HOKKAIDO UNIVERSITY

Title	PCoM-DB Update: A Protein Co-Migration Database for Photosynthetic Organisms
Author(s)	Takabayashi, Atsushi; Takabayashi, Saeka; Takahashi, Kaori et al.
Citation	Plant and Cell Physiology, 58(1), e10 https://doi.org/10.1093/pcp/pcw219
Issue Date	2017-01-06
Doc URL	https://hdl.handle.net/2115/68044
Rights	This is a pre-copyedited, author-produced PDF of an article accepted for publication in Plant and cell physiology following peer review. The version of record PCoM-DB Update : A Protein Co-Migration Database for Photosynthetic Organisms, Atsushi Takabayashi,Saeka Takabayashi, Kaori Takahashi, Mai Watanabe, Hiroko Uchida, Akio Murakami, Tomomichi Fujita, Masahiko Ikeuchi, Ayumi Tanaka, Plant Cell Physiol (2017) 58 (1): e10 is available online at: https://doi.org/10.1093/pcp/pcw219
Type	journal article
File Information	PCoM-DB_revised_manuscript_3.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Cover page

Title: PCoM-DB Update: a Protein Co-Migration Database for Photosynthetic Organisms

Running title: PCoM-DB Update

Corresponding Author: Dr. A. Takabayashi

Institute of Low Temperature Science, Hokkaido University, N19 W8 Kita-Ku, Sapporo 060-0819, Japan,

E-mail: takabayashi@pop.lowtem.hokudai.ac.jp

Tel: +81-75-706-5493; Fax: +81-75-706-5493

Subject Areas:

(4) proteins, enzymes and metabolism

(5) photosynthesis, respiration and bioenergetics

Number of black and white figures: 0

Number of color figures: 6

Number of tables: 2

Number of supplementary data: 9

Title: PCoM-DB Update: a Protein Co-Migration Database for photosynthetic organisms

Running title: PCoM-DB Update

Atsushi Takabayashi[1,2], Saeka Takabayashi[1,3], Kaori Takahashi[1], Mai Watanabe[4], Hiroko Uchida[5], Akio Murakami[5], Tomomichi Fujita[6], Masahiko Ikeuchi[4], and Ayumi Tanaka[1,2]

[1]Institute of Low Temperature Science, Hokkaido University, N19 W8 Kita-ku, Sapporo 060-0819, Japan

[2]CREST, JST, N19 W8 Kita-ku, Sapporo 060-0819, Japan

[3]Department of Public Health, Graduate School of Medicine Hokkaido University, N15 W7 Kita-ku, Sapporo 060-8638, Japan

[4]Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Tokyo 153-8902, Japan

[5]Kobe University Research Center for Inland Seas, Awaji 656-2401, Japan

[6]Faculty of Science, Hokkaido University, N10 W8 Kita-ku, Sapporo 060-0810, Japan

Abbreviations:

Affinity purification combined with mass spectrometry; AP-MS, blue-native; BN, liquid chromatography-tandem mass spectrometry; LC-MS/MS, protein co-migration database; PCoM-DB, protein-protein interaction; PPI, ribulose 1,5-bisphosphate carboxylase/oxygenase; RubisCO, yeast two-hybrid system; Y2H.

[Abstract]

The identification of protein complexes is important to understand protein structure and function, and the regulation of cellular processes. We used blue-native PAGE and tandem mass spectrometry to identify protein complexes systematically, and built a web database, the protein co-migration database (PCoM-DB, http://pcomdb.lowtem.hokudai.ac.jp/proteins/top), to provide prediction tools for protein complexes. PCoM-DB provides migration profiles for any given protein of interest, and allows users to compare them with migration profiles of other proteins, showing the oligomeric states of proteins and thus identifying potential interaction partners. The initial version of PCoM-DB (launched in January 2013) included protein complex data for *Synechocystis* whole cells and *Arabidopsis thaliana* thylakoid membranes. Herein, we report PCoM-DB version 2.0, which includes new data sets and

analytical tools. Additional data are included from whole cells of the pelagic marine picocyanobacterium *Prochlorococcus marinus*, the thermophilic cyanobacterium *Thermosynechococcus elongatus*, the unicellular green alga *Chlamydomonas reinhardtii*, and the bryophyte *Physcomitrella patens*. The Arabidopsis protein data now include data for intact mitochondria, intact chloroplasts, chloroplast stroma and chloroplast envelopes. The new tools comprise a multiple-protein search form and a heat map viewer for protein migration profiles. Users can compare migration profiles of a protein of interest among different organelles or compare migration profiles among different proteins within the same sample. For Arabidopsis proteins, users can compare migration profiles of a protein of interest with putative homologous proteins from non-Arabidopsis organisms. The updated PCoM-DB will help researchers find novel protein complexes and estimate their evolutionary changes in the green lineage.

Key words:

Protein complex database, protein–protein interaction, blue-native PAGE, photosynthetic organisms

[Introduction]

As more genome sequences have become available, mass spectrometry-based proteomics has become increasingly important in many organisms to understand the localization, expression and interactions of proteins. In particular, the importance of protein complexes is becoming more evident, because complex protein networks often underlie cellular biological processes.

To understand protein networks in plants, databases that store plant protein–protein interaction (PPI) information have been developed (Braun et al. 2013). The main sources of data for these databases are Arabidopsis large-scale PPI identification studies using the yeast two-hybrid system (Y2H) and affinity purification combined with mass spectroscopy (AP-MS) (Zhang et al. 2010, Braun et al. 2013). Y2H has been the most widely used method for studying plant PPIs because it is a high-throughput method that does not require plant transformation. However, Y2H has some drawbacks, such as a high error rate (Zhang et al. 2010, Braun et al. 2013, Wodak et al. 2013). In addition, Y2H detects only direct interactions between two proteins (binary interactions) that can be formed in yeast. By contrast, AP-MS can purify whole protein complexes, leading to the identification of additional protein complex components (Zhang et al. 2010, Wodak et al. 2013). However, false positives also affect AP-MS (Zhang et al. 2010). Another

concern is that AP-MS requires plant transformation, which is typically labor-intensive and time-consuming; thus, AP-MS is not a high-throughput method. This is probably the main reason why PPI data are derived from Y2H more frequently than from AP-MS. Note that comparisons of large-scale PPI data produced by different methods are required to identify *bona fide* interactions, given that most methods for finding protein complexes have technical limitations and can produce erroneous and biased data (von Mering et al. 2002, Wodak et al. 2013). In addition, large-scale identification studies have predominantly been limited to a few intensively studied model organisms, like Arabidopsis. Therefore, a database that stores large-scale PPI datasets produced by different methods, including Y2H and AP-MS, and in different species including Arabidopsis, will help to further our understanding of plant protein networks.

Recently, blue-native (BN)-PAGE has been used for large-scale purification of protein complexes because of its favorable characteristics: first, BN-PAGE is a high-resolution electrophoresis method that can separate protein complexes, ranging up to several thousand kDa, while maintaining their structural integrity and enzymatic activity (Wittig and Schagger 2008, 2009). Notably, BN-PAGE can separate not only soluble protein complexes, but also membrane protein complexes using mild detergents like dodecyl maltoside. Second, the oligomeric states of protein complexes are revealed

by BN-PAGE, which is useful to understand the dynamic regulation of protein complexes. Third, the separation of protein complexes by BN-PAGE does not require genetic transformation, which not only saves labor, but also makes it suitable for organisms that lack established genetic transformation methods. BN-PAGE has been used widely to study protein complexes in photosynthetic organisms, including cyanobacteria, algae and plants.

Protein complexes separated by BN-PAGE are typically further separated by 2D-SDS-PAGE so they can be systematically identified. The protein spots are then visualized by staining methods and identified by tandem mass spectrometry (LC-MS/MS) analysis. This workflow is effective for identifying protein complexes systematically, especially in plant organelles (for examples, see Aro et al. 2005, Peltier et al. 2006, Majeran et al. 2008, Klodmann et al. 2011, Senkler and Braun 2012, Behrens et al. 2013). Notably, the GelMap Portal is an online tool for the annotation of protein spots visualized on 2D-BN/SDS-PAGE gels (Senkler and Braun 2012). Data generated for various photosynthetic species and organelles by 2D-BN/SDS-PAGE can be accessed through the GelMap Portal, which is a widely used data source for understanding protein complexes.

Alternatively, BN-PAGE has been directly coupled with LC-MS/MS to analyze

protein complexes (Fandino et al. 2005, Helbig et al. 2009, Wessels et al. 2009, Remmerie et al. 2011, Heide et al. 2012). This method omits the 2D-SDS-PAGE after BN-PAGE, which saves time and labor, making it suitable for high-throughput analysis. In addition, it does not suffer from incomplete visualization of protein spots, because it avoids staining methods that have a limited dynamic range (Remmerie et al. 2011). For this method, a BN-gel is cut horizontally into gel slices and the proteins in each gel slice are identified by LC-MS/MS analysis. The proteins identified in one gel slice represent co-migrating proteins in BN-PAGE, suggesting that they are possible interaction partners. The protein migration profiles, often produced using a label-free semi-quantitative method from the LC-MS/MS data, are useful for finding protein complexes. Proteins with shared peaks in their migration profiles are possible interaction partners, because a peak in the migration profile corresponds to a protein band on BN-PAGE (Helbig et al. 2009, Wessels et al. 2009, Remmerie et al. 2011). This method has shown potential for identifying protein complexes; however, the number of biological samples used in previous reports has been limited. Furthermore, only one public database for data produced using this method, PCoM-DB (Takabayashi et al. 2013), has been released that can help researchers to find novel protein complexes.

The original version of PCoM-DB allowed researchers to find novel protein

complexes from *Arabidopsis thaliana* thylakoids and cyanobacteria *Synechocystis* sp. PCC6803 cells using data produced by BN-PAGE coupled directly with LC-MS/MS (Takabayashi et al. 2013). A unique feature of PCoM-DB is its search and comparison functions for protein migration profiles. Users can find and compare protein migration profiles, enabling them to discover unidentified protein complexes. Another unique feature of PCoM-DB is that comparisons of migration profiles between homologous proteins can be used to understand the evolutionary changes of Arabidopsis thylakoid protein complexes from their cyanobacterial origins. From this perspective, the addition of data from different photosynthetic organisms (e.g. algae and bryophytes) should provide more information to understand the evolutionary changes in protein complexes of the green lineage. The addition of data from Arabidopsis organelles is not only practical, as reliable isolation methods for organelles have been reported in this species, but also informative, because the isolation of organelles increases the detection sensitivity of organelle proteins by mass spectrometry. In fact, when we analyzed Arabidopsis whole leaf proteins, the number of identified proteins was poor (data not shown). Additionally, comparing the exponentially modified protein abundance index (emPAI) values of protein migration profiles from different organelles can help users to predict the localization of proteins of interest, because the emPAI value correlates with

the relative abundance of a protein (Ishihama et al. 2005).

Here, we report a major update of PCoM-DB. We have added data from *Physcomitrella patens* (bryophytes), *Chlamydomonas reinhardtii* (green algae), *Thermosynechococcus elongatus* BP-1 (cyanobacteria), and *Prochlorococcus marinus* (CCMP1986, CCMP1375, and CCMP2773) (cyanobacteria) cells. In addition, we have added data from Arabidopsis organelles, including intact chloroplasts, chloroplast stroma, the chloroplast envelope and mitochondria. We have also added functions to help users search and compare protein migration profiles. This update will help researchers to find unidentified protein complexes and to compare the evolutionary changes in protein complexes of the green lineage. Furthermore, as a case study, we examined the possibility that At3g07480 is a subunit of complex I in Arabidopsis mitochondria. At3g07480 has significant sequence similarity with NUOP3 and thus was suggested to be a subunit of complex I (Subrahmanian et al. 2016); however, there has been no evidence to confirm this. Initially, we observed that it co-migrated with known complex I subunits. We then determined that its mRNA was co-expressed with mRNAs of known complex I subunits, which strongly suggests that At3g07480 is a subunit of complex I.

[Results and Discussion]

The protein data stored in PCoM-DB

In this study, we report additional data for protein complexes newly analyzed by the BN-PAGE coupled with LC-MS/MS method for the following biological materials: the chloroplast envelope (Supplementary Table S1) and mitochondria (Supplementary Table S2) from *Arabidopsis thaliana*; whole cells of *Chlamydomonas reinhardtii* (Supplementary Table S3); whole cells of the pelagic marine picocyanobacterium *Prochlorococcus marinus* (for three strains: CCMP1986, CCMP1375 and CCMP2773) (Supplementary Tables S4-S6); and whole cells of the thermophilic cyanobacterium *Thermosynechococcus elongatus* strain BP-1 (Supplementary Table S7) (Table 1). BN-PAGE profiles of these samples are shown in Fig. 1. We added data from three different strains of *Prochlorococcus marinus* to allow comparisons among them; the three strains possess different amounts of divinyl chlorophyll *b* to adapt to various light conditions. Comparing their protein migration profiles could help reveal their light adaptation mechanisms (Mimuro et al. 2011, Hamada et al. 2015). In addition, we added data for the stroma and intact chloroplasts from *Arabidopsis thaliana* and whole cells of *Physcomitrella patens*, which were reported recently by Takabayashi et al. (2016). Furthermore, we included additional data

for intact chloroplasts from *Arabidopsis thaliana* (Supplementary Table S8) that were generated independently in a previous study (Takabayashi et al. 2016). All of the analyzed data can be accessed in the current version of PCoM-DB.

Web interface

The main function of PCoM-DB is as a protein migration profile viewer. The migration profile of a protein corresponds to its band pattern on a BN-PAGE gel (Fig. 2). Users can find the protein migration profile of a protein of interest easily using the search function in PCoM-DB, with keywords including “gene ID (locus name)”, “symbol”, and “description”, through a “single protein search” interface (Fig. 3A). For example, users can find the large subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) protein using “AtCg00490” (gene ID), RbcL (symbol), and “ribulose-bisphosphate carboxylases” (description). When using the description as a query for a search, any word contained in the description can be used, e.g. “ribulose” can be used to search for RbcL.

Because of the increased number of samples in the updated database, we have implemented a heat map viewer for migration profiles across different samples for a query protein (Figs. 2C and 3B). Each column in the heat map viewer represents a gel

slice from BN-PAGE and gel slices are ordered from left (top of the gel) to right (bottom of the gel). A red column shows that a query protein was identified in the corresponding gel slice of the BN-PAGE gel. The intensity of the red color reflects the relative protein accumulation level, which was estimated using the label-free semi-quantitative method emPAI. Users can compare the migration profile of a query protein across the samples using the viewer. It should be noted that users are required to select a sample to navigate to the protein migration profile viewer page (Fig. 2B). Thus, the comparison of the migration profiles of a protein of interest across different samples can only be performed in the heat map viewer (Fig. 3B). Additionally, the predicted molecular masses of protein complexes in the same position in the gel slices should be different, because the intervals (around 1 mm) and the total numbers (around 60 slices) of gel slices were significantly different among samples, although the heat maps can be useful to compare protein migration profiles across different samples.

In addition to the single protein search function, we have implemented a multiple-protein search function. Users can input multiple-protein locus IDs and select a sample (Fig. 3C). After searching the query items, heat maps of the protein migration profiles in the selected sample will appear (Fig. 3D). Users can compare migration profiles among proteins of interest in the selected sample, which can help users to judge

whether the proteins co-migrate. It should be noted that users must first select a protein to navigate to the page with the protein migration profile viewer.

As an alternative to the search interface, users can also select a protein of interest through the “Browse” interface. First, users select a sample taxon, such *Arabidopsis thaliana* or *Synechocystis* sp. PCC6803, from a drop-down list, and then select a sample name, such as thylakoid or whole cells, from a second drop-down list (Fig. 4A). Users can then select a protein of interest from the list of proteins detected in the selected sample (Fig. 4B). Users can also use a single protein search function from the results page that will only search within the sample selected through the Browse interface. Then, by clicking a protein locus ID, users can navigate to the page showing the protein migration profile viewer.

Through the protein migration profile viewer interface, users can compare profiles with other identified proteins (Fig. 5). As shown in Fig. 2, peaks in a protein migration profile should correspond to protein bands on a BN-PAGE gel. Thus, proteins that share a peak at the same position in the migration profile with the protein of interest are candidate protein interaction partners. To find putative interacting partner proteins, users can view the list of identified proteins in the gel slice that correspond to the peak on the migration profile by clicking the data point. Users can choose the proteins from

the protein list by clicking their checkboxes, and can compare protein migration profiles by clicking “reCharts” (Fig. 5). In cases where the total emPAI values of the compared proteins are quite different, users can compare relative protein migration profiles, which are normalized to the maximum emPAI value, by clicking “max100%”. Note that users can estimate the presence of a peak in the selected gel slice on the protein migration profile to find the presence of “P” in the peak column (Fig. 5).

An example scenario (Arabidopsis mitochondria complex I)

Plant complex I is a respiratory complex in the inner mitochondrial membrane. Complex I is a large, fragile membrane protein complex, and 2D-BN/SDS-PAGE has been useful to maintain its native form during separation (Klodmann and Braun 2011, Meyer 2012, Subrahmanian et al. 2016). Here, we have chosen Arabidopsis mitochondria complex I as an example of the utility of PCoM-DB.

First, to verify the enrichment of mitochondrial proteins in our mitochondria sample, we estimated the abundances of the identified proteins using emPAI values (Ishihama et al. 2005) (Supplementary Tables S2 and S9). Based on comparison with a previous mitochondrial proteome (Lee et al. 2013), we found that the total emPAI values of known mitochondrial proteins accounted for more than 90% of those of all the

identified proteins (Supplementary Table S9), showing that mitochondrial proteins were substantially enriched in the sample.

According to the emPAI-based protein migration profile, the known mitochondrial complex I subunits were identified mainly in gel slices 16 and 17, where the band for mitochondrial complex I is expected to migrate. According to the list of complex I subunits described in Subrahmanian et al. (2016), 36 subunits were identified in these gel slices (Table 2), although the remaining 14 subunits were not identified. We identified protein At3g07480 in both gel slices 16 and 17, in addition to known complex I subunits (Fig. 6). At3g07480 has been suggested to be a subunit of mitochondrial complex I because of its sequence similarity with NUOP3, but it has not been identified in previous Arabidopsis proteomic analyses (Subrahmanian et al. 2016). However, our data showed that At3g07480 co-migrated with complex I subunits (Fig. 6), supporting the notion that At3g07480 (NUOP3) is in fact a subunit of mitochondrial complex I (Subrahmanian et al. 2016). Note that the 14 unidentified subunits from Subrahmanian et al. (2016) were largely membrane-spanning proteins, whereas the identified subunits were mostly matrix and peripheral membrane subunits. We hypothesize that denaturation treatment of BN-PAGE gel strips with SDS and 2-mercaptoethanol before in-gel trypsin digestion might improve the sensitivity of LC-MS/MS for membrane

proteins, which is an avenue of future research we intend to study. Supporting the results from PCoM-DB, the mRNA expression profile of At3g07480 correlates significantly with those of the other complex I genes, according to the ATTED-II database (Aoki et al. 2016). Thus, a comparison of mRNA expression profiles for protein–protein interaction partner candidates using PCoM-DB may be useful to judge whether the proteins actually interact.

[Conclusion and Perspectives]

In this study, we have added large data sets to PCoM-DB. Together with the new functions, we expect that the updated version of PCoM-DB will help researchers to find novel protein complexes. In a case study, we showed that At3g07480 is probably a subunit of mitochondrial complex I using PCoM-DB. Future identification of protein complexes using PCoM-DB will demonstrate the effectiveness of BN-PAGE directly coupled with LC-MS/MS. In addition to the updates mentioned here, we plan to further improve the quality of PCoM-DB. First, we would like to enhance the sensitivity of protein identification. Identifying peaks in protein migration profiles is essential to find novel protein complexes, but these profiles require sufficient numbers of unique peptides to plot. Thus, increased sensitivity in identifying protein migration profile

peaks should enable us to find more novel protein complexes. We believe that a substantial number of novel protein complexes will be found if the sensitivity of protein identification is improved. Future improvements to mass spectrometry will also increase the sensitivity. For instance and as discussed above, the addition of SDS and 2-mercaptoethanol to BN-gel strips before the in-gel tryptic digestion for membrane proteins should be investigated. In addition, multiple fractionation or improved resolution of BN-PAGE might be effective to enhance the sensitivity for unidentified or poorly identified proteins. Second, the use of different detergents for the solubilization of membrane proteins could improve the stability of some protein complexes. One candidate detergent is digitonin, which can stably solubilize certain large and labile protein complexes, including PSI-PSII supercomplexes (Jarvi et al. 2011, Yokono et al. 2015), which have not been stably isolated using dodecyl maltoside. Third, we would like to add data for new species and samples (e.g. organelles) to PCoM-DB. Finally, we would like to introduce functions that help researchers to increase the accuracy of protein complex prediction using data from different biological sources. It should be noted that co-migration of proteins on BN-PAGE by itself does not always indicate an actual interaction. Different biological data sources, including protein–protein interaction data generated by different methods, mRNA expression profiles, and

phylogenetic profiles produced using genome data and mutant phenotypes, will therefore be useful to increase the accuracy of protein complex prediction. Our future efforts will be directed at improving the features of PCoM-DB as mentioned above.

[Materials and Methods]

Organisms and growth conditions

Arabidopsis thaliana (ecotype Columbia) plants were grown in soil for 6 weeks under a 10-h photoperiod of 70 μmol photons m^{-2} s^{-1} at 23 °C. *Chlamydomonas reinhardtii* (cc1931 [arg7]) was cultivated in Tris-Acetate-Phosphate (TAP) medium with 200 mM arginine under a 14-h photoperiod of 40 μmol photons m^{-2} s^{-1} at 24.5 °C (Kunugi et al. 2016). The growth conditions for the three strains of *Prochlorococcus marinus* (CCMP1986, CCMP1375 and CCMP2773) were as described previously (Hamada et al. 2015). The thermophilic cyanobacterium, *Thermosynechococcus elongatus* strain BP-1 (NIES-2133) was grown at 45 °C as described previously (Iwai et al. 2004).

BN-PAGE

Chloroplast envelopes and intact chloroplasts were prepared from rosette leaves according to the methods described by Salvi et al. (2008). Intact mitochondria

were isolated from Arabidopsis rosette leaves, according to Keech et al. (2005). The cells of *Chlamydomonas*, the three species of *Prochlorococcus*, and *Thermosynechococcus* were homogenized using 0.1-mm glass beads on a vortex mixer in ice-cold solubilization buffer (50 mM imidazole-HCl (pH 7.0), 20% glycerol, 5 mM 6-aminocaproic acid, and 1 mM EDTA-2Na). Genomic DNA was removed using Benzonase nuclease (Merck, Darmstadt, Germany).

The suspensions in the solubilization buffer were solubilized with 1% (w/v) n-dodecyl-α-D-maltoside (for envelopes, mitochondria and *Thermosynechococcus* cells) or n-dodecyl-β-D-maltoside (for chloroplasts, *Chlamydomonas* cells, and *Prochlorococcu*s cells) on ice for 5 min. After centrifugation at 18,800 ×g for 10 min at 4 °C, the supernatants were separated in 4–13% polyacrylamide gradient gels at 4 °C, as described previously (Takabayashi et al. 2013).

Nano-LC/MS/MS for proteomics

BN-PAGE gel strips were cut horizontally into approximately 60 pieces from the top to the dye front. All gel pieces were subjected to in-gel digestion with trypsin and analyzed by LC-MS/MS, according to a previously described method (Takabayashi et al. 2013).

Protein identification was performed using Mascot (Matrix Science) by

searching the protein annotation data from TAIR10 (The Arabidopsis Information Resource) (Lamesch et al. 2012) for Arabidopsis proteins, the Chlamydomonas v5.3 annotations from the Phytozome database (http://www.phytozome.net/) (Goodstein et al. 2012) for *Chlamydomonas* proteins, and the protein data from CyanoBase (http://genome.kazusa.or.jp/cyanobase/) (Fujisawa et al. 2014) for *Prochlorococcus* proteins. The search parameters used for Mascot were the same as those described previously (Takabayashi et al. 2013).

Similarity search

Homologs of Arabidopsis proteins in the genomes of the non-Arabidopsis eukaryotic organisms (Chlamydomonas v5.3 and Physcomitrella v1.6) were identified based on the Phytozome database. In addition, we used the BLASTP program (version 2.2) to find putative Arabidopsis proteins (TAIR10) with the highest sequence similarity to cyanobacterial proteins.

Data availability

All of the proteomic data analyzed here are available as Supplementary Tables S1–S8. The proteomic data from Arabidopsis intact chloroplasts, Arabidopsis stroma,

and *Physcomitrella* cells were reported previously in the supplementary tables of Takabayashi et al. (2016). The proteomic data of Arabidopsis thylakoid membranes and *Synechocystis* sp. PCC6803 cells were previously reported in the supplementary tables of Takabayashi et al. (2013).

Database implementation

PCoM-DB has been implemented in a web application framework, Ruby on Rails (http://rubyonrails.org/), with an SQLite database. Protein migration profiles are plotted using HighCharts, which is an open source JavaScript graphing library.

[Funding]

This work was supported by Grants-in-Aid for Young Scientists (23770035 and 26850034 to A. Takabayashi) and Scientific Research (23370013 to A. Murakami and 24370017 to A. Tanaka) from the Japan Society for the Promotion of Science. This work was also supported by research funds and cooperative research funds (H23-47, H24-51) from the Institute of Low Temperature Science, Hokkaido University.

[Disclosures]

No conflicts of interest declared.

[Acknowledgements]

We thank Dr. Ryosuke Kadoya and Mr. Masayoshi Kuwano for performing the

LC-MS/MS analysis.

[References]

Aoki, Y., Okamura, Y., Tadaka, S., Kinoshita, K. and Obayashi, T. (2016) ATTED-II in 2016: A Plant Coexpression Database Towards Lineage-Specific Coexpression. *Plant & cell physiology* 57: e5.

Aro, E.M., Suorsa, M., Rokka, A., Allahverdiyeva, Y., Paakkarinen, V., Saleem, A., et al. (2005) Dynamics of photosystem II: a proteomic approach to thylakoid protein complexes. *J Exp Bot* 56: 347-356.

Behrens, C., Blume, C., Senkler, M., Eubel, H., Peterhansel, C. and Braun, H.P. (2013) The 'protein complex proteome' of chloroplasts in Arabidopsis thaliana. *Journal of proteomics* 91: 73-83.

Braun, P., Aubourg, S., Van Leene, J., De Jaeger, G. and Lurin, C. (2013) Plant protein interactomes. *Annual review of plant biology* 64: 161-187.

Fandino, A.S., Rais, I., Vollmer, M., Elgass, H., Schagger, H. and Karas, M. (2005) LC-nanospray-MS/MS analysis of hydrophobic proteins from membrane protein complexes isolated by blue-native electrophoresis. *Journal of mass spectrometry : JMS* 40: 1223-1231.

Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiya-Kanegae, H., et al. (2014) CyanoBase and RhizoBase: databases of manually curated annotations

for cyanobacterial and rhizobial genomes. *Nucleic acids research* 42: D666-670.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* 40: D1178-1186.

Hamada, F., Murakami, A. and Akimoto, S. (2015) Comparative Analysis of Ultrafast Excitation Energy-Transfer Pathways in Three Strains of Divinyl Chlorophyll a/b-Containing Cyanobacterium, Prochlorococcus marinus. *The journal of physical chemistry. B* 119: 15593-15600.

Heide, H., Bleier, L., Steger, M., Ackermann, J., Drose, S., Schwamb, B., et al. (2012) Complexome Profiling Identifies TMEM126B as a Component of the Mitochondrial Complex I Assembly Complex. *Cell Metab* 16: 538-549.

Helbig, A.O., de Groot, M.J., van Gestel, R.A., Mohammed, S., de Hulster, E.A., Luttik, M.A., et al. (2009) A three-way proteomics strategy allows differential analysis of yeast mitochondrial membrane protein complexes under anaerobic and aerobic conditions. *Proteomics* 9: 4787-4798.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., et al. (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & cellular proteomics : MCP* 4: 1265-1272.

Iwai, M., Katoh, H., Katayama, M. and Ikeuchi, M. (2004) Improved genetic transformation of the thermophilic cyanobacterium, Thermosynechococcus elongatus BP-1. *Plant & cell physiology* 45: 171-175.

Jarvi, S., Suorsa, M., Paakkarinen, V. and Aro, E.M. (2011) Optimized native gel systems for separation of thylakoid protein complexes: novel super- and mega-complexes. *The Biochemical journal* 439: 207-214.

Keech, O., Dizengremel, P. and Gardestrom, P. (2005) Preparation of leaf mitochondria from Arabidopsis thaliana. *Physiologia Plantarum* 124: 403-409.

Klodmann, J. and Braun, H.P. (2011) Proteomic approach to characterize mitochondrial complex I from plants. *Phytochemistry* 72: 1071-1080.

Klodmann, J., Senkler, M., Rode, C. and Braun, H.P. (2011) Defining the protein complex proteome of plant mitochondria. *Plant physiology* 157: 587-598.

Kunugi, M., Satoh, S., Ihara, K., Shibata, K., Yamagishi, Y., Kogame, K., et al. (2016) Evolution of Green Plants Accompanied Changes in Light-Harvesting Systems. *Plant & cell physiology* 57: 1231-1243.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic acids research* 40: D1202-1210.

Lee, C.P., Taylor, N.L. and Millar, A.H. (2013) Recent advances in the composition and heterogeneity of the Arabidopsis mitochondrial proteome. *Front Plant Sci.* 4: 4.

Majeran, W., Zybailov, B., Ytterberg, A.J., Dunsmore, J., Sun, Q. and van Wijk, K.J. (2008) Consequences of C_4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. *Molecular & cellular proteomics : MCP* 7: 1609-1638.

Meyer, E.H. (2012) Proteomic investigations of complex I composition: how to define a subunit? *Front Plant Sci* 3: 106.

Mimuro, M., Murakami, A., Tomo, T., Tsuchiya, T., Watabe, K., Yokono, M., et al. (2011) Molecular environments of divinyl chlorophylls in Prochlorococcus and Synechocystis: differences in fluorescence properties with chlorophyll replacement. *Biochimica et biophysica acta* 1807: 471-481.

Peltier, J.B., Cai, Y., Sun, Q., Zabrouskov, V., Giacomelli, L., Rudella, A., et al. (2006) The oligomeric stromal proteome of Arabidopsis thaliana chloroplasts. *Molecular & Cellular Proteomics* 5: 114-133.

Remmerie, N., De Vijlder, T., Valkenborg, D., Laukens, K., Smets, K., Vreeken, J., et al. (2011) Unraveling tobacco BY-2 protein complexes with BN PAGE/LC-MS/MS and

clustering methods. *Journal of proteomics* 74: 1201-1217.

Salvi, D., Rolland, N., Joyard, J. and Ferro, M. (2008) Purification and proteomic analysis of chloroplasts and their sub-organellar compartments. *Methods Mol Biol* 432: 19-36.

Senkler, M. and Braun, H.P. (2012) Functional Annotation of 2D Protein Maps: The GelMap Portal. *Front Plant Sci* 3: 87.

Subrahmanian, N., Remacle, C. and Hamel, P.P. (2016) Plant mitochondrial Complex I composition and assembly: A review. *Biochimica et biophysica acta* 1857: 1001-1014.

Takabayashi, A., Kadoya, R., Kuwano, M., Kurihara, K., Ito, H., Tanaka, R., et al. (2013) Protein co-migration database (PCoM -DB) for Arabidopsis thylakoids and Synechocystis cells. *SpringerPlus* 2: 148.

Takabayashi, A., Niwata, A. and Tanaka, A. (2016) Direct interaction with ACR11 is necessary for post-transcriptional control of GLU1-encoded ferredoxin-dependent glutamate synthase in leaves. *Scientific reports* 6: 29668.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., et al. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417: 399-403.

Wessels, H.J.C.T., Vogel, R.O., van den Heuvel, L., Smeitink, J.A., Rodenburg, R.J., Nijtmans, L.G., et al. (2009) LC-MS/MS as an alternative for SDS-PAGE in blue native analysis of protein complexes. *Proteomics* 9: 4221-4228.

Wittig, I. and Schagger, H. (2008) Features and applications of blue-native and clear-native electrophoresis. *Proteomics* 8: 3974-3990.

Wittig, I. and Schagger, H. (2009) Native electrophoretic techniques to identify protein-protein interactions. *Proteomics* 9: 5214-5223.

Wodak, S.J., Vlasblom, J., Turinsky, A.L. and Pu, S. (2013) Protein-protein interaction networks: the puzzling riches. *Current opinion in structural biology* 23: 941-953.

Yokono, M., Takabayashi, A., Akimoto, S. and Tanaka, A. (2015) A megacomplex composed of both photosystem reaction centres in higher plants. *Nat Commun* 6.

Zhang, Y., Gao, P. and Yuan, J.S. (2010) Plant protein-protein interaction network and interactome. *Current genomics* 11: 40-46.

[Tables]

Table 1 Summary of the proteomic data stored in the updated version of PCoM-DB.

Organisms	Classifications	Samples	Identified proteins	Data sources
Arabidopsis thaliana	Vascular plants	chloroplast (1)	804	Takabayashi et al. (2016)
		chloroplast (2)	1,159	In this study
		envelope	309	In this study
		stroma	452	Takabayashi et al. (2016)
		thylakoid	244	Takabayashi et al. (2013)
		mitochondria	1,264	In this study
Physcomitrella patens	Bryophytes	whole cells	1,663	Takabayashi et al. (2016)
Chlamydomonas reinhardtii	Green algae	whole cells	1,115	In this study
Synechosystis sp. PCC6803	Cyanobacteria	whole cells	1,458	Takabayashi et al. (2013)
Prochlorococcus marinus CCMP1986		whole cells	761	In this study
Prochlorococcus marinus CCMP1375		whole cells	891	In this study
Prochlorococcus marinus CCMP2773		whole cells	892	In this study
Thermosynechococcus elongates BP-1		whole cells	1,001	In this study

Table 2 List of complex I subunits identified in this study.

Accession number	Name	Localization
AT5G08530	51 kDa	matrix
AT1G47260	CA2	mb peripheral
ATMG00070	Nad9/30 kDa	matrix
AT2G20360	39 kDa	matrix
AT5G52840	B13	matrix
AT4G16450	20.9 kDa/MNLL	mb
AT5G11770	PSST/20 kDa	matrix
AT5G63510	CAL1	mb peripheral
AT5G37510	75 kDa	matrix
AT3G48680	CAL2	mb peripheral
AT2G33220	B16.6/GRIM-19	mb
ATMG00510	Nad7/49 kDa	matrix
AT5G66510	CA3	mb peripheral
AT4G02580	24 kDa	matrix
AT1G19580	CA1	mb peripheral
AT2G42210	B14.7	mb
AT4G20150	NDU9/B14.5b	mb
AT5G47890	B8	matrix
AT5G67590	18 kDa	matrix
AT2G46540	B9	mb
AT3G12260	B14	mb
AT3G08610	MWFE	mb
AT1G14450	B12	mb
AT2G31490	NDU8/B15	mb
AT3G07480*	At3g07480*	unidentified
AT4G00585	NDU10/KFYI	mb peripheral
AT2G27730	P2/16 kDa	mb
AT3G03070	13 kDa	matrix
AT1G79010	TYKY/23 kDa	matrix
AT1G49140	PDSW	mb peripheral
AT5G08060	B14.5a	matrix

AT1G16700	TYKY/23 kDa	matrix
ATMG00285	Nad2	mb
ATMG00060	Nad5	mb
AT4G34700	B22	mb
AT2G42310	NDU12/ESSS	mb
AT3G18410	PDSW	mb peripheral
AT5G47570	ASHI	mb
AT2G47690	15 kDa	mb

Names and localizations of the subunits are according to Subrahmanian et al. 2016.

* At3g07480 had not been identified as a complex I subunit in Arabidopsis; however, a *Chlamydomonas* homolog (NUOP3) has been identified as a complex I subunit.

[Legends to Figures]

Fig. 1 Protein band patterns of the analyzed samples on BN-PAGE gels. The separation patterns of protein complexes solubilized by dodecyl maltoside and separated by BN-PAGE in samples of Arabidopsis intact chloroplasts, Arabidopsis envelopes, Arabidopsis mitochondria, *Chlamydomonas* cells, three strains of *Prochlorococcus* cells, and *Thermosynechococcus* cells, are shown. These BN-gels were cut into approximately 1-mm pieces and then subjected to LC-MS/MS analyses.

Fig. 2 The relationships among the band pattern on the BN-PAGE gel, the protein migration profile, and the heat map view of the protein migration profile. (A) The separation of the protein complexes in Arabidopsis intact chloroplasts. The RubisCO

protein band is indicated. (B) The protein migration profile of RubisCO in Arabidopsis chloroplasts. (C) The heat map view of the protein migration profile of RubisCO in Arabidopsis chloroplasts.

Fig. 3 The search interfaces and subsequent sample/protein selection interfaces in PCoM-DB. (A) The single protein search window interface. Users can input the gene locus ID, gene symbol and descriptions. (B) After searching a protein through the single protein search window, heat map views of the protein migration profiles will appear. Users can click a sample name to navigate to the protein migration profile viewer page. When searching for Arabidopsis proteins, the heat map views of the migration profiles of the putative homologous proteins in non-Arabidopsis organisms will be shown. (C) The multiple-protein search window interface. Users can input multiple gene locus IDs, although they must select a sample. (D) After searching a protein through the multiple-protein search window, the heat map views of the protein migration profiles in the selected sample will appear. Users can click a gene locus ID to move to the protein migration profile viewer page.

Fig. 4 The browse interface in PCoM-DB. (A) Users can select a sample from the

browse tab using a drop-down list. (B) After selecting a sample through the browse window, the list of the proteins that were identified in the sample will be shown. Users can select a protein from the list and click it to move to the protein migration profile viewer page. Users can search a protein from the selected sample using a search window.

Fig. 5 The migration profiles of Rubisco large (ATCG00490) and small (AT1G67090) subunit proteins through the protein migration profile viewer. The shared peaks in their migration profiles in gel slice 24 correspond to the band of a RubisCO protein complex separated by the BN-PAGE (Fig. 2).

Fig. 6 Comparison of the migration profile of AT3G07480 with the migration profiles of known mitochondrial complex I subunits (AT5G08530 and ATMG00070). All of the proteins have peaks in their migration profiles in gel slices 16 and 17, showing that they co-migrate in BN-PAGE.