3). The findings reported here show the composition of the photosynthetic apparatus of *P. abies* and other Pinaceae members to be unique among land plants.

Keywords: Blue native gel, conifer, light harvesting, photosystem, *Picea abies* (Norway spruce), Pinaceae, thylakoid protein complexes.

Introduction

Evergreen conifers are gymnosperms comprising the families (and orders) Pinaceae (Pinales), Araucariaceae, and Podocarpaceae (Araucariales), and Sciadopityaceae, Taxaceae, and Cupressaceae (Cupressales) (Christenhusz *et al.*, 2011). Many conifers have a paramount role as carbon sinks in boreal forest ecosystems of the northern hemisphere and are of substantial economic importance (Shorohova *et al.*, 2011; Gauthier *et al.*, 2015), yet the effects of climate change on the photosynthesis capacity of boreal forests and their capacity to sequester CO_2 from the atmosphere remain largely unknown.

Here we focus on the photosynthetic machinery of Norway spruce (*Picea abies*) as a representative species of Pinaceae with high acclimation capacity to harsh environmental conditions occurring in the northern hemisphere or at high latitudes. Like all plants, Pinaceae perform photosynthetic light

reactions using two photosystems (PSII and PSI), to produce NADPH and ATP used to reduce atmospheric CO_2 in the Calvin–Benson–Bassham cycle. PSII and PSI have their own tightly bound light-harvesting complex (LHC) antenna systems, LHCII (LHCB) and LHCI (LHCA), respectively. Additionally, the thylakoid membrane accommodates a large amount of trimeric LHCII complexes that can deliver excitation to both photosystems. Linear electron transport from PSII to PSI is mediated by the cytochrome b_6f complex (Cyt-bf) and electron carriers plastoquinone and plastocyanin, leading to reduction of ferredoxin and production of NADPH via ferredoxin–$NADP^+$ reductase. Electron transport is coupled to proton translocation across the thylakoid membrane from stroma to lumen, which builds a proton-motive force that is utilized by the chloroplastic ATP synthase to produce ATP. Fluent function of photosynthetic light reactions requires strict co-regulation by a number of different mechanisms, according to environmental cues, in order to avoid harmful side reactions (for reviews, see Rochaix, 2014; Schöttler and Tóth, 2014; Tikkanen and Aro, 2014; Demmig-Adams *et al.*, 2017).

Nordic climate conditions are characterized by harsh winters, which expose the photosynthetic machinery of evergreen needles to a severe imbalance between the supply and utilization of light energy. During sunny days in cold seasons, the evergreen needles absorb light that cannot be utilized in CO_2 assimilation, since the biochemical reactions are largely restricted by low temperatures. This imbalance can damage the photosynthetic apparatus if the needles are not sufficiently acclimated (Öquist and Huner, 2003). Photosynthetic acclimation of evergreen needles to winter conditions has been extensively investigated in Pinaceae. During winter and early spring a sustained form of non-photochemical light energy dissipation is activated, which takes several days to fully relax even in favorable conditions (Verhoeven, 2013). During winter quenching, Pinaceae also retain large amounts of the xanthophyll cycle pigments zeaxanthin and antheraxanthin (Ottander *et al.*, 1995; Verhoeven *et al.*, 1996, 1999, 2009; Merry *et al.*, 2017). The acclimation process also involves changes in thylakoid protein phosphorylation and in the relative abundance of photosynthetic proteins (Ottander *et al.*, 1995; Verhoeven *et al.*, 2009; Merry *et al.*, 2017). However, results in the literature on this topic vary depending on the species investigated and on environmental conditions during the studies (for review see Verhoeven, 2014).

The current picture on thylakoid operation is largely based on comprehensive research on the model angiosperm *Arabidopsis thaliana* using wild type and mutant plants lacking thylakoid regulatory proteins (for reviews, see Tikkanen and Aro, 2014; Alric and Johnson, 2017; Armbruster *et al.*, 2017). Dynamics, acclimation, composition, and organization of the photosynthetic apparatus in Arabidopsis provide a reference point for investigation of Pinaceae. Nevertheless, the unique characteristics of the gymnosperm photosynthetic membranes, including fundamental differences from Arabidopsis, make it imperative to first provide the tools to investigate these aspects in Pinaceae directly.

Making use of all available genomic and transcript data for *P. abies* (Nystedt *et al.*, 2013), we first created a thylakoid protein database for this species that enabled analysis of the composition of the thylakoid protein complexes by two-dimensional large pore blue-native (2D lpBN/SDS-PAGE) coupled with tandem mass spectrometry (MS/MS) protein identification. This generated a detailed map of the subunits of the protein complexes in *P. abies* thylakoids. Further, the analysis of the LHC family members was extended to sequences of other gymnosperms revealing that LHCB4.1 and LHCB4.2 have been lost from the genomes of all Pinaceae members, while LHCB4.3, which is now called LHCB8, has been conserved. Pinaceae have also lost LHCA5 and have substantial amounts of a PSI subpopulation called PSI⋆. This study provides new tools to pave the way to resolving the molecular mechanisms that govern the function and survival of Pinaceae photosynthetic apparatus in extreme environmental conditions.

Materials and methods

Plant material

Needles for all experiments were collected from *P. abies* (L.) Karst. trees grown in a natural forest in Turku, Southern Finland (60°27′N, 22°16′E). South-facing branches (up to 2 m in height) were cut from five different trees at noon on 18 June 2014, 24 June 2015 and 9 June 2016. Cut branches were placed in a light-proof plastic bag, transported to the laboratory, and immediately used for thylakoid isolations. *Arabidopsis thaliana* (Col-0) was grown in a controlled environment (8 h/16 h day/night cycle, 23°C, 50% relative humidity, 125 μmol photons m^{-2} s^{-1} light intensity). Four- to five-week-old Arabidopsis plants were used for thylakoid isolations.

Thylakoid isolation and chlorophyll determination

Healthy, mature *P. abies* needles were harvested from all five branches under dim light and pooled on ice. Ten grams (fresh weight) of needles was transferred into an ice-cold homogenizer (2 inch blades and 200 ml stainless steel chamber, Omni-Inc, GA, USA) and 100 ml grinding buffer was added (50 mM Hepes–KOH, pH 7.5; 330 mM sorbitol; 5 mM $MgCl_2$; 10 mM NaF; 10% (w/v) polyethylene glycol, 6000 kDa; 0.075% (w/v) bovine serum albumin; 0.065% (w/v) Na-ascorbate). All steps were performed in a cold room at 4°C with ice-cold reagents. Needles were homogenized for 90 s at 8000 rpm, filtered through two layers of Miracloth, and centrifuged at 4600 *g* for 6 min. The pellet was re-suspended with a soft paint brush in 25 ml shock buffer (50 mM Hepes–KOH, pH 7.5; 5 mM $MgCl_2$; 10 mM NaF), transferred into a new centrifuge tube, and the thylakoids were pelleted at 4600 *g* for 6 min. Thylakoids were re-suspended in 15 ml storage buffer (50 mM Hepes–KOH, pH 7.5; 5 mM $MgCl_2$; 100 mM sorbitol; 10 mM NaF), starch and residual PEG were removed by brief centrifugation (2 min, 200 *g*) and thylakoids were again pelleted at 4600 *g* for 6 min. Thylakoid pellets were re-suspended with a brush in 200 μl of storage buffer, aliquoted, and immediately frozen in liquid nitrogen for later use. All buffers contained 10 mM NaF to preserve the *in vivo* state of the thylakoid supercomplexes.

Isolation of thylakoids from Arabidopsis was carried out according to Suorsa *et al.* (2015). Concentration of chlorophyll extracted from thylakoids in 80% buffered acetone was determined according to Porra *et al.* (1989).

2D-lpBN-PAGE and protein staining

Thylakoids were solubilized with *n*-dodecyl β-D-maltoside (β-DM) at 1% and 2% (w/v) final concentration for Arabidopsis and *P. abies*, respectively. Two per cent (w/v) β-DM was required to effectively solubilize *P. abies* thylakoids (see Supplementary Fig. S1 at *JXB* online). The solubilized thylakoids were subjected first to lpBN-PAGE for separation of the protein complexes and subsequently to SDS-PAGE for identification of

the protein subunits of each complex as described in Järvi *et al.* (2011). SYPRO Ruby (Invitrogen) staining of the gels was performed according to the instructions supplied. Silver staining was carried out according to Blum *et al.* (1987).

Mass spectrometry analysis

Protein spots were excised from stained gels and subjected to in-gel tryptic digestion as described by Suorsa *et al.* (2015). Eluted peptides were identified by nanoscale liquid chromatography–electrospray ionization MS/MS using either a Q-Exactive or a Q-Exactive-HF mass spectrometer (Thermo Scientific) and applying a gradient from solvent A (0.1% formic acid) to B (80% acetonitrile, 0.1% formic acid) of 8–43% for 10 min followed by a step to 100% for 2 min and 8 min 100% solvent B. The 10 most intense peaks in every full MS scan (range 300–2000 m/z) with a resolution of 120 000 were selected for fragmentation in MS2 with a dynamic exclusion window of 10 s. The acquired spectra were matched against a custom protein database (see below) using Proteome Discoverer 2.2 (Thermo Scientific) with an in-house installation of the Mascot server (v. 2.5.1), allowing Cys carbamidomethylation as fixed modification and Met oxidation, Asn/Gln deamination and protein N-terminal acetylation as dynamic modifications.

The mass spectrometry proteomics data were deposited at the ProteomeXchange Consortium (PXD010071, http://proteomecentral.proteomexchange.org; accessed 15/04/2019) via the PRIDE partner repository (Vizcaíno *et al.*, 2013).

Custom protein database for mass spectrometry analysis

A custom protein database was constructed from three different public databases via procedures shown in Fig. 1. *Picea abies* peptide sequences were derived from Database 1 (chloroplast-encoded gene models, NCBI, NC_021456.1), Database 2 (nuclear-encoded gene models, ConGenIE, Nystedt *et al.*, 2013) and Database 3 (transcripts, ConGenIE). Only transcripts from thylakoid-associated proteins were included, since the MS analysis focused on this subset of proteins.

Candidate transcript sequences from *P. abies* were identified using tBLASTn homology searches with Arabidopsis or *Physcomitrella patens* reference sequence queries (Phytozome V11, Joint Genome Institute). The 10 highest scoring BLAST hits were translated to candidate protein sequences and manually checked for redundancy and possible sequencing errors. Full-length protein sequence candidates were trimmed to the predicted N-terminal methionine, and truncated sequences were only used when no full-length protein sequences were found. To exclude contamination (e.g. from lichens; Delhomme *et al.*, 2015) phylogenies of orthologous sequences of representative model species were constructed (*A. thaliana*, *Selaginella moellendorffii*, *Ph. patens, Chlamydomonas reinhardtii*, *Synechocystis* sp. PCC 6803 and *Anabaena* sp. PCC 7120 collected from

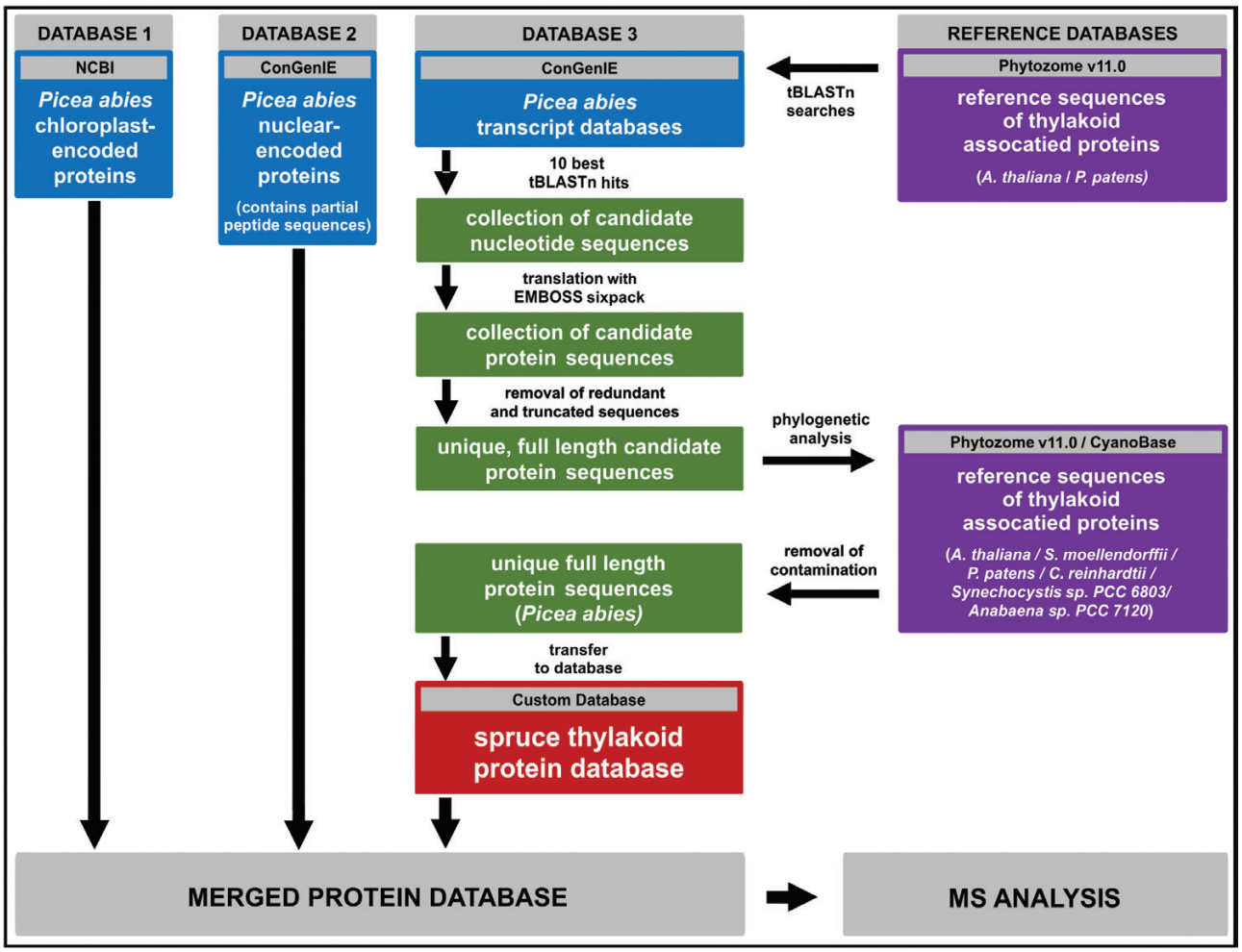

Fig. 1. Procedure for generation of the merged protein database for MS/MS analysis of thylakoid proteins in *P. abies*. Database 1 and Database 2 with chloroplast- and nuclear-encoded protein sequences, respectively, were merged with protein sequences derived from transcripts (Database 3). Transcript nucleotide sequences were selected by tBLASTn searches with known thylakoid associated protein sequences from reference species. These candidate sequences were translated to protein sequences and contaminating as well as truncated sequences were removed. This procedure led to an annotated spruce thylakoid protein database (red box, Supplementary dataset S1), which was combined with Database 1 and Database 2 to form a merged protein database (Supplementary dataset S2) that was used for MS/MS identification of *P. abies* thylakoid proteins.

Phytozome V11 and CyanoBase) and candidate sequences occurring together with algal or cyanobacterial orthologues were removed. In cases where multiple unique *P. abies* sequences were obtained for a single protein, all sequence variations were considered equally valid and therefore included in the custom database (Supplementary dataset S1). Finally, all three databases were combined into a merged protein database that was used for MS spectrum searches (Supplementary dataset S2).

Identification of LHC family proteins from land plants

LHC protein sequences from land plants were identified by homology searches using reference LHC protein sequences from Arabidopsis (LHCB1,AT1G29920.1;LHCB2,AT2G05100.1;LHCB3,AT5G54270.1; LHCB4.1 (LHCB4),AT5G01530.1;LHCB4.3 (LHCB8),AT2G40100.1; LHCB5,AT4G10340.1;LHCB6,AT1G15820.1;LHCB7,AT1G76570.1; LHCA1,AT3G54890.1;LHCA2,AT3G61470.1;LHCA3,AT1G61520.1; LHCA4,AT3G47470.1;LHCA5,AT1G45474.1;LHCA6,AT1G19150.1) and from *Ph. patens* (LHCB9, Pp3c5_22920v3.1) against databases from 84 different species (Supplementary Table S1). The 10 highest scoring BLAST hits for each reference LHC sequence were translated to amino acid sequences and aligned with the reference. Redundant and truncated sequences (≤20% of reference sequence) were discarded and candidate sequences were trimmed to the predicted N-terminal methionine.

A multiple sequence alignment containing all reference LHCA and LHCB sequences was constructed and six diagnostic regions were identified (Fig. 2) to clarify the identities of candidate LHC sequence from other species. Candidate sequences were considered true orthologues when they shared >75% sequence identity with a particular reference sequence within at least one diagnostic region. Strong conservation of selected diagnostic regions across species was illustrated by constructing sequence logos from a multiple sequence alignment of each region within each LHC orthologue group (see Results and Supplementary Table S2).

Excluded from this procedure were LHCBM sequences, which severely disrupted the LHC multiple sequence alignment. LHCBM sequences of non-vascular plants were assigned according to homology to LHCBM sequences from *Ph. patens*. Algal LHCA proteins were assigned to their closest land plant orthologue group (Alboresi *et al*., 2008; Iwai and Yokono, 2017).

Multiple sequence alignments and phylogenetic reconstruction

Multiple sequence alignments were performed in MEGA7 (Kumar *et al*., 2016) using the MUSCLE algorithm (Edgar, 2004) and visualized with BioEdit v7.0.5. Sequence logos constructed from multiple sequence alignments were created with WebLogo3 (Schneider and Stephens, 1990;

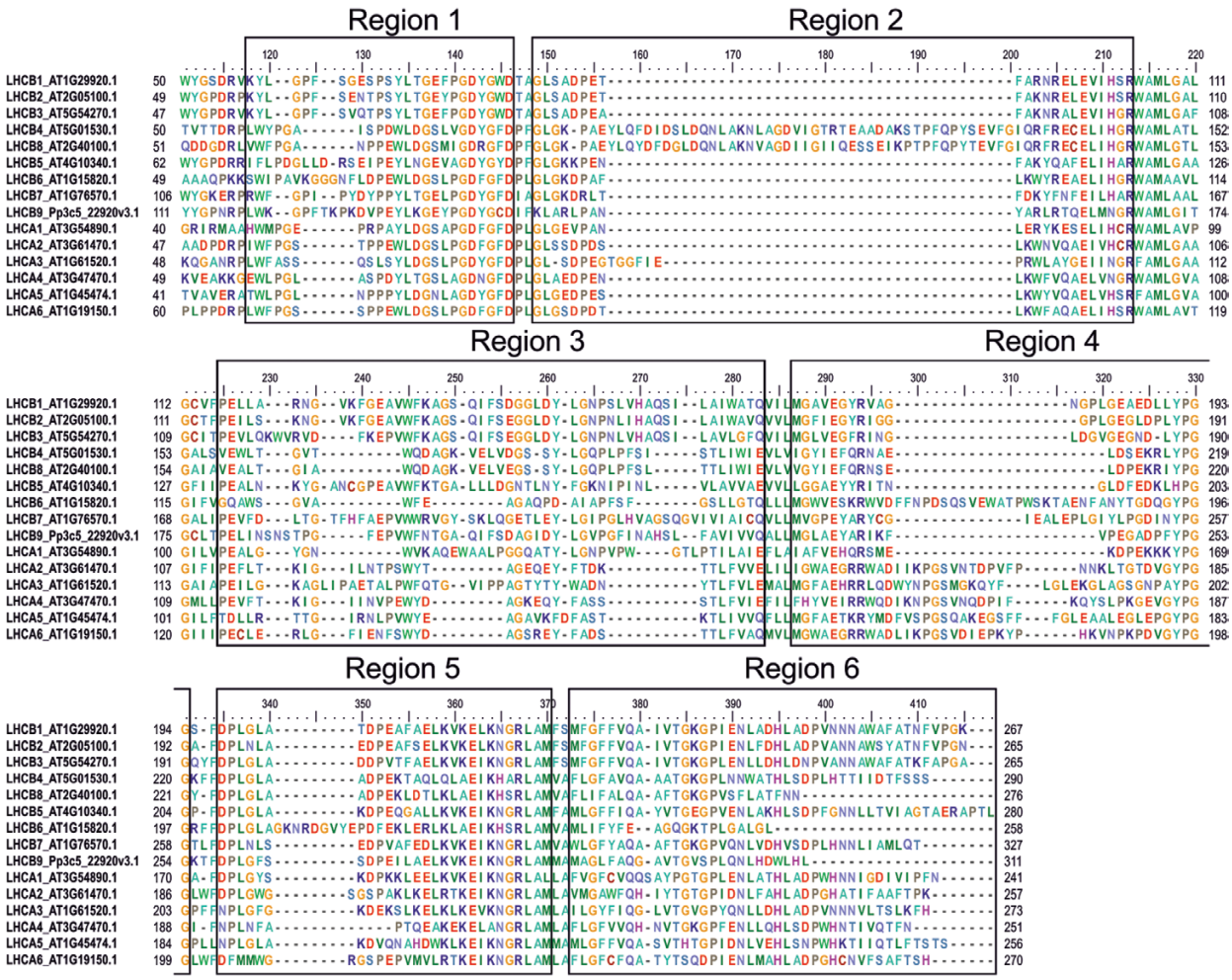

Fig. 2. Diagnostic regions for LHC homologue identification. Multiple sequence alignment of LHCB1–8 and LHCA1–6 proteins from Arabidopsis and LHCB9 from *Physcomitrella patens* was generated and manually adjusted for maximal sequence overlap. N-terminal protein sequences are not shown. Boxes indicate diagnostic regions 1–6, which were used to classify LHC protein sequences. Each LHC candidate sequence was considered a true orthologue if it shared >75% sequence identity with a corresponding reference sequence. Strong conservation of selected diagnostic regions across species was illustrated by constructing sequence logos from a multiple sequence alignment of each region within each LHC orthologue group (see Results and Supplementary Table S2).

Crooks *et al.*, 2004). Phylogenetic reconstruction from polypeptide sequences was performed in MEGA7 with maximum likelihood method.

Results

Picea abies *thylakoid protein database*

Although the sequences of chloroplast- and nuclear-encoded proteins are available from the sequenced *P. abies* genome (Nystedt *et al.*, 2013), sequences derived from automated gene model annotation in publicly available databases were often found not to contain the full-length proteins. Comprehensive analysis of *P. abies* thylakoid protein complexes by MS demanded full-length sequences of the respective proteins; hence it was necessary to build and curate a customized protein database to complement the existing *P. abies* database.

The merged custom database (Fig. 1) contained 168 unique polypeptide sequences from transcripts (spruce thylakoid protein database, Supplementary dataset S1), 74 polypeptide sequences from the chloroplast encoded gene models (Database 1) and 66 632 polypeptide sequences from the nuclear-encoded gene models (Database 2). In total 66 874 *P. abies* polypeptide sequences and 115 additional polypeptide sequences from the common Repository of Adventitious Proteins (cRAP) database (https://www.thegpm.org/crap/; accessed 15/04/2019) were used for identification via MS (merged protein database, Supplementary dataset S2).

In silico *comparison of light-harvesting complex family proteins in gymnosperms and other land plants*

While building the custom database, it became evident that deep analysis was required to determine the correctly assigned LHC family protein sequences (e.g. LHCB1 or LHCB2) to be included. To this end, the distribution of LHC family proteins LHCA1–6 and LHCB1–9 were investigated in 84 different species ranging from green algae (Chlorophyta), liverworts (Marchantiophyta), mosses (Bryophyta), hornworts (Anthocerotophyta), Lycophytes and ferns (Monilophytes) to seed plants (gymnosperms and angiosperms). Deep comparison was focused on differences between angiosperm and gymnosperm LHC sequences. In the case of Pinaceae, every genus (*Picea*, *Pinus*, *Abies*, *Cedrus*, *Tsuga*, *Pseudotsuga*, *Nothotsuga*, *Larix*, *Pseudolarix*, *Keteleeria*, and *Cathaya*) was represented by at least one species.

Since all LHC proteins are homologous, the LHC proteins were identified by homology searches of various genome and transcriptome databases using reference LHC sequences from Arabidopsis (LHCA1–6, LHCB1–8) and *Ph. patens* (LHCB9). To this end, six manually determined diagnostic regions of Arabidopsis LHC proteins (see 'Material and methods') were defined (Fig. 2). This allowed the identification of truncated LHC sequences that would not have been found by relying only on the best homology hits from BLAST searches. This analysis yielded a total of 1366 LHC protein sequences in 84 species. The designation of orthology between each LHC protein and its corresponding reference sequence was supported by strong homology of selected diagnostic regions within each orthologous group, as illustrated in sequence logos for each region in Supplementary Table S2.

The identified LHC proteins in each species were used to create a broad overview of the distribution of LHC homologues in plants. The comparison revealed large variation in the occurrence of LHC proteins between different evolutionary groups, as described in detail below. The results for 35 representative species are shown in Fig. 3 and for all 84 investigated species are shown in Supplementary Table S3.

Among the LHCB proteins, LHCB4 and LHCB8 appeared to have the most distinct distribution between gymnosperms and angiosperms. There are three isoforms of LHCB4 known in Arabidopsis: LHCB4.1 and LHCB4.2 (hereafter LHCB4, as their amino acid sequences are 89% identical and 92% similar) with a longer C-terminus, and LHCB4.3 (hereafter LHCB8 as earlier suggested by Klimmek *et al.*, 2006) with a shorter C-terminus. In the current study, analysis of angiosperms found that LHCB4 was completely conserved, while LHCB8 occurred only in Eurosids (Malvids and Fabids) and in the order Caryphyllales. In the gymnosperm species investigated, LHCB4 and LHCB8 proteins were also differentially present, with both proteins identified in members of Araucariaceae and Podocarpaceae (Araucariales), Sciadopityaceae, Taxaceae, and Cupressaceae (Cupressales), and Ginkgoaceae (Ginkgoales), while LHCB4 was not found in Pinaceae (Pinales), Gnetaceae (Gnetales), or Welwitschiaceae (Welwitschiales) and LHCB8 was not found in the evolutionarily older Cycadaceae and Zamiacae (Cycadales). Interestingly, LHCB4, but not LHCB8, was found in the evolutionarily older non-seed plants (see Fig. 3).

Isolation of LHCB4 and LHCB8 sequences from 63 and 43 different species, respectively, allowed the characterization of distinct differences in the C-termini of orthologues from different evolutionary groups (Fig. 4). The C-terminus of LHCB4 was found to be highly conserved in both length and amino acid composition across angiosperms, gymnosperms (excepting *P. abies* and other members of Pinaceae in which LHCB4 was not found) and non-seed plants. The C-terminus of LHCB8 in Eurosids and gymnosperms was approximately 10 amino acids shorter than in LHCB4, while the Caryphyllales-type LHCB8 was more similar in length to LHCB4. LHCB8 C-termini also showed substantial amino acid sequence variation between gymnosperms, Eurosids and Caryphyllales, although they were well conserved within each orthologue group, as illustrated in the sequence logos in Fig. 4. Importantly, a 15-amino-acid motif that was strictly conserved in LHCB4 C-terminus (WxTHLxDPLHTTIxD; residues 271–285 in Lhcb4.1 AT5G01530.1, Arabidopsis), was absent from all LHCB8 sequences isolated here. Another unique feature identified in LHCB4 and LHCB8 was a sequence insertion, relative to other LHC sequences, between amino acids 85 and 134 (Lhcb4.1 AT5G01530.1, Arabidopsis). This region harbored considerable sequence variability, both between LHCB4 and LHCB8 and between orthologues from different species (see diagnostic region 2, Supplementary Table S2).

LHCB3 and LHCB6 proteins were not identified in Pinaceae, Gnetaceae, and Welwitschiaceae, but were present in other gymnosperms, as reported earlier (Kouřil *et al.*, 2016), and in all angiosperm species investigated, as well as in evolutionarily early non-seed land plants. These LHC homologues

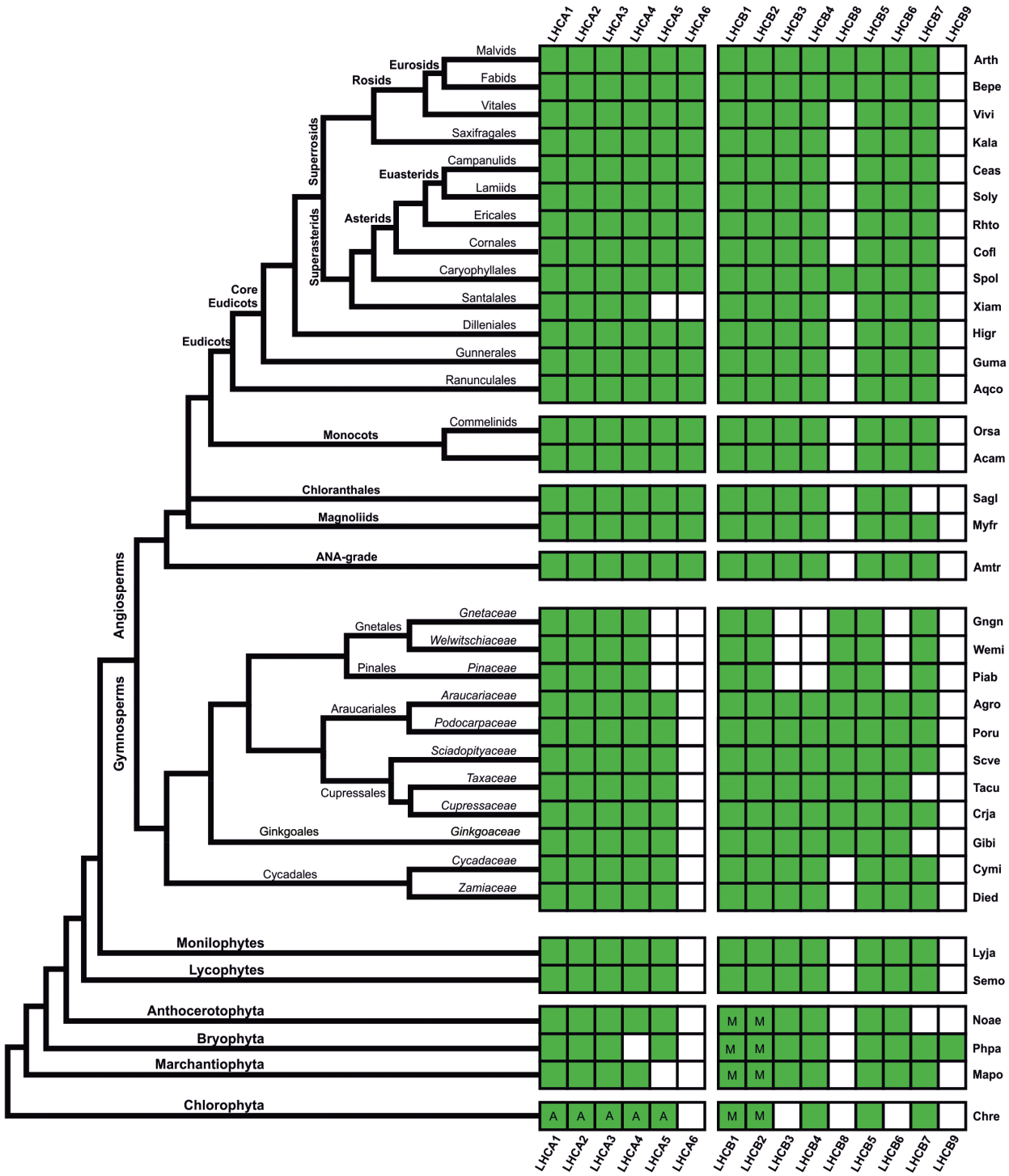

Fig. 3. Overview of LHCA1–6 and LHCB1–9 protein distribution in 35 land plant species. Green boxes indicate LHC homologue was identified, white boxes indicate LHC homologue was not identified. Phylogenetic tree drawn after Clarke *et al.* (2011) with classifications for gymnosperms after Christenhusz *et al.* (2011) and for angiosperms after APG IV system (The Angiosperm Phylogeny Group *et al.*, 2016). Different hierarchies are indicated by typeface: bold (divisions, class and clade), normal (order) and italic (families). Different evolutionary groups are represented by 35 selected species (for all investigated species see Supplementary Table S3): Acam, *Acorus americanus*; Agro, *Agathis robusta*; Amtr, *Amborella trichopoda*; Aqco, *Aquilegia coerulea*; Arth, Arabidopsis; Bepe, *Betula pendula*; Ceas, *Centella asiatica*; Chre, *Chlamydomonas reinhardtii*; Cofl, *Cornus florida*; Crja, *Cryptomeria japonica*; Cymi, *Cycas micholitzii*; Died, *Dioon edule*; Gibi, *Ginkgo biloba*; Gngn, *Gnetum gnemon*; Guma, *Gunnera manicata*; Higr, *Hibbertia grossulariifolia*; Kala, *Kalanchoe laxiflora*; Lyja, *Lygodium japonicum*; Mapo, *Marchantia polymorpha*; Myfr, *Myristica fragrans*; Noae, *Nothoceros aenigmaticus*; Orsa, *Oryza sativa*; Phpa, *Physcomitrella patens*; Piab, *P. abies*; Poru, *Podocarpus rubens*; Rhto, *Rhododendron tomentosum* (*Ledrum palustre*); Sagl, *Sarcandra glabra*; Scve, *Sciadopitys verticillata*; Semo, *Selaginella moellendorffii*; Soly, *Solanum lycopersicum*; Spol, *Spinacia oleracea*; Tacu, *Taxus cuspidate*; Vivi, *Vitis vinifera*; Wemi, *Welwitschia mirabilis*; Xiam, *Ximenia americana*. 'A' indicates algae-specific LHCA isoforms adapted to land plant LHCA nomenclature (Alboresi *et al.*, 2008; Iwai and Yokono, 2017); 'M' indicates LHCBM orthologues instead of LHCB1 and LHCB2 in Chlorophyta, Marchantiophyta, Bryophyta, and Anthocerotophyta.

Fig. 4. Sequence logos of LHCB4 and LHCB8 C-termini in different plant groups. Number of species and sequences used for logos as well as reference sequence with accession number and amino acid position for each plant group were the following: LHCB4 – Angiosperms, 41 species, 52 sequences: Arabidopsis AT5G01530.1, 253–290; LHCB4 – Gymnosperms, 16 species, 16 sequences: *Sequoia sempervirens* UCSsempervirens_isotig06909, 262–297; LHCB4 – non-seed plants 7 species, 12 sequences: *Physcomitrella patens* Pp3c4_5680V3.1, 260–296; LHCB8 – Eurosids, 8 species, 9 sequences: Arabidopsis AT2G40100.1, 253–276; LHCB8 – Caryophyllales, 5 species, 5 sequences: *Spinacia oleracea* XP_021841502, 250–284; LHCB8 – Gymosperms, 15 species, 15 sequences: *P. abies* comp95233_c3_seq1, 275–300. Sequence logos were generated with WebLogo3.6 (Schneider and Stephens, 1990; Crooks *et al.*, 2004).

were found to be absent from green algae, as previously established (Alboresi *et al.*, 2008). LHCB5 orthologues were found in every species investigated (Supplementary Table S3), while the presence of LHCB1, LHCB2, and LHCB7 was variable. LHCB9 is a specific LHC homologue of the moss *Ph. patens* (Alboresi *et al.*, 2008) and was identified only in that species, while LHCBM sequences replaced LHCB1 and LHCB2 in Chlorophyta, Marchantiophyta, Bryophyta, and Anthocerotophyta.

In contrast to the LHCB proteins, which showed varying distribution in different evolutionary plant groups, the LHCA proteins LHCA1–4 were found in all species investigated with the known exception of LHCA4 in *Ph. patens* (Alboresi *et al.*, 2008). LHCA isoforms, which are more abundant in algae than in land plants, were here grouped according to their land plant orthologues (Alboresi *et al.*, 2008; Iwai and Yokono, 2017). LHCA5 and LHCA6 were present in angiosperms, except in *Ximenia americana* (Olacaceae, Santales) and *Lophophora williamsii* (Cataceae, Caryophyllales), which lacked both proteins. Generally, LHCA6 was only identified in angiosperm species (Fig. 3; Supplementary Table S3).

LHCA5 was present in most gymnosperm species; however, it was not identified in any members of Pinales, Gnetales, or Welwitschiales, nor in *Cycas rhumpii* (Supplementary Table S3). Among non-seed plants, LHCA5 was also missing from *Equisetum hyemale* (Equisetales, Monoliophytes) and *Marchantia polymorpha* (Marchantiales, Marchantiophyta; Ueda *et al.*, 2012).

Thylakoid protein complexes in Picea abies

The *in silico* identification of the LHC family proteins was expanded to a biochemical analysis with identification of thylakoid protein complexes isolated from summer needles (see 'Material and methods') of *P. abies* (chlorophyll *a*/*b* ratio 3.42 ± 0.09), which were compared with those of Arabidopsis (chlorophyll *a*/*b* ratio 3.19 ± 0.01). The isolated thylakoids of both species were solubilized with β-DM, using 2% β-DM for *P. abies* and 1% β-DM for Arabidopsis as final concentrations, and the photosynthetic protein complexes were separated via lpBN-PAGE (Fig. 5). The entire thylakoid membrane, including the appressed grana partitions, is solubilized by β-DM (Järvi *et al.*, 2011; Suorsa *et al.*, 2015), while stronger protein–protein interactions remain intact, thus allowing the investigation of the basic building blocks of native thylakoid protein complexes (Rantala *et al.*, 2017). The concentration of 2% β-DM for *P. abies* was based on selecting the most optimal detergent concentration that on one hand solubilized all the protein complexes and on the other hand kept the larger supercomplexes intact (see Supplementary Fig. S1).

The lpBN-PAGE separation showed a typical pattern for thylakoid protein complexes in both species (Fig. 5). The bands were named according to the latest Arabidopsis nomenclature (Rantala *et al.*, 2017). In the high molecular mass region in Arabidopsis and *P. abies*, four bands of the PSII–LHCII supercomplexes (PSII–LHCII sc) were visible (Caffarri *et al.*, 2009; Suorsa *et al.*, 2015; Rantala *et al.*, 2017). They comprise two PSII core complexes (C2, PSII core dimer) with associated LHCII trimers (LHCB1–3) in different stoichiometric ratios. The LHCII trimers are categorized as strongly bound LHCII trimers (S) or moderately bound LHCII trimers (M), both of which are connected to the PSII core via LHCB5 or LHCB4/8 and LHCB6, respectively (Caffarri *et al.*, 2009). PSII core dimers together with various LHCII species give rise to C2S2M2, C2S2M1, C2S2, and C2S1 supercomplexes in the

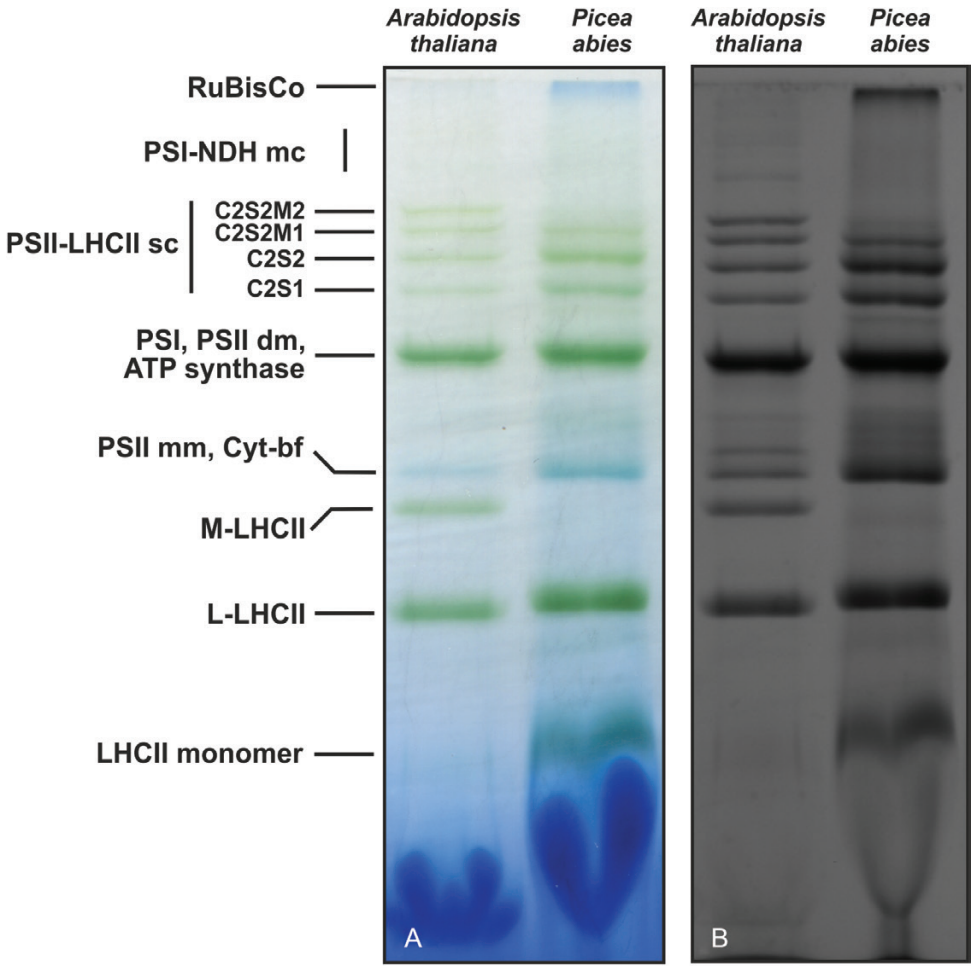

Fig. 5. Comparison of thylakoid protein complexes isolated from Arabidopsis and *P. abies*. After solubilization with *n*-dodecyl-β-D-maltoside (β-DM), the thylakoid complexes were (A) separated by lpBN-PAGE and (B) subsequently stained with Coomassie. Arabidopsis and *P. abies* thylakoids were solubilized with 1% and 2% β-DM, respectively. Samples were loaded with 8 µg chlorophyll per lane. dm, dimer; mc, megacomplex; mm, monomer; sc, supercomplex.

thylakoid membrane. lpBN-PAGE of solubilized thylakoids from both Arabidopsis and *P. abies* achieved sufficient separation of lower molecular mass complexes PSI, PSII dimer (PSII dm), and ATP synthase from the smaller PSII monomer (PSII mm) and Cyt-bf complexes, as well as loosely bound L-LHCII, comprising only trimeric LHCII, and LHCII monomers.

Despite apparent similarity of the major protein complexes in the lpBN gel, distinct differences between *P. abies* and Arabidopsis were found in thylakoid protein complex composition. *Picea abies* completely lacked the M-LHCII band (Fig. 5A), which in Arabidopsis is composed of the minor antenna proteins LHCB4 and LHCB6 as well as trimeric LHCII (Bassi and Dainese, 1992). The PSI–NDH megacomplexes were also absent from *P. abies* thylakoids (Fig. 5B), consistent with the loss of all plastid encoded subunits of the NDH-1 complex from *P. abies* (Nystedt *et al.*, 2013). Notably, a large chlorophyll-free protein complex was also visible in *P. abies* (Fig. 5A), which was identified as ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo).

Protein subunit composition of major thylakoid complexes in Picea abies

In order to study the protein composition of the thylakoid complexes described above, individual lanes containing the complexes in lpBN-PAGE (first dimension) were solubilized in Laemmli buffer (Järvi *et al.*, 2011) and subjected to separation of the protein subunits of each complex by SDS-PAGE (second dimension). After SYPRO Ruby and silver staining, protein spots from the resulting two-dimensional protein map of *P. abies* thylakoid proteomes were identified by in-gel tryptic digestion followed by MS/MS analysis of the eluted peptides. The generated spectra were matched to the custom *P. abies* protein database (see Supplementary Table S4 for MS/MS identification and Supplementary Fig. S2 for spot numbering). Identities of the Arabidopsis proteins in the 2D map were assigned according to previous MS identifications (Aro *et al.*, 2005).

Protein spots in 2D maps often contain multiple proteins (Thiede *et al.*, 2013), the number of which can be identified by MS/MS depending on the sensitivity of the mass spectrometer used. A further level of complexity arises in the analysis of transmembrane proteins that tend to lack Lys or Arg residues, which are required for trypsin-mediated protein cleavage for generation of peptides suitable for the typical scan range used in MS/MS. Thus, the MS analysis of an extrinsic and a membrane-embedded protein co-migrating in the same spot would potentially yield a higher number of peptides for the extrinsic protein, even though this number may not reflect the true relative abundance of the two proteins in the spot. This is especially true for proteins present at low abundance, or those with few or no unique tryptic peptides and several peptides shared with paralogues, which is the case for the LHC proteins (Friso *et al.*, 2004). To consider a protein correctly identified, we applied (i) a requirement of at least two unique peptides identified with high or medium confidence and (ii) a combination of match-crossing rules: Mascot score equal to at least half of the score of the first hit in the same spot; similar position/co-migration to the Arabidopsis reference 2D map; and co-migration in the first dimension with other subunits from the same complex. In rare cases, a protein was considered present even though the Mascot score was below the threshold, as clearly indicated in Supplementary Table S4). These rules were not applied for the LHCII-containing spots, since the high number of homologues made it impossible to apply such strict rules; all the LHCII subunits identified are listed in Supplementary Table S4.

In general, the subunit pattern of thylakoid protein complexes in Arabidopsis and *P. abies* showed many similarities due to conserved PSII and PSI core complexes (Fig. 6), yet distinct differences in migration of lower molecular mass proteins was observed. A longer development step was applied to the silver staining of the lower molecular mass region of the second dimension protein gels (Fig. 7) in order to visualize low molecular mass proteins of *P. abies* (e.g. PSAH) without saturating the signal from more abundant proteins (e.g. LHCII).

In the PSII–LHCII sc of *P. abies*, the PSII subunits PsbB, PsbC, PSBO, PsbD, PsbA, PsbE, and PSBH as well as the LHCII subunits LHCB8, LHCB1, LHCB2, and LHCB5 were identified by MS. The LHCB protein pattern of *P. abies* PSII–LHCII sc was distinct from that in Arabidopsis. In agreement with the *in silico* analysis, the LHCB3 and LHCB6 protein spots visible in Arabidopsis were not identified in all *P. abies* PSII–LHCII sc and all other LHCII complexes. In the L-LHCII band of *P. abies*, LHCB1 and LHCB2 proteins were identified as predominant components of the complex, while a significant number of unique peptides from LHCB5 were also identified. Therefore, LHCB5 is potentially part of the L-LHCII in *P. abies*, which is also supported by the apparent molecular mass of LHCB5 in the PSII–LHCII sc (spot 8, Supplementary Fig. S2), which is compatible with that of the fastest migrating protein subunit of L-LHCII (spot 35, Supplementary Fig. S2).

The core PSI subunits PsaA and PsaB, as well as the small subunits PSAD, PSAF, PSAL, PSAE, PSAH, and PSAG were identified from the main PSI band. It is worth noting that many of the small PSI subunits (PSAD, PSAF, PSAL, PSAE, PSAG) formed a clearly different migration pattern, and thus these subunits are likely to have different molecular masses, with respect to the corresponding PSI subunits in Arabidopsis. LHCA1 and LHCA3 were identified in the major PSI complex of *P. abies*, along with considerably fewer peptides corresponding to LHCA4. Interestingly, a pigment–protein complex migrating in the first dimension between the PSI/PSII dm band and the PSII mm/Cyt-bf/ATP synthase band (Fig. 6) was found to contain the same PSI core subunits as the main PSI complex, except PSAH and PSAG, and was also devoid of the LHCA proteins (Fig. 7). PSAL in the smaller PSI complex could only be identified with one unique peptide (spot 23, Supplementary Table S4) compared with two unique peptides in the main PSI complex (spot 14, Supplementary Table S4). This smaller PSI complex was similar to the PSI* (PSI assembly intermediate) complex, previously reported in algae (Ozawa *et al.*, 2010), Arabidopsis (Suorsa *et al.*, 2015; Järvi *et al.*, 2016), and tobacco (Wittenberg *et al.*, 2017). In contrast to PSI* in tobacco (Wittenberg *et al.*, 2017), PSAF appeared to be part

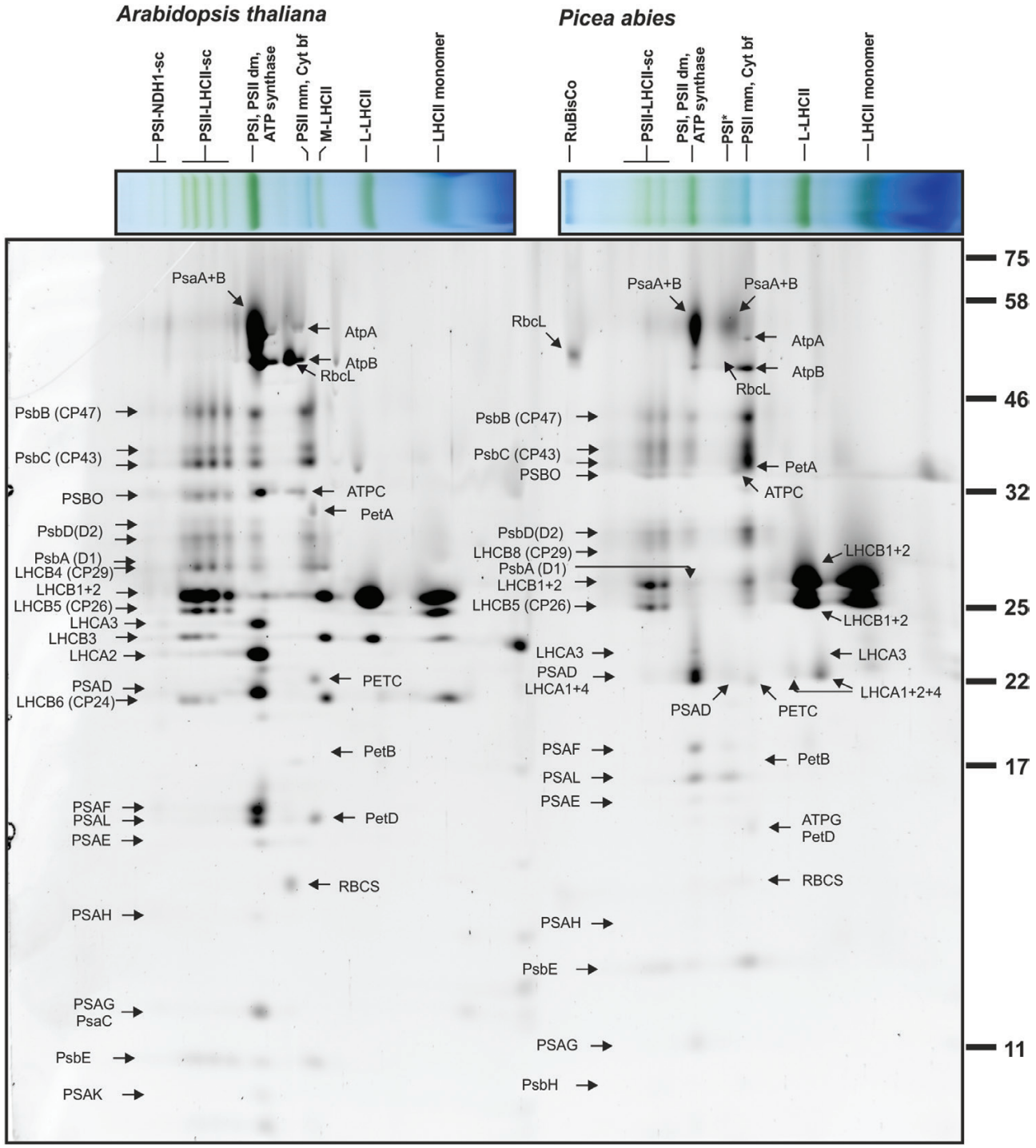

Fig. 6. Comparison of the subunit composition of the thylakoid protein complexes separated via 2D lpBN/SDS-PAGE from Arabidopsis and *P. abies* and subsequently stained with SYPRO Ruby. The molecular masses are indicated in kDa. Arabidopsis proteins are indicated according to Aro *et al.* (2005), and *P. abies* proteins are indicated according to the MS/MS identifications listed in Supplementary Table S4. The corresponding MS/MS files have been deposited in PRIDE repository (PXD010071).

of *P. abies* PSI*-like complex, although apparently at lower abundance with respect to PSAL. However, according to the MS analysis presented in Wittenberg *et al.* (2017), presence of low amounts of PSAF in PSI* in tobacco is likewise possible. Besides the main PSI and PSI*-like complex, free LHCI antenna proteins LHCA1–4 were also identified in two spots in the 2D maps, corresponding to the molecular mass region of L-LHCII in the first dimension separation.

Subunits PetA, PETC, PetB, and PetD of the Cyt-bf complex and subunits AtpA, AtpB, ATPC, ATPG, and AtpE of ATP synthase were identified (Figs 6, 7).

In both Arabidopsis and *P. abies*, the RuBisCo complex was visible in between the PSI, PSII dm, ATP synthase band and the PSII mm, Cyt-bf band after separation by lpBN-PAGE (Fig. 6). Interestingly, in *P. abies* a chlorophyll-free high molecular mass band in the first dimension was identified as RuBisCo.

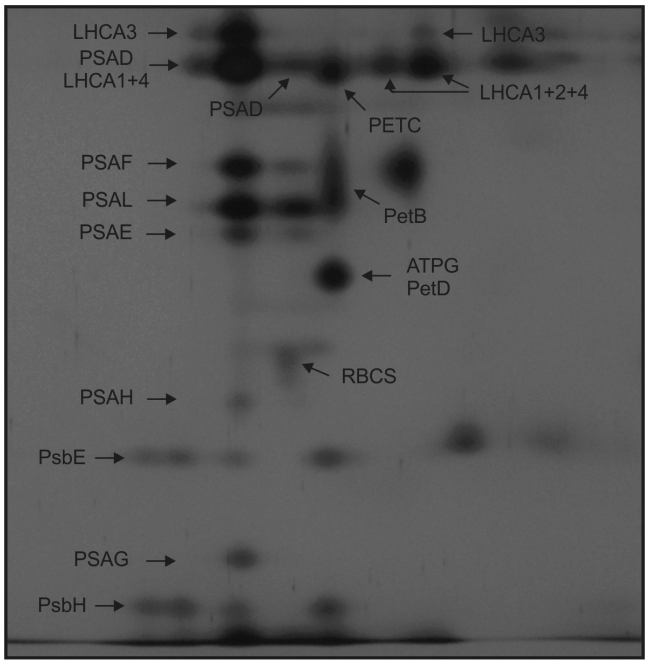

Fig. 7. Magnified image of low molecular mass region of *P. abies* 2D lpBN/SDS-PAGE after silver staining. *Picea abies* proteins are indicated according to the MS/MS identifications listed in Supplementary Table S4. The corresponding MS/MS files have been deposited in PRIDE repository (PXD010071).

Discussion

Recent major advances in sequencing of gymnosperm giga-genomes (De La Torre *et al.*, 2014) have created new possibilities for investigating the dominant species of boreal forests, in particular Pinaceae species like Norway spruce (*P. abies*, Nystedt *et al.*, 2013), white spruce (*Picea glauca*, Birol *et al.*, 2013; Warren *et al.*, 2015), loblolly pine (*Pinus taeda*, Neale *et al.*, 2014; Zimin *et al.*, 2014), and Douglas fir (*Pseudotsuga menziesii*, Neale *et al.*, 2017), by utilizing modern transcriptomic and proteomic methods.

This is particularly important for solving the survival strategies of evergreen species in harsh environmental conditions of boreal forests. To this end, we have elucidated the protein composition of the photosynthetic light-harvesting and energy conversion machinery in thylakoids of the evergreen gymnosperms, which are equipped with an extremely flexible acclimation capacity (for reviews see Öquist and Huner, 2003; Verhoeven, 2014). Although the main components of the photosynthetic machinery, i.e. PSII, PSI, Cyt-bf and ATP synthase, are generally highly conserved in all photosynthetic organisms, the LHC complexes of both photosystems have evolved differently during plant evolution (Büchel, 2015).

To compare the composition of the LHC antenna of *P. abies* with other land plants, we retrieved and classified all available LHC homologues from genomes and transcriptomes of 84 plant species, utilizing homology between selected diagnostic sequence regions. This analysis identified major differences in LHC composition between members of Pinaceae and other gymnosperms, as well as angiosperms (Fig. 3). In some cases, the apparent lack of LHC members might be attributed to incomplete coverage of genome sequences or low transcript concentration in transcriptome databases; however, the analysis of 84 species that were chosen to represent the broad spectrum of extant land plants has allowed us to identify reliable instances of variability in LHC families in plants.

Based on the *in silico* identification of LHC homologues in land plants, it became apparent that the gymnosperm families Pinaceae, Gnetaceae, and Welwitschiaceae (Pinales, Gnetales, and Welwitschiales) have a unique LHC protein composition. The lack of LHCB3 and LHCB6 from *P. abies* and 14 other Pinaceae species, as well as from two Gnetaceae and one Welwitschiaceae species, identified above, supports and expands on previous work (Kouřil *et al.*, 2016). In addition, LHCB4 was also lost from Pinaceae, Gnetidae, and Welwitschiaceae species, while LHCB8, which has a shorter C-terminus containing relatively high sequence variability, has been retained. Including the differences in the LHCA5 and LHCA6 proteins, our results collectively demonstrate that the LHC antenna composition of both photosystems in *P. abies* (and in Pinaceae, Gnetidae, and Welwitschiaceae in general) is distinct not only from angiosperms and evolutionarily early land plants, but also from other gymnosperms. Additionally, the analysis shows that LHCB8 is only present in the angiosperm clades Eurosids and Caryophylles, while it is universally present in all gymnosperms except the early gymnosperm families Cyadaceae and Zamiacae (Cycadales), suggesting independent evolution of 'LHCB8' in angiosperms and gymnosperms.

Distinct protein composition of the light harvesting antenna of PSII in Pinaceae

Strong support for the different occurrences of the LHCB proteins in *P. abies* was obtained from 2D lpBN/SDS-PAGE analysis (Fig. 6), where LHCB3, LHCB4, and LHCB6 spots were absent from *P. abies* thylakoid complexes, while LHCB8 was identified.

The absence of LHCB3 and LHCB6, as well as the presence of LHCB8 (instead of LHCB4) in *P. abies* thylakoids is likely a reason for the lack of the pentameric M-LHCII band in lpBN-gel (Fig. 5). In Arabidopsis, M-LHCII is composed of the M-trimer (i.e. a moderately bound LHCII trimer of the PSII–LHCII sc) together with LHCB4 and 6 (Bassi and Dainese, 1992). While in Arabidopsis the M-LHCII band is the result of thylakoid solubilization with mild detergents (Caffarri *et al.*, 2009; Rantala *et al.*, 2017), no corresponding stable M-LHCII was obtained from solubilized *P. abies* thylakoids. This is because in *P. abies* the M-trimer in the PSII–LHCII sc lacks both LHCB3 and the link provided by LHCB6 (Kouřil *et al.*, 2016), while LHCB4 is replaced with LHCB8 (Figs 5, 6). Instead, a detached 'M-trimer' was found to migrate together with trimeric L-LHCII in the lpBN gel. Interestingly, a significant number of unique peptides from LHCB5 were found in the L-LHCII band (spot 35, Supplementary Table S4), leaving open the possibility that LHCB5 is part of the LHCII trimers. Previous studies have shown that in Arabidopsis *lhcb2* antisense lines, which are deficient in the expression of both LHCB2 and LHCB1, LHCB5 forms homo-trimers and LHCB5/3 hetero-trimers in place of the LHCB1/2 subunits (Ruban *et al.*, 2003,

2006), strengthening the idea that LHCB5 could be part of a LHCII trimer in *P. abies*.

Evolutionary changes in the composition of LHCII subunits suggest that *P. abies* and other Pinaceae have tuned their LHCII composition to maintain a smaller PSII antenna in comparison with Arabidopsis. The differences between these species can be pinpointed to the structural role of the M-trimer and its binding partners in the PSII–LHCII sc, which in Arabidopsis comprise LHCB4/8 and LHCB6 (van Bezouwen *et al.*, 2017). In particular, new evidence is emerging on the special role of LHCB8, often referred to as LHCB4.3, which has already been shown to have a unique sequence and expression profile despite close homology with LHCB4 (Jansson, 1999; Klimmek *et al.*, 2006; Sawchuk *et al.*, 2008). Detailed studies of Arabidopsis knock-out mutants lacking LHCB members that are missing from *P. abies*, namely LHCB3 (Damkjær *et al.*, 2009), LHCB4 (de Bianchi *et al.*, 2011), and LHCB6 (Kovács *et al.*, 2006; de Bianchi *et al.*, 2008), revealed smaller PSII antenna size, suggesting that these individual LHC subunits affect each other's stability in the PSII–LHCII sc (Andersson *et al.*, 2001; Kovács *et al.*, 2006; de Bianchi *et al.*, 2008, 2011). Like in *P. abies* thylakoids (Fig. 5), the M-LHCII complex has been shown to be missing from Arabidopsis mutants lacking LHCB3, 4, or 6 (de Bianchi *et al.*, 2008, 2011; Betterle *et al.*, 2009; Caffarri *et al.*, 2009). Furthermore, LHCB8 cannot alone compensate for the lack of LHCB4 in Arabidopsis (de Bianchi *et al.*, 2011). Instead, LHCB8 seems to have a specific role in long-term high light acclimation, increasing at both transcript (Arabidopsis; Floris *et al.*, 2013) and protein (*Pisum sativum*; Albanese *et al.*, 2016, 2018) levels, unlike LHCB4.

In addition, decrease in PSII antenna size is also well described in wild type plants during long-term high-light acclimation (Anderson *et al.*, 1988; Walters and Horton, 1994; Murchie and Horton, 1997). At the level of single subunit stoichiometry, a strong decrease in LHCB3 and LHCB6 proteins was observed in Arabidopsis after high light acclimation, which consequently led to reduced amounts of the M-LHCII complex (Ballottari *et al.*, 2007; Kouřil *et al.*, 2013; Wientjes *et al.*, 2013; Bielczynski *et al.*, 2016). Thus, it is conceivable that *P. abies* and other Pinaceae have adapted their PSII–LHCII sc structure at the genome level to be similar to that observed in Arabidopsis after long-term high light acclimation. In this respect, the substitution of LHCB8 for LHCB4 in Pinaceae, Gnetaceae, and Welwitschiaceae gives a new focus on the role of the LHCB8 isoform. However, since only south-facing branches were sampled in the current study, further analyses are required to clarify whether this feature is identical in north-facing branches.

On the amino acid level, LHCB8 is distinguished from LHCB4 by its different C-terminal sequence (Fig. 4). According to the recently solved PSII–LHCII sc structure of Arabidopsis (van Bezouwen *et al.*, 2017), the LHCB4 C-terminus is in close contact with LHCB6 and likely involved in stabilizing the attachment of LHCB6 to the PSII–LHCII sc. The permanent replacement of LHCB4 with the C-terminally shorter LHCB8 in Pinaceae, Gnetaceae, and Welwitschiaceae may have evolved in parallel with the loss of LHCB6 in these species. LHCB8 lacks a 15-amino-acid motif that contains the mixed chlorophyll *a*/*b*-binding site b3 for chlorophyll b614 (Bassi *et al.*, 1999; Pan *et al.*, 2011), which is conserved in the C-terminus of LHCB4. Absence of this chlorophyll from the LHCB8 C-terminus could have ramifications for energy-transfer routes (Cinque *et al.*, 2000; Salverda *et al.*, 2003) and proposed quenching mechanisms (Ioannidis and Kotzabasis, 2015; van Bezouwen *et al.*, 2017) in the PSII–LHCII sc. Additionally, LHCB4 and LHCB8 have a longer N-terminal domain that is not found in other LHC subunits (see diagnostic region 2, Supplementary Table S2). This sequence overlaps with motif II of the 'knot' structure of paired PSII–LHCII sc, recently characterized in *Pisum sativum* (Albanese *et al.*, 2017). Thus, the absence of LHCB4 and the distinct presence of only LHCB8 with a unique motif II 'knot' structure could also be involved in unique photosynthetic adaptation in Pinaceae, Gnetaceae, and Welwitschiaceae.

Because of their close homology, it is likely that LHCB8 is a product of LHCB4 gene duplication, with new functions gained through modification of the C-terminus. Variability in the C-terminus of LHCB8 in gymnosperms, Caryophyllales and Eurosids (Fig. 4) and the lack of LHCB8 sequences found in modern species representing the common ancestor of these groups (Fig. 3; Supplementary Table S3) suggest that LHCB8 orthologues may have evolved through independent LHCB4 gene duplications. It is reasonable to assume that the presence of both LHCB4 and LHCB8 increased the functional flexibility of the PSII antenna system, allowing more efficient adaptation to changes in the environment. In contrast, the loss of LHCB3, 4, and 6 from species of Pinaceae, Gnetaceae, and Welwitschiaceae argues for a highly specialized PSII antenna system that may be more capable of long-term high light acclimation at the expense of functional flexibility. Comparative analysis of Pinaceae, Gnetaceae, and Welwitschiaceae species is needed to resolve how the above-described antenna modifications affect the collection, distribution, and dissipation of excitation energy and the general performance and tolerance of the photosynthetic machinery.

Unique characteristics of PSI in Pinaceae

In addition to the differences in the LHCII antenna proteins, differences in the PSI antenna in Pinaceae were also found. Pinaceae (as well as Gnetaceae and Welwitschiaceae) appear to have lost LHCA5 during evolution, but it is present in all other gymnosperms studied (Fig. 3). LHCA5 and LHCA6 are needed for stable formation of the PSI–NDH complex in angiosperms (Kouřil *et al.*, 2014; Otani *et al.*, 2018), and the absence of LHCA5 in Pinaceae is in line with the loss of the NDH-1 genes in *P. abies* (Nystedt *et al.*, 2013) and other members of Pinaceae, Gnetaceae, and Welwitschiaceae (Braukmann *et al.*, 2009). Loss of NDH-1 genes in land plants is not uncommon and has also been described in Geraniales, Alismantales, and Orchidaceae (Ruhlman *et al.*, 2015) as well as individual members of Salantales (Petersen *et al.*, 2015) and Cactaceae (Sanderson *et al.*, 2015).

In the 2D lpBN/SDS-PAGE of *P. abies* thylakoids (Fig. 6), a PSI*-like complex lacking PSAG and the LHCI antenna was identified. LHCI antenna proteins were found in two

additional spots, likely representing dimeric conformations of LHCA proteins separated from the PSI–LHCI complex (Fig. 7). In *P. abies*, the PSI*-like complex appears to be more abundant compared with other seed plants (Suorsa *et al.*, 2015; Järvi *et al.*, 2016; Wittenberg *et al.*, 2017), which could be related to an increased turnover of PSI in *P. abies* (for reviews see Schöttler *et al.*, 2011; Yang *et al.*, 2015). Because *P. abies*, like other members of Pinaceae, is an evergreen species with needles retained for multiple years, it is conceivable to expect a more constantly active assembly of PSI compared with deciduous plants, especially considering that, in contrast to PSII, no PSI repair cycle has been identified. On the other hand, the *P. abies* PSI*-like complex contained PSAF (Fig. 7), which is largely missing from the tobacco PSI* complex (Wittenberg *et al.*, 2017). PSAF forms the docking site for plastocyanin in PSI (Hippler *et al.*, 1989) and therefore the PSI*-like complex in *P. abies* could still be functional in electron transport. This would make the complex similar to PSI* found in algae, where PSAF is part of the PSI* complex and participates in electron transfer (Ozawa *et al.*, 2010). Yet, based on the 2D separation of thylakoid proteins (Fig. 5), the major difference between the mature PSI and PSI*-like complex is their antenna size. Since mature PSI has LHCI attached while PSI*-like complex has no LHCI bound, the benefit of having this PSI subpopulation in the thylakoid membrane of *P. abies* is not yet clear. One potential role for the separate PSI subpopulation in *P. abies* could be connected to its unique composition of photoprotective mechanisms among seed plants, such as the presence of flavodiiron proteins that accept electrons from PSI in a reaction that protects PSI. Such flavodiiron proteins are absent from all angiosperms (Allahverdiyeva *et al.*, 2015; Yamamoto *et al.*, 2016; Ilík *et al.*, 2017). Alternatively, PSI* may serve as a pool of excess PSI that can supplement photosynthetic activity under demanding conditions (Zhang and Scheller, 2004; Lima-Melo *et al.*, 2019).

Conclusions

Our comprehensive analysis of the photosynthetic proteins and complexes of *P. abies* has revealed a unique LHC composition of the photosynthetic apparatus in this and other members of Pinaceae (together with Gnetaceae and Welwitschiaceae) that markedly differ from other gymnosperms, angiosperms, and evolutionarily older land plants. We speculate that the unique LHC composition is related to different regulation of light harvesting, suggesting that the model describing regulation of photosynthetic light reactions in angiosperms cannot be simply superimposed onto Pinaceae. This study also provides high-quality databases and specialized experimental tools that will facilitate further in-depth photosynthetic characterizations of the economically important conifers.

Supplementary data

Supplementary data are available at *JXB* online.

Dataset S1. *Picea abies* thylakoid protein database v1.0.

Dataset S2. Merged *Picea abies* protein database.

Fig. S1. Effect of increasing concentrations of β-DM on thylakoid solubilization of *Picea abies.*

Fig. S2. MS/MS spot numbering of 2D lpBN/SDS-PAGE from *Picea abies*.

Table S1. Databases for LHC identification.

Table S2. Validation of LHC diagnostic regions.

Table S3. Overview of LHC proteins in land plants.

Table S4. List of proteins identified by MS/MS.

Acknowledgements

This project was funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 675006 and the Academy of Finland (project no. 307335 and 303757). Special thanks go to Roberta Croce for helpful discussions about LHCB8. The authors thank Virpi Paakkarinen, Jyrki Rasmus, and Ville Käpylä for their excellent technical assistance.

Data availability

The MS proteomics data of this study have been deposited with the ProteomeXchange Consortium (PXD010071; http://proteomecentral.proteomexchange.org; accessed 15/04/2019) via the PRIDE partner repository.

References

Albanese P, Manfredi M, Meneghesso A, Marengo E, Saracco G, Barber J, Morosinotto T, Pagliano C. 2016. Dynamic reorganization of photosystem II supercomplexes in response to variations in light intensities. Biochimica et Biophysica Acta **1857**, 1651–1660.

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. The Plant Journal **96**, 786–800.

Albanese P, Melero R, Engel BD, *et al*. 2017. Pea PSII-LHCII supercomplexes form pairs by making connections across the stromal gap. Scientific Reports **7**, 10067.

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.

Allahverdiyeva Y, Isojärvi J, Zhang P, Aro EM. 2015. Cyanobacterial oxygenic photosynthesis is protected by flavodiiron proteins. Life **5**, 716–743.

Alric J, Johnson X. 2017. Alternative electron transport pathways in photosynthesis: a confluence of regulation. Current Opinion in Plant Biology **37**, 78–86.

Anderson J, Chow W, Goodchild D. 1988. Thylakoid membrane organisation in sun/shade acclimation. Australian Journal of Plant Physiology **15**, 11.

Andersson J, Walters RG, Horton P, Jansson S. 2001. Antisense inhibition of the photosynthetic antenna proteins CP29 and CP26: implications for the mechanism of protective energy dissipation. The Plant Cell **13**, 1193–1204.

Armbruster U, Correa Galvis V, Kunz HH, Strand DD. 2017. The regulation of the chloroplast proton motive force plays a key role for photosynthesis in fluctuating light. Current Opinion in Plant Biology **37**, 56–62.

Aro EM, Suorsa M, Rokka A, Allahverdiyeva Y, Paakkarinen V, Saleem A, Battchikova N, Rintamäki E. 2005. Dynamics of photosystem II: a proteomic approach to thylakoid protein complexes. Journal of Experimental Botany **56**, 347–356.

Ballottari M, Dall'Osto L, Morosinotto T, Bassi R. 2007. Contrasting behavior of higher plant photosystem I and II antenna systems during acclimation. The Journal of Biological Chemistry **282**, 8947–8958.

Bassi R, Croce R, Cugini D, Sandonà D. 1999. Mutational analysis of a higher plant antenna protein provides identification of chromophores bound into multiple sites. Proceedings of the National Academy of Sciences, USA **96**, 10056–10061.

Bassi R, Dainese P. 1992. A supramolecular light-harvesting complex from chloroplast photosystem-II membranes. European Journal of Biochemistry **204**, 317–326.

Betterle N, Ballottari M, Zorzan S, de Bianchi S, Cazzaniga S, Dall'osto L, Morosinotto T, Bassi R. 2009. Light-induced dissociation of an antenna hetero-oligomer is needed for non-photochemical quenching induction. The Journal of Biological Chemistry **284**, 15255–15266.

Bielczynski LW, Schansker G, Croce R. 2016. Effect of light acclimation on the organization of photosystem II super- and sub-complexes in *Arabidopsis thaliana*. Frontiers in Plant Science **7**, 105.

Birol I, Raymond A, Jackman SD, *et al*. 2013. Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics **29**, 1492–1497.

Blum H, Beier H, Gross HJ. 1987. Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis **8**, 93–99.

Braukmann TW, Kuzmina M, Stefanović S. 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. Current Genetics **55**, 323–337.

Büchel C. 2015. Evolution and function of light harvesting proteins. Journal of Plant Physiology **172**, 62–75.

Caffarri S, Kouril R, Kereïche S, Boekema EJ, Croce R. 2009. Functional architecture of higher plant photosystem II supercomplexes. The EMBO Journal **28**, 3052–3063.

Christenhusz MJM, Reveal JL, Farjon A, Gardner MF, Mill RR, Chase MW. 2011. A new classification and linear sequence of extant gymnosperms. Phytotaxa **19**, 55–70.

Cinque G, Croce R, Holzwarth A, Bassi R. 2000. Energy transfer among CP29 chlorophylls: calculated Förster rates and experimental transient absorption at room temperature. Biophysical Journal **79**, 1706–1717.

Clarke JT, Warnock RCM, Donoghue PCJ. 2011. Establishing a timescale for plant evolution. New Phytologist **192**, 266–301.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Research **14**, 1188–1190.

Damkjær JT, Kereïche S, Johnson MP, Kovacs L, Kiss AZ, Boekema EJ, Ruban AV, Horton P, Jansson S. 2009. The photosystem II light-harvesting protein Lhcb3 affects the macrostructure of photosystem II and the rate of state transitions in *Arabidopsis*. The Plant Cell **21**, 3245–3256.

de Bianchi S, Betterle N, Kouril 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. The Plant Cell **23**, 2659–2679.

de Bianchi S, Dall'Osto L, Tognon G, Morosinotto T, Bassi R. 2008. Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits and the electron transport rate in grana membranes of *Arabidopsis*. The Plant Cell **20**, 1012–1028.

De La Torre AR, Birol I, Bousquet J, *et al*. 2014. Insights into conifer giga-genomes. Plant Physiology **166**, 1724–32.

Delhomme N, Sundström G, Zamani N, *et al*. 2015. Serendipitous meta-transcriptomics: the fungal community of Norway spruce (*Picea abies*). PLoS ONE **10**, e0139080.

Demmig-Adams B, Stewart JJ, Adams WW 3rd. 2017. Environmental regulation of intrinsic photosynthetic capacity: an integrated view. Current Opinion in Plant Biology **37**, 34–41.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **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 Molecular Biology **82**, 147–154.

Friso G, Giacomelli L, Ytterberg AJ, Peltier JB, Rudella A, Sun Q, Wijk KJ. 2004. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. The Plant Cell **16**, 478–499.

Gauthier S, Bernier P, Kuuluvainen T, Shvidenko AZ, Schepaschenko DG. 2015. Boreal forest health and global change. Science **349**, 819–822.

Hippler M, Ratajczak R, Haehnel W. 1989. Identification of the plastocyanin binding subunit of photosystem I. FEBS Letters **250**, 280–284.

Ilík P, Pavlovič A, Kouřil R, Alboresi A, Morosinotto T, Allahverdiyeva Y, Aro EM, Yamamoto H, Shikanai T. 2017. Alternative electron transport mediated by flavodiiron proteins is operational in organisms from cyanobacteria up to gymnosperms. New Phytologist **214**, 967–972.

Ioannidis NE, Kotzabasis K. 2015. Could structural similarity of specific domains between animal globins and plant antenna proteins provide hints important for the photoprotection mechanism? Journal of Theoretical Biology **364**, 71–79.

Iwai M, Yokono M. 2017. Light-harvesting antenna complexes in the moss *Physcomitrella patens*: implications for the evolutionary transition from green algae to land plants. Current Opinion in Plant Biology **37**, 94–101.

Jansson S. 1999. A guide to the *Lhc* genes and their relatives in *Arabidopsis*. Trends in Plant Science **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. The Biochemical Journal **439**, 207–214.

Järvi S, Suorsa M, Tadini L, Ivanauskaite A, Rantala S, Allahverdiyeva Y, Leister D, Aro EM. 2016. Thylakoid-bound FtsH proteins facilitate proper biosynthesis of photosystem I. Plant Physiology **171**, 1333–1343.

Klimmek F, Sjödin A, Noutsos C, Leister D, Jansson S. 2006. Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. Plant Physiology **140**, 793–804.

Kouřil 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 Phytologist **210**, 808–814.

Kouřil R, Strouhal O, Nosek L, Lenobel R, Chamrád I, Boekema EJ, Šebela M, Ilík P. 2014. Structural characterization of a plant photosystem I and NAD(P)H dehydrogenase supercomplex. The Plant Journal **77**, 568–576.

Kouřil R, Wientjes E, Bultema JB, Croce R, Boekema EJ. 2013. High-light vs. low-light: effect of light acclimation on photosystem II composition and organization in *Arabidopsis thaliana*. Biochimica et Biophysica Acta **1827**, 411–419.

Kovács L, Damkjaer J, Kereïche 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. The Plant Cell **18**, 3106–3120.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular Biology and Evolution **33**, 1870–1874.

Lima-Melo Y, Gollan PJ, Tikkanen M, Silveira JAG, Aro E-M. 2019. Consequences of photosystem I damage and repair on photosynthesis and carbon utilisation in *Arabidopsis thaliana*. The Plant Journal, doi: 10.1111/tpj.14177.

Merry R, Jerrard J, Frebault J, Verhoeven A. 2017. A comparison of pine and spruce in recovery from winter stress; changes in recovery kinetics, and the abundance and phosphorylation status of photosynthetic proteins during winter. Tree Physiology **37**, 1239–1250.

Murchie EH, Horton P. 1997. Acclimation of photosynthesis to irradiance and spectral quality in British plant species: chlorophyll content, photosynthetic capacity and habitat preference. Plant, Cell and Environment **20**, 438–448.

Neale DB, McGuire PE, Wheeler NC, *et al*. 2017. The Douglas-fir genome sequence reveals specialization of the photosynthetic apparatus in Pinaceae. G3 **7**, 3157–3167.

Neale DB, Wegrzyn JL, Stevens KA, *et al*. 2014. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. Genome Biology **15**, R59.

Nystedt B, Street NR, Wetterbom A, *et al*. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature **497**, 579–584.

Öquist G, Huner NP. 2003. Photosynthesis of overwintering evergreen plants. Annual Review of Plant Biology **54**, 329–355.

Otani T, Kato Y, Shikanai T. 2018. Specific substitutions of light-harvesting complex I proteins associated with photosystem I are required for supercomplex formation with chloroplast NADH dehydrogenase-like complex. The Plant Journal **94**, 122–130.

Ottander C, Campbell D, Oquist G. 1995. Seasonal-changes in photosystem-II organization and pigment composition in *Pinus sylvestris*. Planta **197**, 176–183.

Ozawa S, Onishi T, Takahashi Y. 2010. Identification and characterization of an assembly intermediate subcomplex of photosystem I in the green alga *Chlamydomonas reinhardtii*. The Journal of Biological Chemistry **285**, 20072–20079.

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. Nature Structural & Molecular Biology **18**, 309–315.

Petersen G, Cuenca A, Seberg O. 2015. Plastome evolution in hemiparasitic mistletoes. Genome Biology and Evolution **7**, 2520–2532.

Porra RJ, Thompson WA, Kriedemann PE. 1989. Determination of accurate extinction coefficients and simultaneous-equations for assaying chlorophyll-a and chlorophyll-b extracted with 4 different solvents – verification of the concentration of chlorophyll standards by atomic-absorption spectroscopy. Biochimica et Biophysica Acta **975**, 384–394.

Rantala M, Tikkanen M, Aro EM. 2017. Proteomic characterization of hierarchical megacomplex formation in Arabidopsis thylakoid membrane. The Plant Journal **92**, 951–962.

Rochaix JD. 2014. Regulation and dynamics of the light-harvesting system. Annual Review of Plant Biology **65**, 287–309.

Ruban AV, Solovieva S, Lee PJ, *et al*. 2006. Plasticity in the composition of the light harvesting antenna of higher plants preserves structural integrity and biological function. The Journal of Biological Chemistry **281**, 14981–14990.

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.

Ruhlman TA, Chang WJ, Chen JJ, *et al*. 2015. NDH expression marks major transitions in plant evolution and reveals coordinate intracellular gene loss. BMC Plant Biology **15**, 100.

Salverda JM, Vengris M, Krueger BP, Scholes GD, Czarnoleski AR, Novoderezhkin V, van Amerongen H, van Grondelle R. 2003. Energy transfer in light-harvesting complexes LHCII and CP29 of spinach studied with three pulse echo peak shift and transient grating. Biophysical Journal **84**, 450–465.

Sanderson MJ, Copetti D, Búrquez A, *et al*. 2015. Exceptional reduction of the plastid genome of saguaro cactus (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. American Journal of Botany **102**, 1115–1127.

Sawchuk MG, Donner TJ, Head P, Scarpella E. 2008. Unique and overlapping expression patterns among members of photosynthesis-associated nuclear gene families in Arabidopsis. Plant Physiology **148**, 1908–1924.

Schneider TD, Stephens RM. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Research **18**, 6097–6100.

Schöttler MA, Albus CA, Bock R. 2011. Photosystem I: its biogenesis and function in higher plants. Journal of Plant Physiology **168**, 1452–1461.

Schöttler MA, Tóth SZ. 2014. Photosynthetic complex stoichiometry dynamics in higher plants: environmental acclimation and photosynthetic flux control. Frontiers in Plant Science **5**, 188.

Shorohova E, Kneeshaw D, Kuuluvainen T, Gauthier S. 2011. Variability and dynamics of old-growth forests in the circumboreal zone: Implications for conservation, restoration and management. Silva Fennica **45**, 785–806.

Suorsa M, Rantala M, Mamedov F, Lespinasse M, Trotta A, Grieco M, Vuorio E, Tikkanen M, Järvi S, Aro EM. 2015. Light acclimation involves dynamic re-organization of the pigment–protein megacomplexes in non-appressed thylakoid domains. The Plant Journal **84**, 360–373.

The Angiosperm Phylogeny Group, **Chase MW, Christenhusz MJM, *et al***. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society **181**, 1–20.

Thiede B, Koehler CJ, Strozynski M, Treumann A, Stein R, Zimny-Arndt U, Schmid M, Jungblut PR. 2013. High resolution quantitative proteomics of HeLa cells protein species using stable isotope labeling with amino acids in cell culture (SILAC), two-dimensional gel electrophoresis (2DE) and nano-liquid chromatograpohy coupled to an LTQ-Orbitrap mass spectrometer. Molecular & Cellular Proteomics **12**, 529–538.

Tikkanen M, Aro EM. 2014. Integrative regulatory network of plant thylakoid energy transduction. Trends in Plant Science **19**, 10–17.

Ueda M, Kuniyoshi T, Yamamoto H, Sugimoto K, Ishizaki K, Kohchi T, Nishimura Y, Shikanai T. 2012. Composition and physiological function of the chloroplast NADH dehydrogenase-like complex in *Marchantia polymorpha*. The Plant Journal **72**, 683–693.

van Bezouwen LS, Caffarri S, Kale RS, Kouřil R, Thunnissen AWH, Oostergetel GT, Boekema EJ. 2017. Subunit and chlorophyll organization of the plant photosystem II supercomplex. Nature Plants **3**, 17080.

Verhoeven AS. 2013. Recovery kinetics of photochemical efficiency in winter stressed conifers: the effects of growth light environment, extent of the season and species. Physiologia Plantarum **147**, 147–158.

Verhoeven A. 2014. Sustained energy dissipation in winter evergreens. New Phytologist **201**, 57–65.

Verhoeven AS, Adams WW, Demmig-Adams B. 1996. Close relationship between the state of the xanthophyll cycle pigments and photosystem II efficiency during recovery from winter stress. Physiologia Plantarum **96**, 567–576.

Verhoeven AS, Adams Iii WW, Demmig-Adams B. 1999. The xanthophyll cycle and acclimation of *Pinus ponderosa* and *Malva neglecta* to winter stress. Oecologia **118**, 277–287.

Verhoeven A, Osmolak A, Morales P, Crow J. 2009. Seasonal changes in abundance and phosphorylation status of photosynthetic proteins in eastern white pine and balsam fir. Tree Physiology **29**, 361–374.

Vizcaíno JA, Côté RG, Csordas A, *et al*. 2013. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Research **41**, D1063–D1069.

Walters RG, Horton P. 1994. Acclimation of *Arabidopsis thaliana* to the light environment: changes in composition of the photosynthetic apparatus. Planta **195**, 248–256.

Warren RL, Keeling CI, Yuen MM, *et al*. 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. The Plant Journal **83**, 189–212.

Wientjes E, van Amerongen H, Croce R. 2013. Quantum yield of charge separation in photosystem II: functional effect of changes in the antenna size upon light acclimation. The Journal of Physical Chemistry. B **117**, 11200–11208.

Wittenberg G, Järvi S, Hojka M, Tóth SZ, Meyer EH, Aro EM, Schöttler MA, Bock R. 2017. Identification and characterization of a stable intermediate in photosystem I assembly in tobacco. The Plant Journal **90**, 478–490.

Yamamoto H, Takahashi S, Badger MR, Shikanai T. 2016. Artificial remodelling of alternative electron flow by flavodiiron proteins in *Arabidopsis*. Nature Plants **2**, 16012.

Yang H, Liu J, Wen X, Lu C. 2015. Molecular mechanism of photosystem I assembly in oxygenic organisms. Biochimica et Biophysica Acta **1847**, 838–848.

Zhang S, Scheller HV. 2004. Light-harvesting complex II binds to several small subunits of photosystem I. The Journal of Biological Chemistry **279**, 3180–3187.

Zimin A, Stevens KA, Crepeau MW, *et al*. 2014. Sequencing and assembly of the 22-gb loblolly pine genome. Genetics **196**, 875–890.