



Title	Unique Peripheral Antennas in the Photosystems of the Streptophyte Alga <i>Mesostigma viride</i>
Author(s)	Aso, Michiki; Matsumae, Renon; Tanaka, Ayumi et al.
Citation	Plant and Cell Physiology, 62(3), 436-446 https://doi.org/10.1093/pcp/pcaa172
Issue Date	2021-01-08
Doc URL	https://hdl.handle.net/2115/83775
Rights	This is a pre-copyedited, author-produced version of an article accepted for publication in Plant and Cell Physiology following peer review. The version of record is available online at: https://academic.oup.com/pcp/article/62/3/436/6070176 , http://doi.org/10.1093/pcp/pcaa172 .
Type	journal article
File Information	Plant and Cell Physiology62(3)_436-446.pdf



1 **This is a “clear copy” of the revised manuscript. The content is exactly the same as**
2 **“marked copy” described above.**

3

4 **Cover page**

5

6 **Title:** Unique peripheral antennas in the photosystems of the streptophyte alga

7 *Mesostigma viride*.

8

9 **Running Title:** Unique Photosystems in *Mesostigma viride*

10

11 **Corresponding Author:** Atsushi Takabayashi

12 Institute of Low Temperature Science, Hokkaido University, N19 W8 Kita-Ku, Sapporo

13 060-0819, Japan

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

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

16

17 **Subject Area:** (5) photosynthesis

18 **Number of black and white figures:** 0

19 **Number of color figures:** 6

20 **Number of tables:** 0

21 **Type and number of supplementary materials:** 2 figures and 4 tables

22 **Title:** Unique peripheral antennas **in the photosystems of the streptophyte alga**

23 *Mesostigma viride*.

24

25 **Running Title:** The Unique Photosystems in *Mesostigma viride*

26

27 **Michiki Aso¹, Renon Matsumae¹, Ayumi Tanaka¹, Ryouichi Tanaka¹, Atsushi**

28 **Takabayashi¹**

29 ¹Institute of Low Temperature Science, Hokkaido University, N19 W8 Kita-ku,

30 Sapporo 060-0819, Japan

31

32 **Abbreviations:** CN, clear-native; α -DDM, α -dodecyl maltoside; Isoseq, isoform

33 sequencing; LHC, light-harvesting complex; PAGE, polyacrylamide gel electrophoresis

34

35

36

37 **[Abstract]**

38 Land plants evolved from a single group of streptophyte algae. One of the key factors
39 needed for adaptation to a land environment is the modification of the peripheral
40 antenna systems of photosystems. Here, the photosystems of *Mesostigma viride*, an
41 earliest-branched streptophyte alga, were analyzed to gain insight into their evolution.
42 Iso-seq and phylogenetic analyses of Light-Harvesting Complexes (LHCs) revealed that
43 *M. viride* possesses three algae-specific LHCs, including algae-type LHCA2, LHCA9,
44 and LHCP; while the streptophyte-specific LHCB6 was not identified. These data
45 suggest that the acquisition of LHCB6 and the loss of algae-type LHCs occurred after
46 the *M. viride* lineage branched off from other streptophytes. Clear-native (CN)-PAGE
47 resolved the photosynthetic complexes, including the PSI-PSII megacomplex, PSII-
48 LHCII, two PSI-LHCI-LHCII, PSI-LHCI, and the LHCII trimer. Results indicated that
49 the higher-molecular weight PSI-LHCI-LHCII likely had more LHCII than the lower-
50 molecular weight one, a unique feature of *M. viride* photosystems. CN-PAGE coupled
51 with mass spectrometry strongly suggested that the LHCP was bound to PSII-LHCII,
52 while the algae-type LHCA2 and LHCA9 were bound to PSI-LHCI, both of which are
53 different from those in land plants. Results of the present study strongly suggest that *M.*

54 *viride* photosystems possess unique features that were inherited from a common
55 ancestor of streptophyte and chlorophyte algae.

56

57 **[Introduction]**

58 Photosystems (PSs) in oxygenic photosynthetic organisms are composed of core and
59 peripheral antenna complexes. While the peripheral antenna complex in these organisms
60 is highly divergent, the core complex is highly conserved (Green and Durnford, 1996;
61 Neilson and Durnford, 2010). The diversity of the peripheral antenna system contributes
62 to the ability of photosynthetic organisms to adapt to different light environments due to
63 the multiple roles it plays in photosynthesis, including the harvesting of light energy and
64 transferring it to the core complex. The variety of light spectra available in different
65 environments is one of the driving forces behind the evolution of different peripheral
66 antenna systems in photosynthetic organisms. (Croce and Amerongen, 2014; Stomp et
67 al., 2007). The thermal dissipation of excess light energy, which occurs mainly in
68 peripheral antennas (Ruban, 2018), is also necessary to avoid photooxidative damage.
69 Therefore, photosynthetic organisms have also evolved a variety of thermal dissipation
70 mechanisms in response to different environments (Giovagnetti and Ruban, 2018; Goss
71 and Lepetit, 2015; Niyogi and Truong, 2013; Wobbe et al., 2016).

72 Notable differences in peripheral antenna systems have been reported, even
73 within the green plants of photosynthetic organisms, and especially for PSI (Pan et al.,
74 2020; Suga and Shen, 2020). In vascular plants, PSI retains four LHCI (Lhca1-Lhca4)
75 proteins as a peripheral antenna and transiently gains a few extra LHCII proteins
76 (Mazor et al., 2017; Pan et al., 2018; Qin et al., 2015). PSI in chlorophyte algae
77 typically have a greater number of LHCI proteins forming their peripheral antenna than
78 vascular plants. For example, PSI in *Chlamydomonas reinhardtii* has ten LHCI (two
79 Lhca1 and one Lhca2-Lhca9) proteins and one mobile LHCII trimer that can transiently
80 bind to the antenna (Su et al., 2019; Suga et al., 2019). A similar antenna structure has
81 also been reported in *Bryopsis corticulans* (Qin et al., 2019). Chlorophyte algae may
82 typically have a larger size PSI antenna than vascular plants as an adaptation to aquatic
83 environments where light intensity is weaker than in terrestrial environments (Suga and
84 Shen, 2020). In contrast, *Dunaliella salina* has a “mini-PSI” that does not possess a
85 second LHCI belt unlike other reported chlorophyte algae (Perez-Boerema et al. 2020).
86 Since *D. salina* has the ability to survive high salinity and high light environments
87 (Perez-Boerema et al. 2020), the unique PSI structure may represent an adaptation to an
88 extreme habitat. In the model moss plant *P. patens*, PSI can possess five (Pinnola et al.,
89 2018) or eight (Iwai et al., 2018) more LHCI proteins than occurs in PSI-LHCI in

90 vascular plants, plus one additional mobile LHCII trimer. The number of LHC proteins
91 in *P. patens* PSI can change in response its photosynthetic status (Iwai et al., 2015;
92 Pinnola et al., 2018). The changes in the PSI antenna size reflect the variable light
93 environment that is typically experienced by mosses, which is characterized by low
94 irradiance complemented by sunflecks (Pinnola et al., 2018). These reports reveal that
95 the peripheral antenna system of PSI I-in green plants exhibit large species-specific
96 differences that apparently reflect their adaptation to different light environments.
97 However, to the best of our knowledge, detailed investigations on the peripheral antenna
98 system of PSI in streptophyte algae have not been reported so far.

99 Significant diversity is also present in thermal dissipation mechanisms in green
100 plants. Thus far, PsbS- and LHCSR-dependent thermal dissipation mechanisms have
101 been reported (Niyogi and Truong, 2013). The LHCSR-dependent mechanism is the
102 primary process used for thermal dissipation in chlorophyte algae, although they
103 possess both PsbS and LHCSR proteins (Correa-Galvis et al., 2016; Tibiletti et al.,
104 2016). In contrast, vascular plants only possess PsbS-dependent thermal dissipation,
105 having lost the LHCSR protein during their evolution (Giovagnetti and Ruban, 2018;
106 Niyogi and Truong, 2013; Pinnola, 2019). Notably, both PsbS and LHCSR play an
107 essential role in thermal dissipation in the model moss plant, *P. patens* (Pinnola, 2019).

108 Phylogenetically, *P. patens* represents an early-branched land plant and possesses
109 thermal dissipation characteristics that are intermediate between chlorophyte algae and
110 vascular plants. In addition, the contribution of PsbS to the thermal dissipating capacity
111 in land plants is much higher than it is in chlorophyte algae. How thermal dissipation
112 mechanisms have changed over the course of evolution from ancestral green algae to
113 land plants, however, has not been fully elucidated. This can be mainly attributed to the
114 absence of a suitable streptophyte alga species that could serve as a model for
115 photosynthesis research.

116 *M. viride* is one of the earliest-branched freshwater streptophyte algae,
117 although its exact phylogenetic position in green plants is still under debate (Lemieux et
118 al., 2007, 2000; Li et al., 2020; Wang et al., 2020). Nevertheless, it is worth testing
119 whether this basally branching alga retains a feature present in a common ancestor of
120 streptophytes and chlorophytes. Therefore, in the present study, the photosystems of the
121 earliest-branched *M. viride* were characterized to gain insight into the structure of
122 photosystems of a common ancestor of streptophytes and chlorophytes. Results revealed
123 that considerable changes occurred in photosystems after the *M. viride* lineage branched
124 off from other streptophyte algae.

125

126 **[Results]**

127 **Separation of photosystems by ClearNative (CN)-PAGE**

128 Clear-Native (CN)-PAGE is a powerful technique that enables one to separate
129 protein complexes while retaining their structure. Here, the *M. viride* photosystems
130 were separated using amphipol-based CN-PAGE (Furukawa et al., 2019) after
131 solubilization with a mild detergent, dodecyl maltoside (α -DDM). As a result, the PSI-
132 PSII megacomplex, the PSII-LHCII supercomplexes, the PSI-LHCI-LHCII bands, the
133 PSI-LHCI, and the LHCII trimer were resolved (Fig. 1A). The identification of the
134 separated bands was accomplished using 2D-CN/SDS-PAGE followed by immunoblot
135 analysis (Fig. 1B) and silver-staining (Fig. 1C), as described in previous studies (Järvi et
136 al., 2011; Takabayashi et al., 2011). The identification of two PSI-LHCI-LHCII bands
137 were confirmed by further analysis described in a later section of this report. The overall
138 band profile (Fig. 1A) was similar to the profile for *P. patens* presented by Furukawa et
139 al. (2019), however, a substantial difference was evident for PSI-LHCI-LHCII. Two
140 PSI-LHCI-LHCII bands were found in *M. viride* (Fig. 1A), whereas only one PSI-
141 LHCI-LHCII band was found in the profile of *P. patens* presented by Furukawa et al.
142 (2019). The presence of two PSI-LHCI-LHCII bands appears to be a unique
143 characteristic of the PSs of *M. viride* since land plants also exhibit one PSI-LHCI-

144 LHCII band in BN-PAGE and CN-PAGE gels (Järvi et al., 2011; Pesaresi et al., 2009;
145 Pinnola et al., 2018).

146

147 **Iso-seq analysis to provide a protein database for identification of *M. viride* LHC**
148 **proteins by MS**

149 No information is available on whether commercially-available LHC antibodies react
150 with *M. viride* LHCs. Therefore, mass spectrometry (MS) analysis of the protein
151 complexes resolved in the CN-PAGE was conducted, as described in our previous
152 papers (Takabayashi et al., 2017, 2013), to elucidate the composition of the LHC
153 proteins in the resolved PSI and PSII supercomplexes. This approach can be used to
154 estimate the positions of the bands of protein complexes separated by CN-PAGE,
155 especially for the relatively high expressed proteins. The number of *M. viride*
156 photosynthetic protein sequences available in public databases was limited at the time
157 of our MS analysis. Therefore, Isoform sequencing (Iso-seq) analysis was used to obtain
158 a transcriptome using PacBio sequencing, which provides longer and more complete
159 sequence information relative to short-read sequencing platforms such as Illumina
160 (Zhao et al., 2019). Full-length cDNA sequences enabled us to estimate full-length

161 amino acid sequences with high reliability, a feature that is advantageous for protein
162 identification by MS.

163 The iso-seq analysis (see Materials and Methods) allowed us to identify nine
164 PSI core proteins, 11 PSII core proteins, and 14 LHC proteins (Table S1). The small
165 subunits of PSI and PSII were not identified because the TransDecoder software
166 identified ORFs that are at least 100 amino acids long using the default settings.
167 Notably, the rate of false positives drastically increased when a shorter minimum length
168 was used. Therefore, default settings were used to predict ORFs in our analysis to obtain
169 reliable results.

170

171 **Phylogenetic analysis of LHC proteins**

172 A BLAST query identified 14 putative *M. viride* LHC proteins (Table S1). The
173 annotation of the LHC proteins derived from the BLAST query, however, may not be
174 reliable as considerable sequence similarity has been reported among different LHC
175 proteins. Therefore, a phylogenetic tree of LHC sequences was constructed based on the
176 alignment shown in Fig. S1 and each LHC sequence of *M. viride* was annotated based
177 on the nomenclature of *A. thaliana* LHC proteins, except for LHCBMs (corresponding
178 to major LHCII (LHCB1, LHCB2, and LHCB3)) and three algae-specific LHCs (algae-

179 type LHCA2, LHCA9, and LHCP). Algae-type LHCA2 is conserved among
180 chlorophytes, however, it is not closely related to plant-type LHCA2 (Fig. 2). Therefore,
181 we annotated it as algae-type LHCA2 in this study to avoid any misrepresentation.
182 Based on the phylogenetic analysis, *M. viride* possesses LHCBMs, LHCB4, LHCB5,
183 LHCA1, two LHCA2, LHCA3, and three algae-specific LHCs as peripheral antenna
184 proteins. The *M. viride* LHC sequences identified by the Iso-seq analysis exhibited good
185 correspondence with previously reported *M. viride* LHCs (Koziol et al., 2007) (Fig. 2).
186 Some differences, however, were observed relative to the previous classification (Koziol
187 et al., 2007). These differences are likely due to the considerable sequence similarities
188 among LHCs, as the LHCs in green lineages rapidly diversified during the early
189 evolution of green algae. LHCs that were classified differently in the previous report
190 (Koziol et al., 2007) and our present study have been marked with an asterisk (Fig. 2).

191 The Iso-seq analysis did not detect an LHCB6 gene, even though LHCB6 is
192 highly conserved among streptophyte algae and land plants (Kouřil et al., 2016). This
193 result is consistent, however, with a previous study by Koziol et al. (2007) in which
194 LHCB6 sequences were also not identified in *M. viride*, and with the *M. viride* RNA-
195 seq data generated in the 1000 plant transcriptomes (1kP) project (Carpenter et al.,
196 2019; Leebens-Mack et al., 2019). In contrast, LHCP proteins were found in *M. viride*

197 (Fig.2), which is also consistent with the previous report by Koziol et al. (2007). LHCP
198 is a unique LHC protein found in prasinophyte algae, including *Ostreococcus tauri* and
199 *Micromonas pusilla*. Prasinophyte algae is a group of early-diverging chlorophyte algae,
200 which gave rise to core chlorophyte algae including *C. reinhardtii*. It is very notable that
201 *M. viride*, an early-divergent streptophyte alga, also encodes LHCP. Previous studies of
202 a model prasinophyte alga, *O. tauri*, reported that it contains LHCP as the peripheral
203 antenna of PSI-LHCI (Six et al., 2005; Swingley et al., 2010), although no detailed
204 studies have been conducted on LHCP-containing photosystems. A possible loss of
205 LHCB6 and the presence of LHCP in *M. viride* suggest that the LHC composition in *M.*
206 *viride* is likely inherited from a common ancestor of streptophyte and chlorophyte algae.

207

208 **The distribution of LHC proteins in PSI and PSII**

209 No studies have been conducted on the distribution of *M. viride* LHCs in PSI and PSII,
210 although they have been identified and classified (Koziol et al., 2007). Therefore, we
211 determined the protein composition of the separated photosystems using MS. As
212 expected, PSI, PSII, and LHC proteins were detected in the PSI-PSII megacomplex
213 band (Table S2). Also as expected, PSI and LHC proteins were detected in the PSI-
214 LHCI band (Table S3). Unexpected proteins were rarely detected in these two

215 photosystems, suggesting that contamination from the other protein complexes was
216 limited. The PSII-LHCII band also contained PSII and LHC proteins, although some
217 contamination from PSI subunits and the NDH-like complex subunits were observed to
218 some extent (Table S4).

219 Importantly, the bands representing protein complexes on electrophoresis gels
220 are often distorted by tailing at both ends. The protein complexes in the tailings are also
221 detected by high-resolution of MS, which is the main source of contamination in a band
222 of interest. To predict unknown protein complexes and to estimate the protein
223 compositions of known protein complexes, a protein migration profile was used after
224 the native-PAGE coupled with MS (Takabayashi et al., 2017, 2013). A protein migration
225 profile is a plot where the y-axis represents the amount of protein estimated from MS
226 data, while the x-axis represents the migration distance on a native-PAGE gel. A peak in
227 the protein migration profile has been demonstrated to correspond to the position of the
228 band (Helbig et al., 2009; Remmerie et al., 2011; Wessels et al., 2009). This approach
229 allows one to determine if the proteins in a band detected by MS are contaminants from
230 other protein complexes by verifying their peak positions in protein migration profiles
231 (Helbig et al., 2009; Müller et al., 2016; Takabayashi et al., 2017; Wessels et al., 2009).

232 In the present study, migration profiles were constructed for PSI, PSII, and
233 LHC proteins based on MS data. A normalized spectral abundance factor (NSAF)
234 method, which is a label-free quantification method (Zybaïlov et al., 2006), was used to
235 estimate the amount of protein from MS data. First, the migration profiles of PSI and
236 PSII were compared. The distance between PSII-LHCII and PSI-LHCI-LHCII on the
237 CN-PAGE gel was relatively close, and PSI-LHCI-LHCII bands contained considerable
238 amounts PSI-LHCI-LHCII proteins. Nevertheless, it was possible to isolate those bands
239 using their migration profiles (Fig. 3A). The PSII-LHCII peak was observed in the
240 position of gel slices 8 and 9 (Fig. 3A), where the band was distinctly visible in the CN-
241 PAGE gel (Fig. 1A). The PSI-LHCI-LHCII and the PSI-LHCII peaks were observed in
242 the positions of gel slices 12 to 15 (Fig. 1A), where those bands were also visible in the
243 CN-PAGE gel (Fig. 1A). These data indicate that it is possible to distinguish PSI-bound
244 LHCS and PSII-bound LHCS by comparing their migration profiles with the migration
245 profiles of PSI and PSII.

246 Comparing the migration profiles of LHCS with those of PSI and PSII resulted
247 in the classification of LHCS into two groups. One group included LHCA1, LHCA2,
248 LHCA3, algae LHCA2, and LHCA9 proteins, whose peaks overlapped with the peaks
249 of PSI-LHCI and PSI-LHCI-LHCII (Fig. 3B). The other group included LHCB4,

250 LHCb5, and LHCP proteins, whose peaks overlapped with the peak of PSII-LHCII
251 (Fig. 3C). These data strongly suggest that PSI-LHCI includes the former LHCs,
252 whereas PSII-LHCII includes the latter LHCs. Notably, the migration profile of
253 LHCbM proteins appeared to be different from the profiles of other LHCs. Its highest
254 peak corresponded with LHCII trimers, suggesting that LHCbM is the major
255 component of the LHCII trimer.

256

257 **Separation of PSI-LHCI-LHCII complexes by sucrose density gradient**

258 Two PSI-LHCI-LHCII bands were observed on CN-PAGE gels. A “two-step”
259 separation of those bands was performed to improve the purity of the bands observed on
260 CN-PAGE using sucrose density gradient centrifugation as the first “rough” separation
261 step prior to further separation by CN-PAGE. As a result, three bands (B1-B3)
262 containing photosynthetic pigments were identified after separation by sucrose density
263 gradient centrifugation (Fig. 4). After subsequent separation by CN-PAGE, PSI-LHCI-
264 LHCII, in addition to PSI-LHCI, were found to be present in the middle band (B2) (Fig.
265 4), while the bottom band (B3) primarily contained PSII-LHCII, and the upper band
266 (B1) contained LHCII trimer (Fig. 4).

267 The three PSI-containing bands present in the CN-PAGE gel were further
268 separated by SDS-PAGE after they were cut out of the CN-gel to compare their
269 composition. Two replicates for each PSI-LHCI-LHCII band were subjected to 2D-
270 SDS-PAGE (Fig. 5). Silver-staining of the resulting gels revealed the PsaA/PsaB
271 heterodimer, the PsaA and the PsaB monomers, LHC monomers (c.a. 20kDa), and the
272 small PSI subunits. A portion of PsaA and PsaB remained as a heterodimer and migrated
273 slowly in the gel despite the inclusion of SDS (Fig. 5). A comparison of the three CN-
274 PAGE bands containing PSI indicated that the stoichiometry of the PSI and LHCI
275 subunits were similar, while the amounts of LHCII were greatest in the top band and
276 lowest in the bottom band. These results indicate that variations in the size of the three
277 CN-PAGE bands containing PSI were attributed to differences in the number of LHCII
278 (mainly LHCBM) subunits bound to each PSI-LHCI complex. Importantly, the
279 identification of the LHCII band was based on the separation pattern of the B1 and B3
280 bands containing LHCII trimers by the subsequent 2D-SDS-PAGE (Fig. S2). These data
281 suggest that the larger PSI-LHCI-LHCII possesses more LHCII than the smaller one.

282

283 **[Discussion]**

284 **Unique peripheral antenna systems of *M. viride***

285 We classified the LHCBs in the peripheral antenna of PSI and PSII. LHCB4 and LHCB5
286 were identified as the peripheral antenna of PSII-LHCII, similar to land plants. In
287 contrast, LHCB6 was not identified in the Iso-seq analysis in this study. The absence of
288 LHCB6 is consistent with the reported absence of LHCB6 sequences in *M. viride* LHC
289 sequences (Koziol et al., 2007) and is also consistent with the *M. viride* RNA-seq data
290 generated in the 1KP project (Carpenter et al., 2019; Leebens-Mack et al., 2019).
291 Therefore, it is likely that *M. viride* does not possess LHCB6, although we do not
292 exclude the possibility that the expression level of LHCB6 is relatively low compared to
293 other LHCBs, and thus was undetected. LHCB6 is widely distributed among streptophyte
294 algae and land plants but has not been found in chlorophyte algae, including *C.*
295 *reinhardtii* (Grebe et al., 2019; Kouřil et al., 2016). The putative loss of LHCB6 in *M.*
296 *viride* suggests that LHCB6 was acquired during evolution after the divergence of *M.*
297 *viride*.

298 *M. viride* PSII possesses LHCP (Koziol et al., 2007), the main peripheral
299 antenna protein in the model prasinophyte alga, *O. tauri* (Six et al., 2005; Swingley et
300 al., 2010), whereas LHCP has not been found in land plants. The peaks in the migration
301 profile of LHCP corresponded with the PSI-PSII megacomplex and the PSII-LHCII
302 (Fig. 3C). These data suggest that LHCP binds to PSII-LHCII in *M. viride*, although

303 further biochemical studies will be required to confirm this possibility. It should be
304 noted that detection of the LHCP peak at the position of the PSI-PSII megacomplex
305 could be explained by assuming co-migration of the PSI-PSII megacomplex and PSII-
306 LHCII at the top part of CN-PAGE (Fig. 3C). Such co-migration of these complexes
307 was actually observed in *A. thaliana* (Yokono et al. 2019).

308 A previous report revealed that LHCP, which is conserved among algae in the
309 Mamiellophyceae, including *O. tauri* and *M. pusilla*, is a major LHC antenna protein in
310 PSI-LHCI in the model prasinophyte algae, *O. tauri* (Swingley et al., 2010). It has also
311 been suggested that LHCP functions as the peripheral antenna of PSII-LHCII, although
312 no direct evidence has been provided (Six et al., 2005; Swingley et al., 2010). Our data
313 support the idea that LHCP can function as the peripheral antenna of PSII-LHCII (Fig.
314 3). Collectively, the data suggest that ancestral green algae possessed LHCP proteins as
315 their peripheral antenna because the earlier-branched streptophyte *M. viride* and the
316 prasinophyte *O. tauri* possess LHCP proteins in their photosystems. The LHCP-
317 containing photosystem in *M. viride* was likely inherited from a common ancestor of
318 chlorophyte and streptophyte algae that has been lost during the evolution of
319 streptophyte algae to land plants. Considering that the loss of LHCP and the acquisition
320 of the LHCB6 seemed to occur concomitantly according to our phylogenetic analysis, it

321 is possible to hypothesize that LHCP in streptophyte PSII-LHCII have been replaced by
322 LHCb6 in land plants. Further investigation will be required, however, to confirm this
323 hypothesis.

324 In regards to the peripheral antenna of PSI, two PSI-LHCI-LHCII
325 supercomplexes were stably observed after their separation by CN-PAGE. LHCII was
326 more abundant in the higher molecular weight band than it was in the lower molecular
327 weight band, suggesting that the higher molecular PSI-LHCI-LHCII possesses at least
328 two LHCII trimers. It is not presently known if at least two PSI-bound LHCII trimers
329 change their binding to PSI in response to different light environments, i.e., if they are
330 involved in state transitions. Since the binding of LHCII trimer to PSI was involved in
331 the state transition according to previous studies in green algae and land plants (Drop et
332 al., 2014; Pesaresi et al., 2009; Pinnola et al., 2018, 2018; Pribil et al., 2010; Takahashi
333 et al., 2006), we hypothesize that the additional *M. viride* LHCII trimers are also
334 involved in the state transition. We do not exclude the possibility, however, that the
335 smaller form of PSI-LHCI-LHCII represents an artifact caused by the dissociation of
336 LHCII from the larger form. Further studies are required to confirm this supposition.

337 In addition, PSI-LHCI in *M. viride* possesses algae-type LHCA2 and LHCA9
338 proteins, which are conserved among chlorophytes including *C. reinhardtii* (Su et al.,

339 2019; Suga et al., 2019) and *Bryopsis corticulans* (Qin et al., 2019), but not conserved
340 among streptophytes (Fig. 2). The loss of these algae-specific LHCs must have occurred
341 after the *M. viride* lineage branched off from the other streptophyte algae, although
342 further studies will also be required to confirm this possibility.

343 We constructed a hypothetical model of the antenna structure of PSI-LHCI-
344 LHCII in *M. viride* using the reported structure of PSI-LHCI in chlorophyte algae and
345 land plants as references (Fig. 6). Based on their sequence similarities, the binding
346 manner of LHCA1, two LHCA2s, and LHCA3 proteins to the PSI core in *M. viride* may
347 be similar to the binding of LHCA1, LHCA2, LHCA3, and LHCA4 proteins to the PSI
348 core in vascular plants. Likewise, the binding manner of algae-type LHCA2 and
349 LHCA2 to PSI-LHCI may be similar to the binding mechanism that occurs in PSI-LHCI
350 in *C. reinhardtii* (Su et al., 2019; Suga et al., 2019) and *B. corticulans* (Qin et al., 2019).
351 The binding site of the second LHCII trimer to PSI-LHCI, however, is unknown, as a
352 PSI-LHCI with two LHCII trimers has not, to our knowledge, been reported. Similarly,
353 the binding site of LHCP to PSII-LHCII is also unknown as no structural studies on
354 LHCP-bound PSI-LHCI have been reported. Given the phylogenetic position of *M.*
355 *viride* and its unique antenna system in PSI and PSII, further structural analyses of *M.*
356 *viride* PSI-LHCI-LHCII are warranted and would significantly contribute to our

357 understanding of the changes that occurred in photosystems during the evolution of a
358 common ancestor of chlorophytes and streptophytes to land plants.

359

360 **[Materials and Methods]**

361 **Algal strain and culture conditions**

362 *M. viride* strain, NIES-296, was obtained from the National Institute for Environmental
363 Studies (NIES) (Ibaraki, Japan). The strain was cultured in 500 mL of C medium in a
364 1L Erlenmeyer flasks at 22°C under a 14-h photoperiod of 20 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$, as
365 previously described by (Kunugi et al., 2016).

366

367 **RNA extraction**

368 One liter of a *M. viride* culture was centrifuged at 10,000 x g for 10 min at 4 °C. The
369 pellet was resuspended in 0.5 mL RLT buffer (RNeasy Plant Mini Kit, Qiagen)
370 supplemented with 1% 2-mercaptoethanol. The suspension was then transferred to a 2-
371 ml vial containing 500 mg of glass beads (0.5 mm diameter). The vial was subjected to
372 10s disruption treatments using a Mini-Bead Beater (Merck, Germany) and the
373 suspension was then centrifuged at 21,600xg at 4 °C for 5 min. Supernatants from the

374 vials were used for total RNA isolation with a RNeasy Plant Mini Kit (Qiagen),
375 according to the manufacturer's instructions.

376

377 **Full-length isoform sequencing (Iso-seq) using PacBio data**

378 The integrity of extracted total RNA was assessed using an Agilent 2100 Bioanalyzer
379 (Agilent Technologies Inc). The Iso-seq library was prepared according to the protocol
380 described by Pacific Biosciences (PN 101-070-200) using a SMARTer PCR cDNA
381 Synthesis Kit (Clontech) and SMRTbell Template Prep Kit 1.0 SPv3 (Pacific
382 Biosciences). Sequencing was performed on a PacBio Sequel platform.

383 Raw sequence data were processed using SMRT Link v6.0 (Pacific
384 Biosciences) software and then further processed using IsoSeq3 (version 3.1) software
385 tools (<https://github.com/PacificBiosciences/IsoSeq>) with default parameters. Circular
386 consistency sequence (CCS) reads were generated from subread sequences. Full-length
387 cDNA reads were then selected by finding the 5' and 3' primers or polyA tail using the
388 lima tool in IsoSeq3. After trimming of the polyA tail and removal of the concatemer
389 sequences, 84,545 full-length cDNA reads were obtained. Finally, 7,209 high-quality
390 consensus full-length cDNA sequences were obtained after isoform clustering and
391 polishing.

392

393 **Estimation of full-length amino acid sequences using the Iso-seq data**

394 TransDecoder (<https://github.com/TransDecoder/TransDecoder/>) software was used to
395 identify full-length protein sequences in the 7,209 full-length cDNAs. Candidate open
396 reading frames (ORFs) with a minimum length of 100 amino acids were identified using
397 the TransDecoder.LongOrfs module. Candidate ORFs were validated by BLASTP
398 queries using an e-value cutoff of 10^{-5} against data protein database comprising proteins
399 of *Arabidopsis thaliana*, *Chara braunii*, *Chlamydomonas reinhardtii*, *Coccomyxa*
400 *subellipsoidea*, *Cyanidioschyzon merolae*, *Cyanophora paradoxa*, *Klebsormidium*
401 *flaccidum*, *Oryza sativa*, *Ostreococcus lucimarinus*, *Physcomitrella patens*, and
402 *Selaginella moellendorffii* within the Phytozome database (v12.1)(Goodstein et al.,
403 2012). After the blast queries, 5,829 protein sequences with significant blast hits were
404 obtained using the TransDecoder.Predict module. CD-HIT (Fu et al., 2012) software
405 was employed to remove redundant protein sequences using a cutoff value of 0.9. A
406 total of 3,198 protein sequences were obtained using CD-HIT software. These protein
407 sequences were used as a *M. viride* protein database for MS queries.

408

409 **Phylogenetic analysis of *M. viride* LHC proteins**

410 *M. viride* LHC proteins were identified by NCBI-BLASTP homology queries against
411 the *M. viride* protein sequence database described above. *A. thaliana* and *C. reinhardtii*
412 LHC proteins were used as query sequences in the BLASTP searches. The LHC
413 sequences of *Arabidopsis thaliana*, *Marchantia polymorpha*, *Klebsormidium flaccidum*,
414 *Ostreococcus lucimarinus*, and *Chlamydomonas reinhardtii* were obtained from the
415 Phytozome database (<https://phytozome.jgi.doe.gov/>). The LHCs of *Cyanidioschyzon*
416 *merolae* and *Pyropia yezoensis* were used as the outgroups in the phylogenetic tree
417 analysis. Amino acid sequences of LHCs were aligned using the MAFFT algorithm
418 (Kato and Standley, 2013). Trimming of the alignment was done by a ClipKIT
419 program (Steenwyk et al. 2020) with default parameters. A maximum likelihood (ML)
420 phylogenetic tree including *M. viride* LHCs was constructed using W-IQ-TREE
421 (Trifinopoulos et al., 2016) software under the best-fitting model (LG+F+I+G4).
422 Ultrafast bootstrap values (1,000 replicates) are shown below the branches.

423

424 **Thylakoid membrane preparation**

425 A pellet obtained from the centrifugation (10,000 x *g* for 10 min at 4 °C) of 1 L of a *M.*
426 *viride* culture was suspended in 2 mL BN- solubilization buffer (50 mM imidazole/HCl
427 (pH 7.0), 20% glycerol) with 1% Protease Inhibitor Cocktail for plant cell lysate

428 (Merck, Germany). The suspension was transferred to a 2-ml vial containing 500 mg of
429 glass beads (0.5 mm diameter). The vial was first immersed in liquid nitrogen and the
430 cells were subsequently subjected to three 10s disruption treatments using a Mini-Bead
431 Beater (Merck, Germany). The resulting sample was centrifuged at 21,600×g at 4 °C for
432 5 min, and the supernatant was suspended in BN- solubilization buffer. The suspension
433 was centrifuged at 200×g at 4 °C for 1 min, and the supernatant was centrifuged at
434 21,600×g at 4 °C for 5 min to obtain the thylakoid membranes.

435

436 **Clear-Native (CN)-PAGE**

437 CN-PAGE was performed as previously described (Furukawa et al. 2019). A linear 4–
438 13% gradient polyacrylamide gel was used as the separation gel, and a 3.5%
439 polyacrylamide gel was used as the sample gel. Thylakoid membranes were solubilized
440 in 1% α -DDM (α -dodecyl maltoside) at 4 °C for 1 min. After centrifugation at 21,600×g
441 at 4 °C for 5 min, the supernatant (roughly 5 μ g chlorophyll equivalent) was
442 supplemented with 1% Amphipol A8-35 and loaded onto CN-PAGE gels. An anode
443 buffer (25mM imidazole/HCl (pH 7.0)) and cathode buffer (50mM Tricine, 7.5mM
444 imidazole) were used for the electrophoresis.

445

446 **Two-dimensional (2D)-CN/SDS-PAGE**

447 CN-gel strips were soaked in denaturation buffer (1% SDS, 50mM DTT) for 30 min,
448 and the 2D-SDS-PAGE was performed using a 14% polyacrylamide gel with 4M urea.
449 The proteins on the 2D-CN/SDS-gel were visualized by silver-staining using a Pierce
450 Silver Stain kit (ThermoFisher Scientific, USA), according to the manufacturer's
451 instructions. Alternatively, the obtained 2D-gels were used in the immunoblot analyses.

452

453 **Immunoblot analysis**

454 Proteins from the 2D-CN/SDS-gel were transferred to a polyvinylidene fluoride
455 membrane (PolyScreen PVDF transfer membrane, PerkinElmer Life Sciences, MA,
456 USA). Protein detection using specific antibodies was performed using Western
457 Lightning Plus-ECL (PerkinElmer Life science, MA, USA) reagent. All antibodies,
458 except anti-PsbB (AS04 038)), were purchased from Agrisera (Vännäs, Sweden).

459

460 **Sucrose density gradient**

461 *M. viride* thylakoid membranes were solubilized in 1% α -DDM and then loaded onto a
462 continuous sucrose density gradient (0.3M sucrose to 1.3M sucrose in a buffer
463 containing 25mM MES-KOH (pH 6.5) and 0.1% GDN (GDN101, Anatrace)). The

464 samples were then subjected to ultracentrifugation at 72,000 x *g* using a S-65T rotor
465 (Hitachi-koki, Japan) for 15 h at 4 °C.

466

467 **LC-MS/MS analysis**

468 CN-PAGE gel strips were cut horizontally as shown in Fig.1A. All gel pieces were
469 subjected to in-gel digestion with trypsin and analyzed to identify their peptides by LC-
470 MS/MS using Orbitrap Elite mass spectrometry (Thermo Fisher Scientific, Waltham,
471 MA, USA) coupled with Thermo Easy-nLC (Thermo Fisher Scientific, Waltham, MA,
472 USA). Each sample was loaded onto a C18-reversed phase EASY-Column (0.1 mm ×
473 20 mm, 5 μm particle size, 120 Å pore size), before separation on a C18 Tip column
474 (75 μm × 120 mm; Nikkyo Technos, Tokyo, Japan). The samples were separated by a
475 gradient formed by solvent A (0.1% formic acid) and solvent B (acetonitrile in 0.1%
476 formic acid) at a flow rate of 300 nL/min. The gradient separation setting was as
477 follows: 0-1min, 0–5% B; 1-12 min, 5–35% B; 12–25 min, 35%–90% B; 90% B; 25–45
478 min.

479 Proteins in each sample were identified using SearchGUI (version 3.3.15)
480 software and an andromeda search engine (Barsnes and Vaudel, 2018). The 3,198 *M.*
481 *viride* protein sequences identified in the Iso-seq analysis were used for protein

482 identification. PeptideShaker (version 1.16.40) software with default settings was used
483 to validate the protein identifications (Vaudel et al., 2015).

484 The normalized spectral abundance factor (NASF) (Zybailov et al., 2006)
485 calculated by the PeptideShaker program was used as a label-free quantification method
486 based on the LC-MS/MS data to estimate the abundances of the identified proteins in
487 each gel slice. As previously described, protein migration profiles were then generated
488 by plotting the NASF values on the y-axis, and plotting the gel slice number on the x-
489 axis (Takabayashi et al., 2017, 2013).

490

491 **[Acknowledgement]**

492 **Funding**

493 This work was supported by JSPS (Japan Society for the Promotion of Science)

494 KAKENHI Grant numbers 16H06554 to R. Tanaka.

495

496 **Data availability**

497 Raw data of the Iso-seq analysis have been deposited to the DDBJ Sequence Read

498 Archive (DRA) (<https://www.ddbj.nig.ac.jp/dra/index.html>) under the accession number

499 PRJDB10366. In addition, the assembled full-length cDNA sequences of PSI, PSII, and

500 LHC sequences have been deposited to DDBJ/EMBL/GenBank as Transcriptome
501 Shotgun Assembly (TSA) data under the accession numbers from ICQU01000001 to
502 ICQU01000034.

503

504 **Disclosures**

505 Conflicts of interest: The authors declare no conflicts of interest.

506

507 **[References]**

508 Barsnes, H., and Vaudel, M. (2018) SearchGUI: A Highly Adaptable Common
509 Interface for Proteomics Search and de Novo Engines. *J Proteome Res.* 17: 2552–2555.

510 Carpenter, E.J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., et al. (2019)
511 Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant
512 transcriptomes initiative (1KP). *Gigascience.* 8.

513 Correa-Galvis, V., Redekop, P., Guan, K., Griess, A., Truong, T.B., Wakao, S., et al.
514 (2016) Photosystem II Subunit PsbS Is Involved in the Induction of LHCSR Protein-
515 dependent Energy Dissipation in *Chlamydomonas reinhardtii*. *J Biological Chem.* 291:
516 17478–87.

517 Croce, R., and Amerongen, V.H. (2014) Natural strategies for photosynthetic light
518 harvesting.

519 Drop, B., K.N., S.Y., Boekema, E.J., and Croce, R. (2014) Consequences of state
520 transitions on the structural and functional organization of Photosystem I in the green
521 alga *Chlamydomonas reinhardtii*. *Plant J.* 78: 181–191.

522 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012) CD-HIT: accelerated for clustering
523 the next-generation sequencing data. *Bioinformatics.* 28: 3150–3152.

524 Furukawa, R., Aso, M., Fujita, T., Akimoto, S., Tanaka, R., Tanaka, A., et al. (2019)
525 Formation of a PSI–PSII megacomplex containing LHCSR and PsbS in the moss
526 *Physcomitrella patens*. *J Plant Res.* 132: 867–880.

527 Giovagnetti, V., and Ruban, A.V. (2018) The evolution of the photoprotective antenna
528 proteins in oxygenic photosynthetic eukaryotes. *Biochem Soc T.* 46: 1263–1277.

529 Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., et al. (2012)
530 Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40:
531 D1178–D1186.

532 Goss, R., and Lepetit, B. (2015) Biodiversity of NPQ. *Journal of plant physiology*. 172:
533 13–32.

534 Grebe, S., Trotta, A., Bajwa, A.A., Suorsa, M., Gollan, P.J., Jansson, S., et al. (2019)
535 The unique photosynthetic apparatus of Pinaceae: analysis of photosynthetic complexes
536 in *Picea abies*. *Journal of experimental botany*. 70: 3211–3225.

537 Green, B.R., and Durnford, D.G. (1996) The chlorophyll-carotenoid proteins of
538 oxygenic photosynthesis. *Annu Rev Plant Phys*. 47: 685–714.

539 Helbig, A.O., Groot, M.J.L. de, Gestel, R.A. van, Mohammed, S., Hulster, E.A.F. de,
540 Luttik, M.A.H., et al. (2009) A three-way proteomics strategy allows differential
541 analysis of yeast mitochondrial membrane protein complexes under anaerobic and
542 aerobic conditions. *Proteomics*. 9: 4787–98.

543 Iwai, M., Grob, P., Iavarone, A.T., Nogales, E., and Niyogi, K.K. (2018) A unique
544 supramolecular organization of photosystem I in the moss *Physcomitrella patens*.
545 *Nature plants*. 4: 904–909.

546 Iwai, M., Yokono, M., Kono, M., Noguchi, K., Akimoto, S., and Nakano, A. (2015)
547 Light-harvesting complex Lhcb9 confers a green alga-type photosystem I supercomplex
548 to the moss *Physcomitrella patens*. *Nature plants*. 1: 14008.

549 Järvi, S., Suorsa, M., Paakkarinen, V., and Aro, E.-M. (2011) Optimized native gel
550 systems for separation of thylakoid protein complexes: novel super- and mega-
551 complexes. *Biochemical Journal*. 439: 207–214.

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

555 Koziol, A.G., Borza, T., Ishida, K.-I., Keeling, P., Lee, R.W., and Durnford, D.G.
556 (2007) Tracing the evolution of the light-harvesting antennae in chlorophyll *a/b*-
557 containing organisms. *Plant Physiol*. 143: 1802–1816.

558 Kunugi, M., Satoh, S., Ihara, K., Shibata, K., Yamagishi, Y., Kogame, K., et al. (2016)
559 Evolution of green plants accompanied changes in light-harvesting systems. *Plant Cell*
560 *Physiology*. 57: 1231–43.

561 Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., Deyholos, M.K., Gitzendanner,
562 M.A., Graham, S.W., et al. (2019) One thousand plant transcriptomes and
563 the phylogenomics of green plants. *Nature*. 574: 679–685.

564 Lemieux, C., Otis, C., and Turmel, M. (2000) Ancestral chloroplast genome in
565 *Mesostigma viride* reveals an early branch of green plant evolution. *Nature*. 403: 649–
566 652.

567 Lemieux, C., Otis, C., and Turmel, M. (2007) A clade uniting the green algae
568 *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the
569 Streptophyta in chloroplast genome-based phylogenies. *Bmc Biol*. 5: 2.

570 Li, L., Wang, S., Wang, H., Sahu, S.K., Marin, B., Li, H., et al. (2020) The genome of
571 *Prasinoderma coloniale* unveils the existence of a third phylum within green plants. *Nat*
572 *Ecol Evol*. 1–12.

573 Mazor, Y., Borovikova, A., Caspy, I., and Nelson, N. (2017) Structure of the plant
574 photosystem I supercomplex at 2.6 Å resolution. *Nat Plants*. 3: 17014.

575 Müller, C.S., Bildl, W., Haupt, A., Ellenrieder, L., Becker, T., Hunte, C., et al. (2016)
576 Cryo-slicing blue native-mass spectrometry (csBN-MS), a novel technology for high
577 resolution complexome profiling. *Mol Cell Proteomics*. 15: 669–681.

578 Neilson, J.A.D., and Durnford, D.G. (2010) Structural and functional diversification of
579 the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res*. 106: 57–
580 71.

581 Niyogi, K.K., and Truong, T.B. (2013) Evolution of flexible non-photochemical
582 quenching mechanisms that regulate light harvesting in oxygenic photosynthesis.
583 *Current opinion in plant biology*. 16: 307–14.

584 Pan, X., Cao, P., Su, X., Liu, Z., and Li, M. (2020) Structural analysis and comparison
585 of light-harvesting complexes I and II. *Biochimica Et Biophysica Acta Bba -*
586 *Bioenergetics*. 1861: 148038.

587 Pan, X., Ma, J., Su, X., Cao, P., Chang, W., Liu, Z., et al. (2018) Structure of the maize
588 photosystem I supercomplex with light-harvesting complexes I and II. *Science*. 360:
589 1109–1113.

590 Pesaresi, P., Hertle, A., Pribil, M., Kleine, T., Wagner, R., Strissel, H., et al. (2009)
591 Arabidopsis STN7 kinase provides a link between short- and long-term photosynthetic
592 acclimation. *Plant Cell Online*. 21: 2402–2423.

593

594 Perez-Boerema, A., Klaiman, D., Caspy, I., Netzer-EI, S., Amunts, A., and Nelson, N.
595 (2020) Structure of a minimal photosystem I from the green alga *Dunaliella salina*.
596 *Nature plants*. 6: 321-327.

597

598 Pinnola, A. (2019) The rise and fall of light-harvesting complex stress-related proteins
599 as photoprotection agents during evolution. *J Exp Bot*. 70: 5527–5535.

600 Pinnola, A., Alboresi, A., Nosek, L., Semchonok, D., Rameez, A., Trotta, A., et al.
601 (2018) A LHCB9-dependent photosystem I megacomplex induced under low light in
602 *Physcomitrella patens*. *Nature plants*. 4: 910–919.

603 Pribil, M., Pesaresi, P., Hertle, A., Barbato, R., and Leister, D. (2010) Role of plastid
604 protein phosphatase TAP38 in LHCII dephosphorylation and thylakoid electron flow.
605 *Plos Biol*. 8: e1000288.

606 Qin, X., Pi, X., Wang, W., Han, G., Zhu, L., Liu, M., et al. (2019) Structure of a green
607 algal photosystem I in complex with a large number of light-harvesting complex I
608 subunits. *Nat Plants*. 5: 263–272.

609 Qin, X., Suga, M., Kuang, T., and Shen, J.-R. (2015) Structural basis for energy transfer
610 pathways in the plant PSI-LHCI supercomplex. *Science*. 348: 989–995.

611 Remmerie, N., Vijlder, T.D., Valkenburg, D., Laukens, K., Smets, K., Vreeken, J., et al.
612 (2011) Unraveling tobacco BY-2 protein complexes with BN PAGE/LC–MS/MS and
613 clustering methods. *J Proteomics*. 74: 1201–1217.

614 Ruban, A.V. (2018) Light harvesting control in plants. *Febs Lett*. 592: 3030–3039.

615 Six, C., Worden, A.Z., Rodríguez, F., Moreau, H., and Partensky, F. (2005) New
616 insights into the nature and phylogeny of prasinophyte antenna proteins: *Ostreococcus*
617 *tauri*, a Case Study. *Mol Biol Evol*. 22: 2217–2230.

618

619 Steenwyk, J.L., Buida 3rd, T.J., Li, Y., Shen, XX., Rokas, A. (2020) ClipKIT: A
620 multiple sequence alignment trimming software for accurate phylogenomic inference.
621 *PLoS Biol*. 18: e3001007.

622 Stomp, M., Huisman, J., Stal, L.J., and Matthijs, H.C.P. (2007) Colorful niches of
623 phototrophic microorganisms shaped by vibrations of the water molecule. *Isme J.* 1:
624 271–282.

625 Su, X., Ma, J., Pan, X., Zhao, X., Chang, W., Liu, Z., et al. (2019) Antenna arrangement
626 and energy transfer pathways of a green algal photosystem-I-LHCI supercomplex.
627 *Nature Plants.* 5: 273–281.

628 Suga, M., Ozawa, S.-I., Yoshida-Motomura, K., Akita, F., Miyazaki, N., and Takahashi,
629 Y. (2019) Structure of the green algal photosystem I supercomplex with a decameric
630 light-harvesting complex I. *Nature Plants.* 5: 626–636.

631 Suga, M., and Shen, J.-R. (2020) Structural variations of photosystem I-antenna
632 supercomplex in response to adaptations to different light environments. *Current*
633 *Opinion in Structural Biology.* 63: 10–17.

634 Swingley, W.D., Iwai, M., Chen, Y., Ozawa, S., Takizawa, K., Takahashi, Y., et al.
635 (2010) Characterization of photosystem I antenna proteins in the prasinophyte
636 *Ostreococcus tauri*. *Biochimica Et Biophysica Acta Bba - Bioenergetics.* 1797: 1458–
637 1464.

638 Takabayashi, A., Kadoya, R., Kuwano, M., Kurihara, K., Ito, H., Tanaka, R., et al.
639 (2013) Protein co-migration database (PCoM -DB) for Arabidopsis thylakoids and
640 Synechocystis cells. *Springerplus*. 2: 148.

641 Takabayashi, A., Kurihara, K., Kuwano, M., Kasahara, Y., Tanaka, R., and Tanaka, A.
642 (2011) The oligomeric states of the photosystems and the light-harvesting complexes in
643 the Chl *b*-less mutant. *Plant Cell Physiol*. 52: 2103–2114.

644 Takabayashi, A., Takabayashi, S., Takahashi, K., Watanabe, M., Uchida, H., Murakami,
645 A., et al. (2017) PCoM-DB update: a protein co-migration database for photosynthetic
646 organisms. *Plant Cell Physiol*. 58: e10–e10.

647 Takahashi, H., Iwai, M., Takahashi, Y., and Minagawa, J. (2006) Identification of the
648 mobile light-harvesting complex II polypeptides for state transitions in *Chlamydomonas*
649 *reinhardtii*. *Proc National Acad Sci*. 103: 477–482.

650 Tibiletti, T., Auroy, P., Peltier, G., and Caffarri, S. (2016) *Chlamydomonas reinhardtii*
651 PsbS protein is functional and accumulates rapidly and transiently under high light.
652 *Plant Physiol*. 171: 2717–2730.

653 Vaudel, M., Burkhart, J.M., Zahedi, R.P., Oveland, E., Berven, F.S., Sickmann, A., et
654 al. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat*
655 *Biotechnol.* 33: 22–24.

656 Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y., et al. (2020) Genomes of early-
657 diverging streptophyte algae shed light on plant terrestrialization. *Nat Plants.* 6: 95–106.

658 Wessels, H.J.C.T., Vogel, R.O., Heuvel, L. van den, Smeitink, J.A., Rodenburg, R.J.,
659 Nijtmans, L.G., et al. (2009) LC-MS/MS as an alternative for SDS-PAGE in blue native
660 analysis of protein complexes. *Proteomics.* 9: 4221–8.

661 Wobbe, L., Bassi, R., and Kruse, O. (2016) Multi-level light capture control in plants
662 and green algae. *Trends Plant Sci.* 21: 55–68.

663 Yokono, M., Takabayashi, A., Kishimoto, J., Fujita, T., Iwai, M., Murakami, A., et al.
664 (2019) The PSI–PSII megacomplex in green plants. *Plant Cell Physiol.* 60: 1098–1108.

665 Zhao, L., Zhang, H., Kohnen, M.V., Prasad, K.V.S.K., Gu, L., and Reddy, A.S.N.
666 (2019) Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq
667 and Nanopore-based direct RNA sequencing. *Frontiers Genetics.* 10: 253.

668 Zybaylov, B., Mosley, A.L., Sardu, M.E., Coleman, M.K., Florens, L., and Washburn,
669 M.P. (2006) Statistical analysis of membrane proteome expression changes in
670 *Saccharomyces cerevisiae*. *J Proteome Res.* 5: 2339–2347.

671

672 **[Figure legends]**

673 **Figure 1. Separation of photosynthetic complexes by CN-PAGE**

674 Separation of *M. viride* thylakoid membrane protein solubilized in 1% α -DDM by
675 amphipol A8-35-based CN-PAGE (A). The number in parentheses represents the
676 number of the corresponding gel slice which was further subjected to mass spectrometry
677 analysis (see Fig. 1A). Immunoblot analysis of PsaB (PSI), and PsbB (PSII) proteins
678 after their separation by 2D-CN/SDS-PAGE (B). Anti-PsaB and anti-PsbB antibodies
679 were purchased from Agrisera. 2D- CN/SDS-PAGE of *M. viride* thylakoid membrane
680 protein complexes visualized by silver-staining (C).

681

682 **Figure 2. A phylogenetic tree of *M. viride* LHC proteins**

683 A maximum likelihood phylogenetic tree of *M. viride* LHC proteins. Sequences
684 highlighted in yellow are clades containing the LHC from *M. viride*. The *M. viride* LHC

685 genes that are surrounded by a box represent the LHC genes identified in the Iso-seq
686 analysis within the present study. LHCs that were classified differently between the
687 previous report (Koziol et al., 2007) and our present study are marked with an asterisk.
688

689 **Figure 3. Comparison of LHC migration profiles in PSI and PSII**

690 *M. viride* thylakoid membranes were solubilized in α -DDM and separated by CN-
691 PAGE. After separation, gel regions containing the resolved photosystems were
692 horizontally cut into approximately 1 mm slices. The number shown in parentheses in
693 Fig. 1A represents the number of the corresponding gel slice. (A) Migration profiles of
694 PSI and PSII. The X-axis indicates the number of the gel slice as indicated on the CN-
695 PAGE gel (Fig. 1A). The Y-axis indicates the relative protein abundance estimated using
696 a label-free quantification method on the data obtained in the MS analysis. The PSII
697 peak shown in gel slices 8 and 9 likely correspond to the PSII-LHCII band on the CN-
698 PAGE gel, while the PSI peaks shown in gel slices 11 to 15 likely correspond to the
699 PSI-LHCI-LHCII band and the PSI-LHCI band. (B) Migration profiles of LHCA1,
700 LHCA2, LHCA3, algae-type LHCA2, and LHCA9 proteins. Their highest peaks
701 correspond to the PSI-LHCI-LHCII and PSI-LHCI bands. The protein data of the two
702 algae-type LHCA2 protein were combined to construct the migration profile of the

703 algae-type LHCA2, as the two LHCA2 protein sequences were similar. (C) Migration
704 profiles of LHCB4, LHCB5, and LHCBM proteins. The highest peaks in the LHCB4
705 and the LHCB5 migration profiles correspond with the PSII-LHCII band, while the
706 highest peak in the LHCBM migration profile corresponds to the LHCII trimer. All of
707 the MS data obtained for the LHCBM proteins were combined due their sequence
708 similarities.

709

710 **Figure 4. Separation of PSI-LHCI-LHCII supercomplexes using sucrose density**
711 **gradient centrifugation followed by CN-PAGE.**

712 (A) Separation of *M. viride* thylakoid membrane protein complexes solubilized in 1% α -
713 DDM by sucrose density gradient centrifugation (0.3M-1.3M). The middle band (B2)
714 contained the PSI-LHCI and the PSI-LHCI-LHCII, while the bottom band (B3)
715 primarily contained PSII-LHCII and the upper band (B3) contained LHCII trimer. (B)
716 Further separation of the PSI-enriched fraction (B2) in the sucrose gradient by CN-
717 PAGE.

718

719 **Figure 5. Separation of two PSI-LHCI-LHCII bands and a PSI-LHCI band by 2D-**
720 **SDS-PAGE**

721 Two PSI-LHCI-LHCII bands and a PSI-LHCI band were separated by CN-PAGE,
722 followed by sucrose density gradient centrifugation (Fig. 4), and were then subjected to
723 the 2D-SDS-PAGE. Protein bands were visualized by silver-staining.

724

725 **Figure 6. A hypothetical model of evolutionary changes in the PSI peripheral**
726 **antenna system of green plants.**

727 During evolution, green plants have diverged into two groups: streptophytes and
728 chlorophytes. Land plants are thought to have diverged from one group of freshwater
729 streptophytes, while core chlorophytes, including *C. reinhardtii*, diverged from seawater
730 chlorophytes (prasinophytes). The structure of PSI and its peripheral antennas have been
731 previously reported in vascular plants (Qin et al. 2015; Mazor et al. 2017; Pan et al.,
732 2018), moss plants (Iwai et al., 2018; Pinnola et al., 2018), and core chlorophytes,
733 including *C. reinhardtii* (Su et al., 2019; Suga et al., 2019) and *D. salina* (Perez-
734 Boerema et al. 2020). Notably, the “minimal” PSI-LHCI in *D. salina* lacks several core
735 subunits in addition to the second LHCI belt. The PSI structure of streptophyte algae
736 and prasinophyte algae, however, has not been reported. Based on the data obtained in
737 the present study, *M. viride* PSI possesses LHCA1, two LHCA2s, LHCA3, algae-type
738 LHCA2, and LHCA9. Since *M. viride* LHCA1, two LHCA2s, and LHCA3 exhibit

739 considerable sequence similarity with LHCA proteins in vascular plants, their binding
740 site may be similar to the binding site of the LHCI in vascular plants. Based on
741 sequence similarities, the binding site of algae-type LHCA2 and LHA9 may be similar
742 to the binding site of PSI-LHCI in chlorophytes. The binding site of LHCP to PSII-
743 LHCII is unclear. *M. viride* PSI can bind at least two LHCII trimers, however, the
744 binding site of the additional LHCII trimer is unknown.

745

746 **Fig. S1 A multiple sequence alignment of LHC sequences.** Below is the multiple
747 sequence alignment used to construct a phylogenetic tree in Fig. 2. A MAFFT program
748 (<https://mafft.cbrc.jp/alignment/software/>) was used to align LHC sequences and a
749 ClipKIT program was used to do trimming of the alignment.

750

751 **Figure S2. 2D-SDS-PAGE of the high-molecular-weight PSI-LHCI-LHCII, PSI-
752 LHCI, and LHCII trimer.**

753 The high-molecular-weight PSI-LHCI-LHCII, PSI-LHCI, and LHCII trimer bands
754 separated by CN-PAGE after sucrose density gradient centrifugation were subjected to
755 2D-SDS-PAGE. Protein bands were visualized by silver-staining.

756

757 **Supplementary Table Legends:**

758 Table S1. List of *M. viride* LHC proteins predicted in the Iso-seq analysis.

759 Table S2. Proteins identified in the *M. viride* PSI-PSII band by MS.

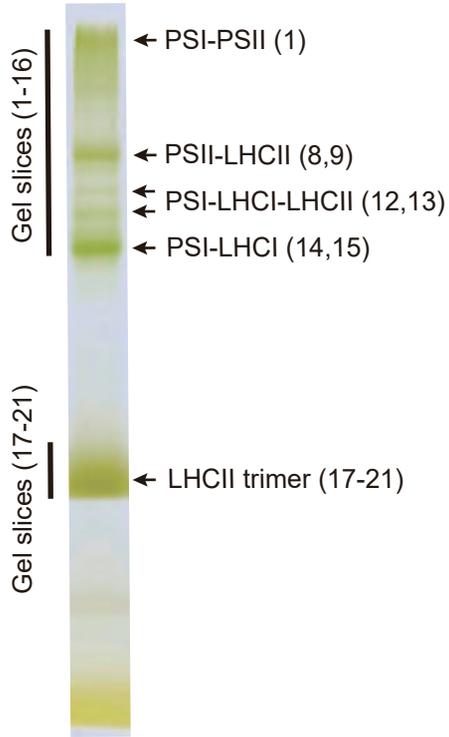
760 Table S3. Proteins identified in the *M. viride* PSII-LHCII band by MS.

761 Table S4. Proteins identified in the *M. viride* PSI-LHCI band by MS.

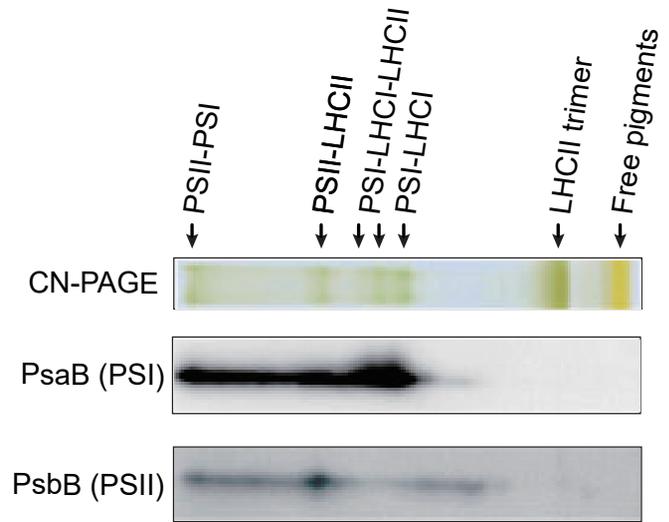
762

Fig. 1

A



B



C

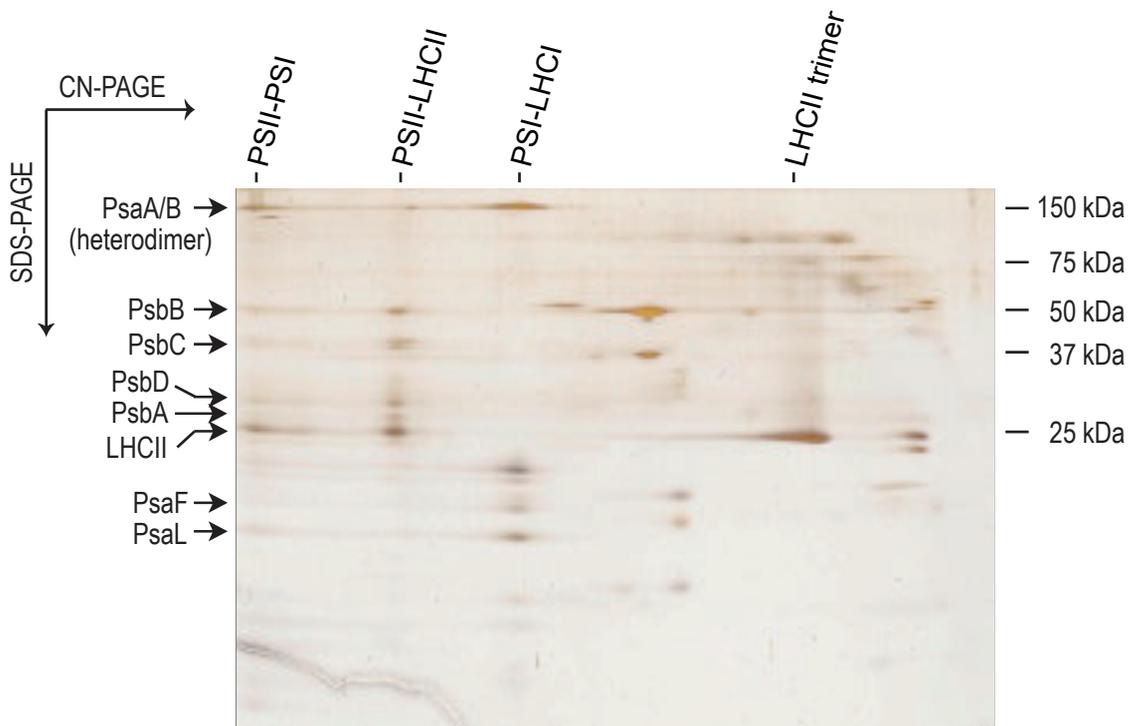


Fig. 2

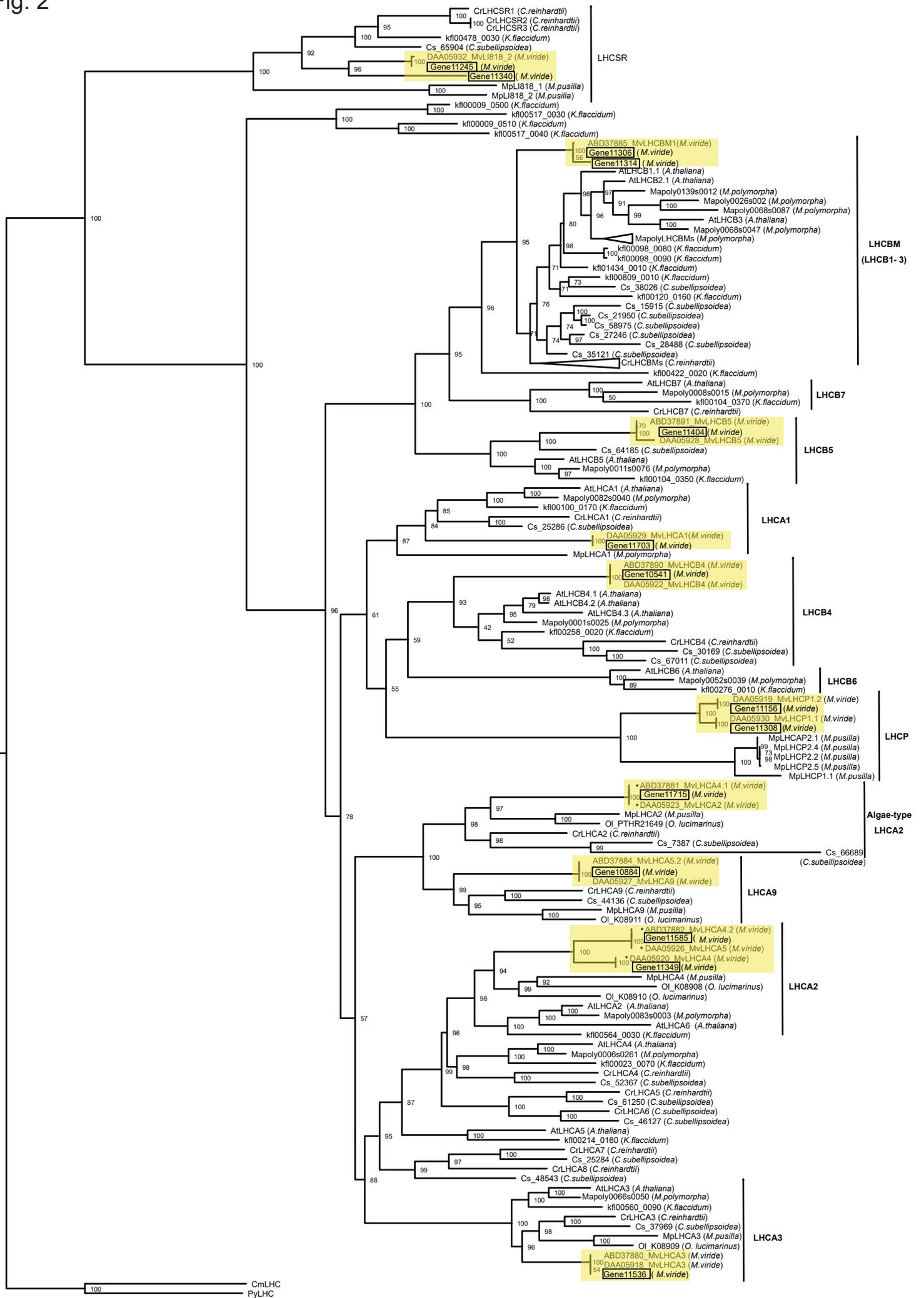
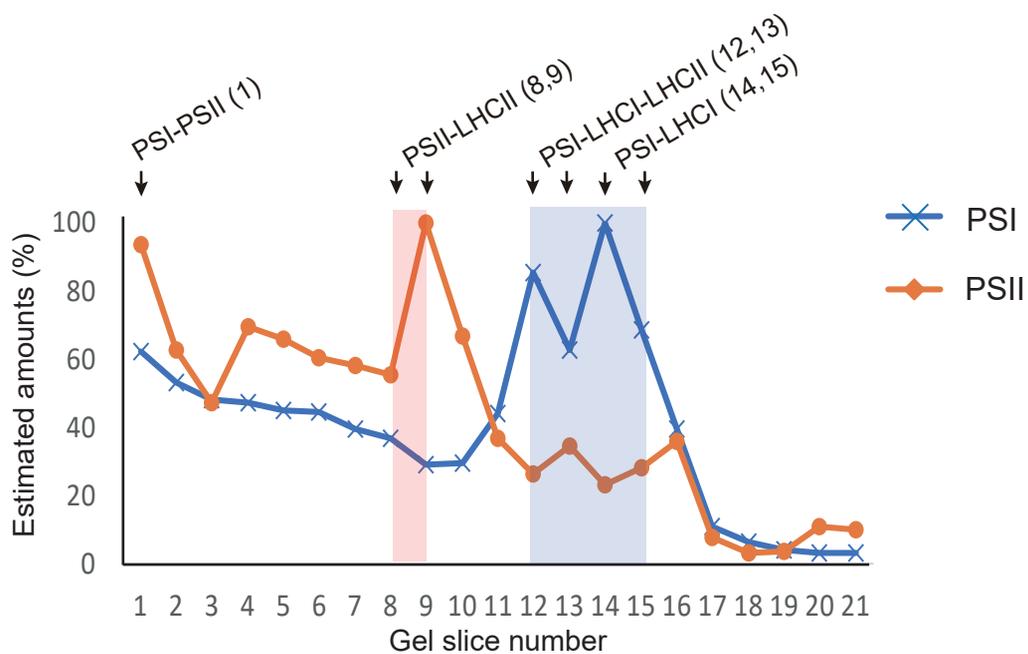
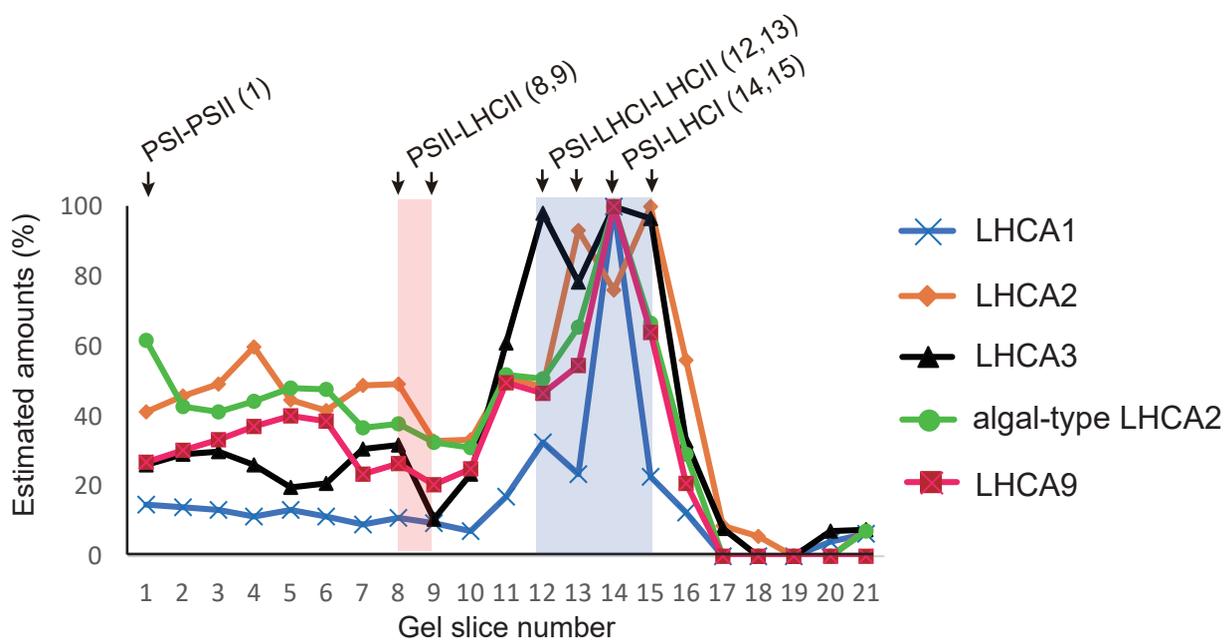


Fig. 3

A



B



C

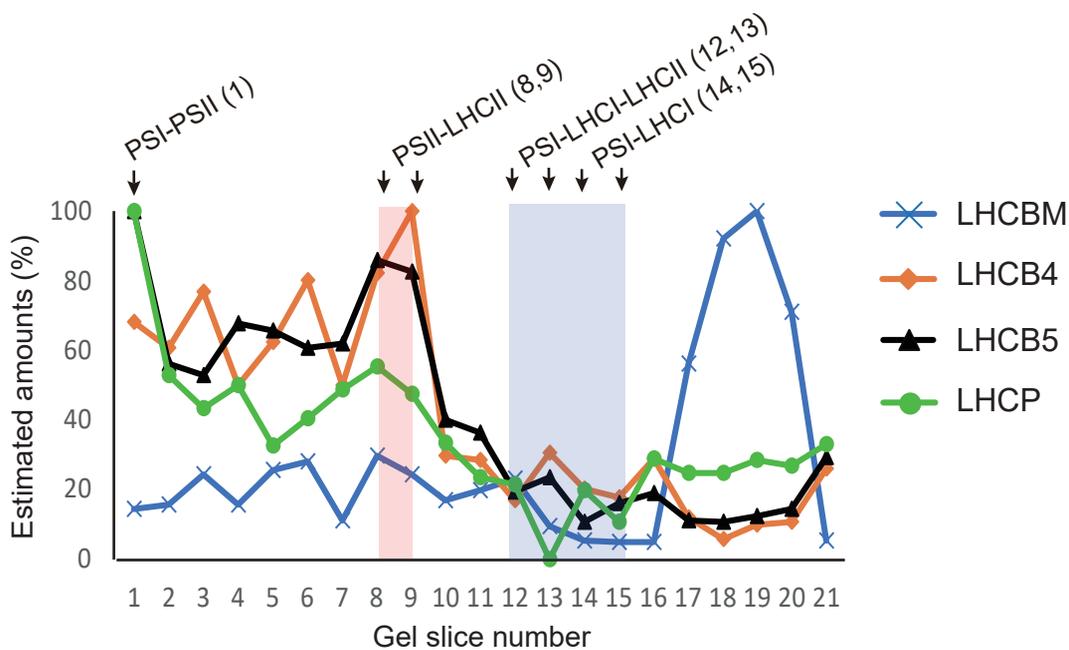
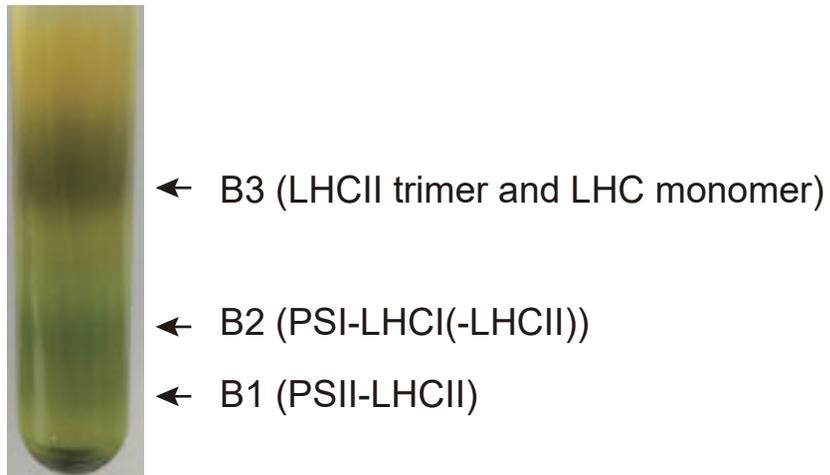


Fig. 4

A



B

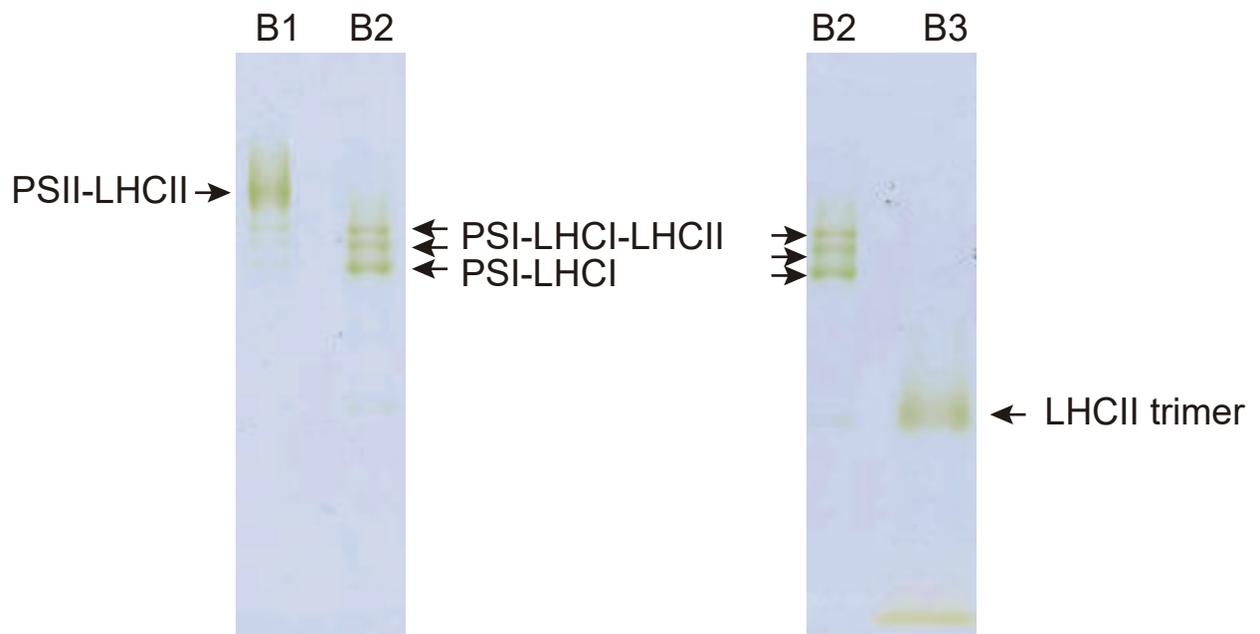


Fig. 5

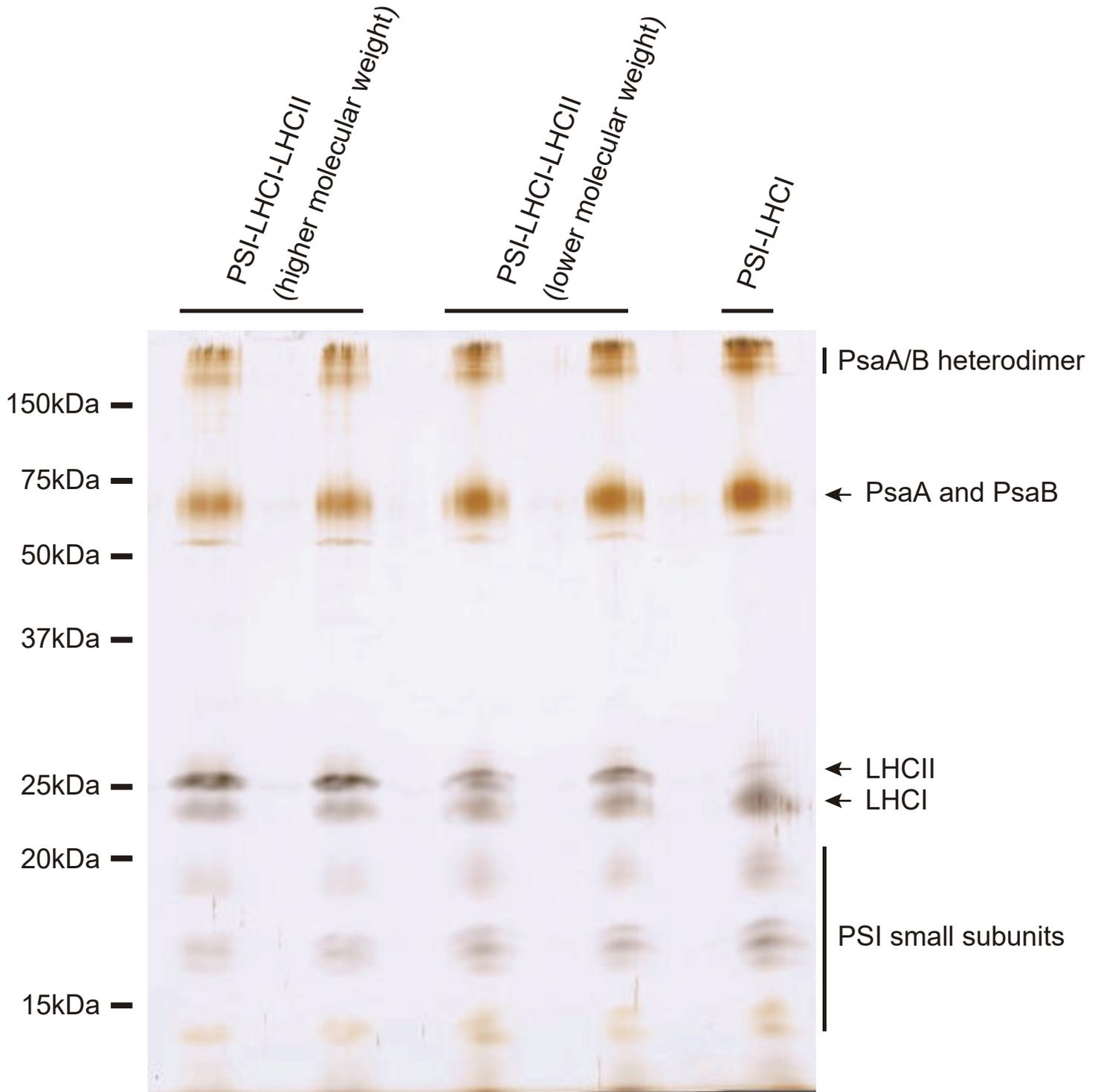


Fig. 6

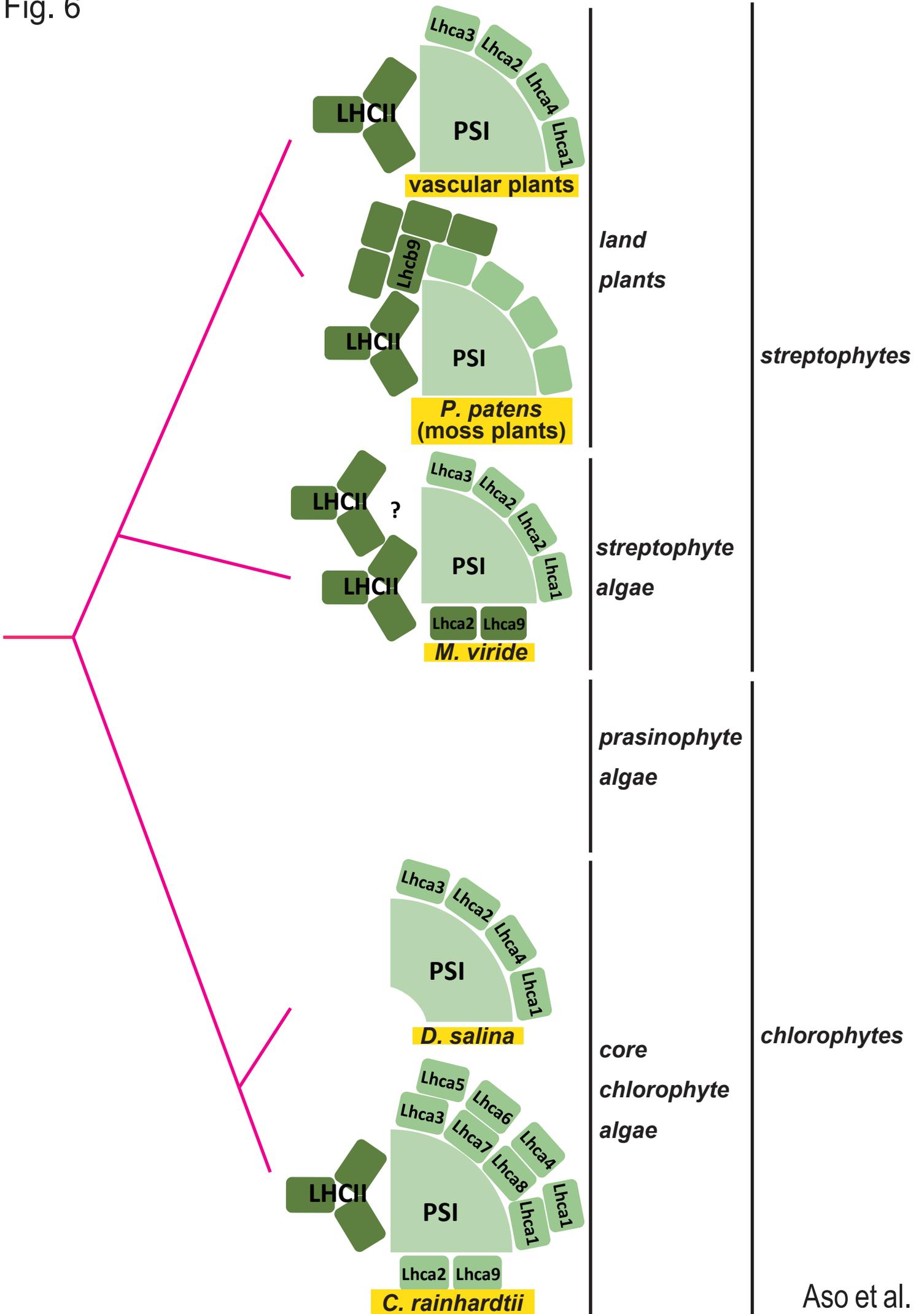


Fig. S1 A multiple sequence alignment of LHC sequences. Below is the multiple sequence alignment used to construct a phylogenetic tree in Fig. 2. A MAFFT program (<https://mafft.cbrc.jp/alignment/software/>) was used to align LHC sequences and a ClipKIT program was used to do trimming of the alignment.

10 20 30 40 50 60

ABD37880_MvLHCA3 MAT--AALASVSVPVVC-GKPF AEKKFLGAR AA--AA--RPQASK---AAKFVV A-
Gene.11536 MAT--AALASVSVPVVC-GKPF AEKKFLGAR AA--AA--RPQASK---AAKFVV A-
AtLHCA3 MAA--QALVSSSLTSV--QT--ARQIFGSKPVAS---ASQK---KSSFVV AA
Mapoly0066s0050 MATTVACRAVAGAVFPA--QE--QRLMGNA RVSFAP--IGTSNA---RSSIVV AS
kfl00560_0090 MAQ--SAAVASRATQLSAARSAASSSFKGTA PA--IP--QLGSCA---RKSVVV AT
CrLHCA3 -----MMLTKSA-----QAAFSG-----KVA--RPAKAN---RARLVC AE
Cs_37969 MAS-----TTMMM SA
MpLHCA3 MAVGQLSVTMASITSSVAVRACGRKTFSGNARVVCAAP--KTTTKR---SFTVRAADE
Ol_K08909 MFA-----RTL--NT TLASAP--ARATKRCA-KRATVVV AE
ABD37881_MvLHCA4.1 MAAVA-SIAASNVAV-----LARKAFVQA KAS-----AKPAGAKP
DAA05923_MvLHCA2 MAAVA-SIAASNVAV-----LARKAFVQA KAS-----AKPAGAKP
Gene.11715 MAAVA-SIAASNVAV-----LARKAFVQA KAS-----AKPAGAKP
ABD37884_MvLHCA5.2 MAA--ATVAMA-----GVSAAPL-----LKA-----SFGV KA
DAA05927_MvLHCA9 MAA--ATVAMA-----GVSAAPL-----LKA-----SFGV KA
Gene.10884 MAA--ATVAMA-----GVSAAPL-----LKA-----SFGV KA
CrLHCA9 -----IAA-----KSQV-----ALGR AP
Cs_44136 MAA--MTQMSRC-----VLG-----ARPVAPSF
MpLHCA9 AAL--TTSSVTAVC--AVRG-AKAKFAG-----S--RAAF-----AAPA TA
Ol_K08911 -----
CrLHCA2 MAM-----L-----K-----SRVSAGVS
Cs_7387 -----
MpLHCA2 MAA--ICTLPVA-----LRASV AP
Ol_PTHR21649 MFS--LTTARHA-----IAN-----RATIVD-----ARSTT AP
ABD37882_MvLHCA4.2 MAA--SVSQMIQ-----ASTFLGSK APV-----ARATAAK
DAA05926_MvLHCA5 MAA--SVSQMIQ-----ASTFLGSK APV-----ARATAAK
Gene.11585 MAA--SVSQMIQ-----ASTFLGSK APV-----ARATAAK
DAA05920_MvLHCA4 -----MIQ-----ASTFLGSK ASV-----ARATPAK
Gene.11349 MAA--SVSQMIQ-----ASTFLGSK ASV-----ARATPAK
AtLHCA2 MAS--SLCASSA-AAI-----SSPSFLGKK RLK-----K-----KLTVPVAVSR
Mapoly0083s0003 MAT--AVCTTSR--ASIAPSTAVKAAFFGKT RSA-----K-----VSAVSK
AtLHCA6 MAF--AIASALT--STLTLSTASTSSTGGRL RERLVVV-RAGK-----
kfl00564_0030 MAQ--CVAAKLA--GSIA-AT-RPAFLSGAT RKS--V--KAVA-----PRAAGNG
MpLHCA4 MAT--IIAQL-----SAG TQR--S--RTVR-----PRSVASK
Ol_K08910 -----
Ol_K08908 M V--SMRISST-----AGLKTRVA--T--KTRA-----
AtLHCA4 MAT--VTTHASA--SIFRPCTSKPRFLTGSSGRLN---R-----DLSFTSIGS
Mapoly0006s0261 MAS--VSMQAMS--VSSAVC--KTSFLSGSSTRFS---R-----NVQVAAAPI
kfl00023_0070 MAS--SVLAKVG--VSTALQLASKTCLVAAPTRGS-----FLAGNKAINA
CrLHCA4 MAF--VLAKSSA--FGVA-----AVKASA PEN--V--KEAR-----EWI--DA
Cs_52367 MGL-----ERPLHA-----S--RAAILRHLELLLTVS---
CrLHCA5 MAA-----LMQKSA SRP-----ACST SS
Cs_61250 MAA--TFTALNA-----APAFLTGRK VSK-----ARAGPAP
CrLHCA6 MML-----VA-----KNA-----VAARPSA
Cs_46127 MAA--LLACNGL-----SRPILGSP AAR---Q-----SRARPVR
AtLHCA1 MASN--SLMSCGI--AA-----VYPSLLSS-----S--KSKF-----VSAGV-P
Mapoly0082s0040 MAAA--TMACVGL-----SAPALLSG-----AQ--RTKF-----FSGHVSS
kfl00100_0170 MAQA--VVARAAA-----SGLVGLTGTAQVLS--Q--KSTY-----LSGS SV
CrLHCA1 MAL-----SMRTL-----A--RTAA-----PRGFSGR
Cs_25286 MAA-----AM TSSFVA--RGAA-----RPAT LP
DAA05929_MvLHCA1 MVAA-----LIQSTFLGAS AAA-----ATAAPVA
Gene.11703 MVAA-----LIQSTFLGAS AAA-----ATAAPVA
CrLHCA7 MAL-----SM AQR-----RAGA-----FSAR AP
Cs_25284 MAN--AVAMSSL-----LGSKSAVA--A--RPAAFSG--KARTARVQA
CrLHCA8 MALT-----M-KRSGVA--A--RSAS-----S
Cs_48543 MNT-----SM TSSFVS--RAAA-----RPAQARP
AtLHCA5 MAV-----VLRGGITG-----RDASSV--ITRRISSVKA
kfl00214_0160 MAS--TLARSQI-----QATSRISGSSRHCSGSAPCRNAVFRGAGLRRHSASPLQ

MpLHCA1 RAL---TISAPIV-----ARAGSFAGRK ATNATN-----
Cs_66689 MNLPTTALSPVT-----VDSIHFLG-----R-----
kfl00009_0500 MA---MTACAVK-----ASALAGSLASVSSST--QSSFCRSF---NVVP G-
kfl00517_0030 MAVS--TLACA IK-----MPALASGLATICRSR--QTSFVQKG---GAVYANA
kfl00009_0510 MAS--STIAASS-----LSKLAGFAATTASTSAARSSFFAGR---ALVS PA
kfl00517_0040 MA A--SSARA IP-----ISGVM TRGAGSAVRTALAGCRLAPQSCVSTGR
ABD37890_MvLHCB4 -----MNTSTFMAQLQVAT--QAIFGTTKGGTVVK--KGTVKK-----AATAVKT
Gene.10541 -----MNTSTFMAQLQVAT--QAIFGTTKGGTVVK--KGTVKK-----AATAVKT
AtLHCB4.1 MAATSAAAAAASSIMT-RVAP-----GHPGSGRFTA--VFGF-----GK KA
AtLHCB4.2 MAATS-TAAAASSIMT-RVVS-----DSSNSSRFTA--RFGF-----GT KA
Mapoly0001s0025 MATAIATSLAASSFCSCREVGSAYSRVF-GPSNAARIVC--RSAIKK-----AASAAGK
AtLHCB4.3 MATT--TAAAASGIFI-----DRPGTGRVQA--RFGF-----SFGK KP
kfl00258_0020 MATSLSSLPSVSKLSL---GS-AARAL-GASNGAAVVA--RVGTTK-----VATTVKK
CrLHCB4 MVFKFPTPPGTQK-----KAGTTATKPAP--KATTKK-----VATSTGT
Cs_30169 MV-----HLSGSFLLGTTALAPQNGATTKIVR-----KAQKAAN
Cs_67011 MVGSS-ALLQRQAFCSSSAIGRRLNALFKGGTKQLKKATAPKTP IKK-----AVPKTSV
ABD37885_MvLHCBM1 MA-----AAL-----ASSTFLGAA VAPAQK--VEAKK--INTVVEA RT
Gene.11306 MA-----AAL-----ASSTFLGAA VAPAQK--AEAKK--INTVVEA RT
Gene.11314 MA-----AAL-----ASSTFLGAA VAPAQK--AEAKK--INTVVEA RT
AtLHCB1.1 MA-----ASTMAL-----SSPAFAGKA NLS-----PAASEVL-GSGRVTM KT
AtLHCB2.1 -----MATSAI-----QQSSFAGQTALKPSNE--LLRKVGVSGGGRVTM RT
Mapoly0057s0073 M-----AAVTAC-----ASATFAGQA GASSNA--LASKVNV-GEARIVM KT
Mapoly0371s000 M-----ASVTAC-----ASTSLAGQA GASSNQ--LAAKVVN-GDARVVM KT
Mapoly0199s0017 M-----AAVTAC-----ASTSLAGQA GASSNA--LASRVNV-GESRVVMAK-
Mapoly0199s0018 M-----ATATAC-----ASTFLAGQA GASSNE--LAKKVVN-GEARVVM K-
Mapoly0199s0019 M-----ATVTAC-----ASSSIAGQA GASSNE--LAKKVVN-GEARVTM K-
Mapoly0199s0012 M-----AAVTAC-----ASTSLAGQA GASSNA--LAAKVVN-GEARVVM KT
Mapoly0199s0014 M-----ASATVC-----ASSTFAGQA GASSNA--LAAKVVN-GEARVVM KT
Mapoly0199s0016 M-----ASATAC-----ASTVLAGQA GASSNA--LAAKVVN-GEARVVM KT
Mapoly0199s0015 M-----AAATAC-----ASSTFAGQA GASSNA--LAAKVVN-GEARVVM KT
Mapoly0199s0013 M-----ASVTAC-----ASTSLAGVA GATSNE--LAKKVVN-GQARVVM KT
Mapoly0199s0020 M-----AAVTAC-----ASTSLAGQA GASSNA--LAGKVVN-GEARVVM K-
Mapoly0057s0082 M-----ASVTAC-----ASTSLAGQA GLSSNA--LAAKVVN-GEARIVM KT
Mapoly0057s0083 M-----AAVTAC-----ASTTLAGQA GSSSNE--LARKVVN-GEARIVM KT
Mapoly0068s0071 -----MAATAV-----SASTFVAGQA GLSGNA--LTQKVVN-GEARVVM KT
Mapoly0139s0012 MA A--ATASAI-----SSTTFAGQS LKGQSE--LSKKVGN-VDARITM RT
Cs_15915 -----MCPA-----
Cs_21950 M-----ASAL-----LCNSFAGAS RSAVP-----TAG-KGNRQVT AA
Cs_58975 -----MCQSFLGAS RSAVP-----TKG-QGNKQVT AA
Cs_27246 M-----ASTM-----MAKTFLGAS AKAVP-----TSG-KGTKQVT AA
Cs_35121 M-----T-----T-----TVM RT
kfl00098_0080 MASALAAASAT-----LVSSFAGKVAAT-----KKISQVSN
kfl00098_0090 MASALAATSVSA-----LVSSFTGKVAAS-----KKVSQVSN
kfl01434_0010 MASSLALRAAQL-----VPSAFVGGTT ASNA-----SEGKIQM RT
CrLHCBM1 MA FAL-----AKSSA-----RAAVSRR-STVKVEA RT
CrLHCBM2 M-----AAI-----MKS AV-RSS-----RPTVSG-RSARVVP AA
CrLHCBM7 M-----AAI-----MKSSV-RSS-----RSTVSS-RSARVVP AA
CrLHCBM3 MA FALSFSRKAL-----QVSAKA-TGKKGTG TA
CrLHCBM4 MA FALATSRKAL-----QVTCKA-TGKKTAA AA
CrLHCBM8 MA FALATSRKAL-----QVTCKA-TGKKTAA AA
CrLHCBM6 MA FALA-SRKAL-----QVTCKA-TGKKTAA AA
CrLHCBM9 MA FALASSRKAL-----QVTCKA-TGKKTAA AA
kfl00809_0010 MASS-ALASSVG-----LAGAFAGKG SN-----AVVKPS-NGTRTVM RT
CrLHCBM5 MM----LSRTVVN-----VQAKLTGGAPKKAAPA-----SAQKKT I EK
Cs_38026 M-----ASAL-----LSNSFV-VSARRTVS-----SPK-SKGKQVT AS
Mapoly0026s002 MAALASQPVAAST-----SVSQSV-SQAPLAAAEP--KQAIS--TTDETL P RV
AtLHCB3 MA STFTSSSSVL-----TPTTFLGQTKASSFNP-----LRDVVSLGSP
Mapoly0068s0047 MAATA-LTKQVV-----SRSTFLGATAAPKDA-----LREQLPVCRA
Mapoly0068s0087 MAALASAPKAAT-----TTTENAPAA SMAAQD--ASSRITSAETSIVNAPKS

Mapoly0199s0018	----GSAPSS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0019	----GSAPSS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0012	---VKSTPTS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0014	---VKSTPTS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0016	---VKSTPTS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0015	---VKSTPTS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0013	---VKSTPTS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0199s0020	----GSAPSS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0057s0082	---SSKSSSS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0057s0083	---SSKSSSS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0068s0071	S--KAAAPAS-IWYGE	RPK	LG	PFS	-GATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Mapoly0139s0012	---KTSTPES-IWYGP	RPK	LG	PFS	-EATP	YLT	GEFP	GDY	GTAGL	ADP	-----T
Cs_15915	-----G	PLS	-SATP	YLN	GEFP	GDY	GTAGL	ADP	-----T		
Cs_21950	-----IEFYGP	RAT	LG	PFT	-DTP	YLN	GEFP	GDY	GTAGL	ADP	-----T
Cs_58975	-----IEFYGP	RAG	LG	PFT	-ESP	YLN	GEFP	GDY	GTAGL	ADP	-----T
Cs_27246	-----IEFYGP	RAG	LG	PFT	-KSP	YLN	GEFP	GDY	GTAGL	ADP	-----T
Cs_35121	V-KGKTTPTS-MWYGE	RPK	LG	PFS	-GNTP	YLS	GEYP	GDY	GTAGL	ADP	-----T
kfl00098_0080	---GARTSMV-EFYGP	RVK	LG	PFS	-GDTP	YLN	GEFAG	GDY	GTAGL	ADP	-----T
kfl00098_0090	---GARTSMV-EFYGP	RVK	LG	PFS	-GDTP	YLN	GEFAG	GDY	GTAGL	ADP	-----T
kfl01434_0010	---VKSAPSS-PWYGP	RLK	LG	PFS	-GETP	YLT	GEFAG	GDY	GTAGL	ADP	-----T
CrLHCBM1	V-SKASTPDS-FWYGP	RPL	LG	AFT	-GEP	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM2	-----IEWYGP	RPK	LG	PFSE	GDTPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM7	-----IEWYGP	RPK	LG	PFSE	GDTPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM3	---AKQAPASIEFYGP	RAK	LG	PYSEN	ATPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM4	-----APKSVEFYGP	RAK	LG	PYSEN	STPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM8	-----APKSVEFYGP	RAK	LG	PYSEN	STPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM6	-----APKSVEFYGP	RAK	LG	PYSEN	ATPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
CrLHCBM9	-----APKGIIEFYGP	RAK	LG	PYSEN	STPAY	YLT	GEFP	GDY	GTAGL	ADP	-----T
kfl00809_0010	V---SSKAS-LWYGP	RAK	LG	PFS	-GDTPNY	YLT	GEYAG	GDY	GTAGL	ADP	-----T
CrLHCBM5	A-NGGNEKLS-AFYGP	RGL	LG	PLS	-GTPAY	YLT	GEFP	GDY	DSAGL	ADP	-----T
Cs_38026	I-----EWYGP	RPK	LG	PYSD	GATP	YLN	GEYP	GDY	GTAGL	ADP	-----T
Mapoly0026s002	P-TETAVDLT-LWYGP	RVK	LG	PKF	-AATPAY	YLN	GEFP	GDY	GTAGL	ADP	-----T
AtLHCB3	---KYTMGND-LWYGP	RVK	LG	PFS	-VQTP	YLT	GEFP	GDY	GTAGL	ADP	-----A
Mapoly0068s0047	---KVSMDAD-YWYGP	RPK	LG	PFS	-AQT	YLN	GEFP	GDY	GTAGL	ADP	-----S
Mapoly0068s0087	D-EEFGVDLS-LWYGAG	RVK	LG	PFS	-ARTP	YLN	GEFAG	GDY	GTAGL	ADP	-----S
kfl00120_0160	GGQPGAGGLG-KWYGP	RAK	LG	PFSE	GGTPAY	LDG	TYPG	GDY	DSSGL	ADP	-----S
Cs_28488	KAAKSASGAS-FWYGA	RPG	LG	AFT	-RSP	YLN	GEFP	GDY	GTAGL	ADP	-----T
ABD37891_MvLHCB5	A---PADGAA-KWYGP	RPK	L-	PTGP	ADIPAY	NGE	AGDY	DPLGLG	-KG-	-----G	
Gene.11404	A---PADGAA-KWYGP	RPK	L-	PTGP	ADIPAY	NGE	AGDY	DPLGLG	-KG-	-----G	
DAA05928_MvLHCB5	A---PADGAA-KWYGP	RPK	L-	PTGP	ADIPAY	NGE	AGDY	DPLGLG	-KG-	-----G	
Cs_64185	KQAGGAQGLD-KWYGPS	RAL	L-	PGG	PSDVPT	YLN	GTAGDY	DPLGLG	KD	TA-----T	
AtLHCB5	S--ETSDELA-KWYGP	RRI	L-	PDGR	SEIPEY	LN	GEAGDY	DPFGLG	KKP	-----N	
Mapoly0011s0076	P--PSNAELA-KWYGP	RRI	L-	PEG	KSDIPAY	LT	GEAGDY	DPFGLG	KKPA	-----D	
kfl00104_0350	RKNASDDELS-KWYGP	RRL	L-	PEG	REEVP	YLT	GEAGDY	DPLGLG	GKKE	-----D	
AtLHCB7	A--QEARNDS-KWYGK	RPR	FG	P	IPIYDYP	YLT	GEAGDY	DIAGL	GKDR	-----T	
Mapoly0008s0015	Q--ADARAKS-LWFGDS	RPK	LG	PLP	-FDYPQ	LQ	GESPGDY	DILGLG	R	TA-----N	
kfl00104_0370	Q--TEARQRS-RWYGP	RPK	LG	PLP	-YEYPG	LS	GEYPADY	DILNLG	VDP	S-----N	
CrLHCB7	A-HAEARRQS-KWYGPS	RPR	--	PGP	-LGSALIGA	AGD	AGD	DPLGLG	KQA	-----S	
kfl00422_0020	LTPKGVDAK-ALYGP	RPK	KG	AFTA	ADV	PDY	LKQYAGDY	IDVLG	FCK	DE-----A	
AtLHCB6	-----AAAQPKS	-	PA	VN	FLDPE	LD	GS	PGD	G	DPLGLG	KDPA-----F
Mapoly0052s0039	A---AKPAF-KGGKGLS	-	PA	IS	FVDPE	LD	GS	PGDY	G	DPLGLG	RDPA-----A
kfl00276_0010	T-----RSTSGRLN	-	PA	YD	LKNP	LD	GS	AGDY	G	DPLGLG	KDPA-----A
DAA05919_MvLHCP1.2	---PARR-----AAR	-	---	---	GTEPLY	---	SGTYS	YD	---	PNWTEG	KTEA-----Q
Gene.11156	---PARR-----AAR	-	---	---	GTEPLY	---	SGTYS	YD	---	PNWTEG	KTEA-----Q
DAA05930_MvLHCP1.1	-----P	---	---	---	---	---	PGTYS	YD	---	PNWTKG	KSDA-----D
Gene.11308	---VARP-----AAR	-	---	---	GSEEVY	---	PGTYS	YD	---	PNWTKG	KSDA-----D
MpLHCAP2.1	-----SVS--T--MI	A	---	---	---	---	DIYP	PGG	GESPI	PF	GSK-----
MpLHCP1.1	-----SAP--S--GS	GTKF	-GSI	-APP	DLYP	---	PGG	GESPI	PF	D	K-----
CmLHC	S-GYATLK-----M	SPAL	-	PF	-LSKPN	LS	SPD	PGYR	G	DPLRF	D-----AFD

PyLHC PV - AARWTMV - - - - KS - - A - PF - MEAPPALDGS IGDVG DPLGF N - - - - YID
CrLHCSR1 - - - - GKRTVS - - - - GKA - GA - - - - PVPAYAKT PGDNV DPAGFLATAS - - - VKD
CrLHCSR3 - - - - GRRTTA - - - - A P - - QT - - - - AAPAYTKN PG GV DPAGFLATAS - - - IKD
Cs_65904 RK - - GLRVLA - - - - LGP - - QAEP - - KAATPAYAKT PGPDI DPANLLS AT - - - IQE
kfl00478_0030 RTFAPIRAVA - - - - D P - - AA - PD - - ATVTPKYAKD PGPNI DPATLLARAA - - PIKE
DAA05932_MvLI818;2 RRNVKVRCTL - - - - G - - PA - PP - - APVSAQFCAG PGGDG DPLNFLGGKT - - - TAE
Gene.11245 RRNVKVRCTL - - - - G - - PA - PP - - APVSAPFCAG PGGDG DPLNFLGGKT - - - TAE
Gene.11340 KT - - EARASK - - - - GS - - GSKPS - - - - - AGN VG DF DPLNL AGK - - - DNE
MpLI818;1 TGVA AAVPRA - - - - GQ - - A - - - - - NVAAA DAG IQPEGI DVFVGFNDI - - - QAE
MpLI818;2 AVTGKTKSTA - - - - GK - - A - - - - - DIKAATQAG VAPEGV DFFGFYNGI - - - EKE

130 140 150 160 170 180
ABD37880_MvLHCA3 PKWLAYS E E I H CR AMLGMAGM APE E LGKLG LIPAETGLVW TGVIP - SEGQYD Y WAS
Gene.11536 PKWLAYS E E I H CR AMLGMAGM APE E LGKLG LIPAETGLVW TGVIP - SEGQYD Y WAS
AtLHCA3 PRWLAYGE E I GR AMLGAAGA APE E LGKAG LIPAETALPW TGVIP - PAGTYT Y WAD
Mapoly0066s0050 PNWLRYA E E I GR AMLGAAGA APE E FGKIG LIPQETAIPW TGVIP - PLGQYS Y WAD
kfl00560_0090 SEWLPYA E E I GR AMLGAAGM APE E LGKMG LIPAKTALPW G VFP - PAGDYS Y WAD
CrLHCA3 PKWL YSE E I H ARWAMLGAAGC APE E LGAAG LIPDATNIKW G VIP - PAGSYN Y WAD
Cs_37969 PAWL YSE E I H ARWAMLGAAGC APE E LATAG VIPQTDEVIW G VIP - PAGVYD Y WLD
MpLHCA3 PQWLAYS E E I H GRWAMLGVAGV APE E LGGMGI IPADTGLVW GAIP - PQGTFD Y WAD
Ol_K08909 PQWLAYS E E I H GRWAMLGVAGM APE E LGGLGI IPQETGLVW AGMIP - AQGT YD Y WAN
ABD37881_MvLHCA4.1 KW GE E GRWAM GVAGM FT ALTQFGISD - - - LPDW AGAVA - - - - - Q - - -
DAA05923_MvLHCA2 KW GE E GRWAM GVAGM FT ALTQFGISD - - - LPDW AGAVA - - - - - Q - - -
Gene.11715 KW GE E GRWAM GVAGM FT ALTQFGISD - - - LPDW AGAVA - - - - - Q - - -
ABD37884_MvLHCA5.2 AW VEA E RT GRWAMAGVAGI FT ALHI - - - - - GSAPW SIGAEE - - - - - S - - -
DAA05927_MvLHCA9 AW VEA E RT GRWAMAGVAGI FT ALHI - - - - - GSAPW SIGAEE - - - - - S - - -
Gene.10884 AW VEA E RT GRWAMAGVAGI FT ALHI - - - - - GSAPW SIGAEE - - - - - S - - -
CrLHCA9 KW AEA E KT GRWAM AVAGI GQE E LGV - - - - - TPAW AGAKE - - - - - Y - - -
Cs_44136 KW TEA E KT GRWAMA GVAGI GQE E LGV - - - - - QPKW AGAKD - - - - - Y - - -
MpLHCA9 KY EA E SRWAM AVAGITTT ALNI - - - - - GGPW AGAGD - - - - - Y - - -
Ol_K08911 KY EA E ARWAM AVAGI GTE E AGI - - - - - EPRW AGTED - - - - - Y - - -
CrLHCA2 KW REA E T GRWAMA AV GI FT A GL - - - - - PK TAGAEK - - - - - Y - - -
Cs_7387 PWLV EGE Y GRVAMLAVAGI LVEAAGL - - - - - GPW SAPFRD - - - - - Y - - -
MpLHCA2 PYL EA E GRWAMHATAGI FT A GL - - - - - PK AGDAA - - - - - L - - -
Ol_PTHR21649 PYL EA E GRWAMAATAGI FT A GL - - - - - PKW AGAVD - - - - - Y - - -
ABD37882_MvLHCA4.2 KWMV AE Q ARWAMLGVAGM SV LQGILPDN - - - VPQW DAPLLA - - - - - DY - - -
DAA05926_MvLHCA5 KWMV AE Q ARWAMLGVAGM SV LQGILPDN - - - VPQW DAPLLA - - - - - DY - - -
Gene.11585 KWMV AE Q ARWAMLGVAGM SV LQGILPDN - - - VPQW DAPLLA - - - - - DY - - -
DAA05920_MvLHCA4 KWMV AE Q ARWAMLGVAGM GA ANKVG LPA - - - AVQW APLVD - - - - - Y - - -
Gene.11349 KWMV AE Q ARWAMLGVAGM GA ANKVG LPA - - - AVQW APLVD - - - - - Y - - -
AtLHCA2 KWNV AE H CRWAMLGAAGI F IPE E F LTKIGILN - - - TPSW TAGEQE - - - - - Y - - -
Mapoly0083s0003 KW V AE H CRWAMLGAAGI F IPE E F LTKIGILN - - - TPSW TAGEQT - - - - - Y - - -
AtLHCA6 KW A AE I H SRWAMLAVTGI IPE E CLERLGFIE - - - NFSW DAGRE - - - - - Y - - -
kfl00564_0030 KW AE Q GRWAMLGAAGI IPE E ALTSAGILH - - - TPHW AGAAD - - - - - Y - - -
MpLHCA4 EW AE Q HARWAMLGVTGMAAPE QNPFVEG - - - AMPNWA APLWD - - - - - GY - - -
Ol_K08910 KW V AE Q HGRWAMLGVAGAAAPE LTNMGISN - - - LPNWH APYD - - - - - GY - - -
Ol_K08908 RQWMV AE Q HARWAMLGVAGC APE E LTKIGVAD - - - LPNWD AATYQ - - - - - Y - - -
AtLHCA4 KW V AE GRWAMLGVAGM LPE E FTKIGIIN - - - VPEW DAGKEQ - - - - - Y - - -
Mapoly0006s0261 NW V AE Q GRWAMLGVAGM VPE E LTKIGLIN - - - APLW DAGKVE - - - - - Y - - -
kfl00023_0070 KW V AE Q GRWAMLGAAGA VPE E LTKAGLWN - - - APLW DAGKAE - - - - - Y - - -
CrLHCA4 KW A AE AR AMLAVAGI VPE E LSNIGFSWPGAGVAW DAGKFE - - - - - Y - - -
Cs_52367 KW AE Q GRWAMLAAGI FV LGHLGAGGPAATPW ASDFT - - - - - Y - - -
CrLHCA5 KW R SE Q HARWAMLGVAGV VQE E KP - - - - - DVY AGLPQ - NLPEP - - -
Cs_61250 KW R AE HARWAMLAVAGI AQE E KP - - - - - GVF NAGLPE - NLPGIN GG P
CrLHCA6 KW E SE H SRWAMA AVAGI VQE E RP - - - - - DVFW NAGKEV - - - - - Y - - -
Cs_46127 AW A E SE R HARWAMLAVAGI AVQE E RP - - - - - DVFW AALKS - PMP - - - GN
AtLHCA1 E E SE I H CRWAMLAVPGI VPE E ALGYGNWVK - - - AQEWAPGGQAT - - - - - Y LGN
Mapoly0082s0040 FA E SE I H CRWAMLAVPGV IPE E ALGLGNWVS - - - AQQWAPGGQAT - - - - - Y LGN

CrLHCBM4 FK **RELE** I H A R W A M L G A G C T P E L A K N G - T K F G - E A V W A G A Q I F S E G G L D Y L G N
CrLHCBM8 FK **RELE** I H A R W A M L G A G C T P E L A K S G - T K F G - E A V W A G A Q I F S E G G L D Y L G N
CrLHCBM6 FK **RELE** I H A R W A M L G A G C T P E L A K S G - T K F G - E A V W A G A Q I F S E G G L D Y L G N
CrLHCBM9 FK **RELE** I H A R W A M L G A G I T P E L Q K N G - V Q F G - E A V W A G A Q I F Q E G G L N Y L G N
kfl00809_0010 FR N **RELE** I H A R W A M L G A G I T P E A L E K A G - V K F G - E A V W A G A Q I F S A D G L N Y L G N
CrLHCBM5 FK **RELE** I H A R W A M L G A G C T P E L A K N G - T P I V - E P V W A G A Q I F A E G G L D Y L G N
Cs_38026 FS **RELE** I H S R W A M L G T G I T P E L E K G G - V K F Q - E A V W A G A Q I F S A D G L N Y L G N
Mapoly0026s002 FA N **RELE** I H S R W A M L G A A G C F P E L A K N A - V S F R - E P V W A G Q I F S D G G L D Y L G N
AtLHCB3 FA N R A L E I H G R W A M L G A F G C T P E L Q K V R - V D F K - E P V W A G Q I F S E G G L D Y L G N
Mapoly0068s0047 FA N R A L E I H G R W A M L G A G C L P E A L V K S R - V T L K - E A V W A G Q I F T D G G L D Y L G N
Mapoly0068s0087 FA N **RELE** I H G R W A M L G W A G C F P E L V K K S - C G L K - E G V W A G Q I F T D G G L D Y L G N
kfl00120_0160 FR N **RELE** I H S R W A M L G A G M L P E L A D S G - I P I K - E P V W A G A Q I F D S D G L N Y L G N
Cs_28488 FS **REIE** I H A R W A M L G A G I V P E L D Q T N - - - - - H V A W D A G A T I F G P K G I Q Y L G V
ABD37891_MvLHCB5 E **RPYE** H G R W A M L G V G M V P E G L Y A N G N T N I K - G A V W D G A V L L D P S T L T A G I
Gene.11404 E **RPYE** H G R W A M L G V G M V P E G L Y A N G N T N I K - G A V W D G A V L L D P S T L T A G I
DAA05928_MvLHCB5 E **RPYE** H G R W A M L G V G M V P E G L Y A N G N T N I K - G A V W D G A V L L D P S T L T A G I
Cs_64185 E **RAYE** I H A R W A M L A A A G I I P E G L Q A N G - A A I K - G G T W T G A E M L N G G T L N Y F A V
AtLHCB5 FA A F E I H A R W A M L G A A G F I P E A L N K Y G A N C G P - E A V W T G A L L L D G N T L N Y F G K
Mapoly0011s0076 FD A Y E I H A R W A M L G A A G F L P E A F N K Y G A V C G P - E A V W T G A L L L E G N T L Q Y F G A
kfl00104_0350 FD **RAAE** I H A R W A M L G A A G I I P E A F N R S G L P C G P - E A V W T G A Q L L E G E S L Q C G I
AtLHCB7 FD F N F E H A R W A M L A A G A I P E F D L T G T F H F A - E P V W R V G Y S K L Q G E T L E Y L G I
Mapoly0008s0015 FD F N F E H A R W A M L G A G A I P E L V R Y G G L S F S - E P V W R V G Y A K L Q G E T L D Y F G I
kfl00104_0370 FQ F N F E H A R W A M L G A G A I P E L Q Y F S A L D F T - E P V W V G Y A K L Q G E D L D Y F G I
CrLHCB7 FA **EAE** H A R W A M L G V G C V P E L A L R G - V D L G - E P V W V G A S K L N S D T L N G G I
kfl00422_0020 FA L **RAQE** F A R W A M L G V G M Y P E F F P A E G - - - - - F E P V W T G A Q I F D P A G I D Y L G A
AtLHCB6 KW **REAE** I H G R W A M A A V G I F V G A W S G - - - - - V A W A G A Q - - - - -
Mapoly0052s0039 KW **REAE** I H G R W A M A A V G I F V G A W S G - - - - - I P W A G A D - - - - -
kfl00276_0010 KW **REAE** I H G R W A M A T G C I G A F S G - - - - - I P W A G A A - - - - -
DAA05919_MvLHCP1.2 A Y N V **EVE** I H G R W A M L G C A G A W A A E Q G T G - - - - - I N W A G A I C T P A D G I H P G E
Gene.11156 A Y N V **EVE** I H G R W A M L G C A G A W A A E Q G T G - - - - - I N W A G A I C T P A D G I H P G E
DAA05930_MvLHCP1.1 T Y N V S V E I H G R W A M L G C A G A W A A E Q G T G - - - - - I N W A G A I C T P A D G I H P G E
Gene.11308 T Y N V S V E I H G R W A M L G C A G A W A A E Q G T G - - - - - I N W A G A I C T P A D G I H P G E
MpLHCAP2.1 - - - N A **EERE** I H G R W A M L G V T G A W A A E N G T G - - - - - I P W T A G T L C T P D D A V K P G A
MpLHCP1.1 - - - N A **EERE** I H G R W A M L G V T G A W A A E N G T G - - - - - I P W T A G T L C T P D D A V K P G A
CmLHC N W L **EGE** K G R V A M L A C H F F V T E F Y Q F - - - - - P F A G A P K L A - - - - A H D Y F V K
PyLHC K F L **REAE** K H C R V T M L A V G L F V Q E F Y T L - - - - - P F S G G P A L A - - - - S H N Y F V T
CrLHCSR1 R **RESE** T H G R V A M L A A G F V G E Q L Q D - - - - - F P L F D G R V S - - - - A I Y F Q Q
CrLHCSR3 R **RESE** T H G R V A M L A A G F V G E Q L Q D - - - - - F P L W D G R V S - - - - A I Y F Q Q
Cs_65904 K **RESE** T H G R V A M L A A G W V G E F A D - - - - - K K L L S D G R I T - - - - A I D F Q Q
kfl00478_0030 K **RESE** H G R V A M L A T G F V G E Q E D - - - - - F P A R F F P H V T - - - - A I Y F Q Q
DAA05932_MvLI818;2 K M L **RESE** H G R V A M L A T G F V G E T F N P - - - - - L - - - F G G I T - - - - A I N Q F Q Q
Gene.11245 K M L **RESE** H G R V A M L A T G F V G E T F N P - - - - - L - - - F G G I T - - - - A I N Q F Q Q
Gene.11340 L L **REAE** T H G R V M L A T A G F A E F N P L - - - - - F N G I K - - - - A I N F Q Q
MpLI818;1 A Q A V E T H G R L A M L A S G F V G E S E G - - - - - S S L F D S Q V T - - - - A I N F Q Q
MpLI818;2 A Q A V E T H G R I A M L A S G F V G E Q E G - - - - - S A L F D A N I T - - - - A I D F Q Q

190 200 210 220 230 240

ABD37880_MvLHCA3 P - - - - - L A F W L M Q F A E L R R W Q D Y R H P G S Q S K Q Y F L G L E Q F F G G S G P S Y P G G
Gene.11536 P - - - - - L A F W L M Q F A E L R R W Q D Y R H P G S Q S K Q Y F L G L E Q F F G G S G P S Y P G G
AtLHCA3 N - - - - - Y T F V A L M G F A E H R R L Q D W Y N P G S M G K Q Y F L G L E K G L A G S G P A Y P G G
Mapoly0066s0050 P - - - - - Y T F V A L M G F A E H R R A Q D Y Y K P G S M G K Q Y F L G F E K V L G G S G P A Y P G G
kfl00560_0090 S - - - - - Y T F G L M A F A E H K R L A D Y R K P G S Q G K V F F L G M E K F L G G S G P A Y P G G
CrLHCA3 P - - - - - Y T F F A M Q F A E L R R L Q D F R Y P G S M G Q Q Y F L G L E A I F K G S G A A Y P G G
Cs_37969 P - - - - - Y S F F L Q F A E L R R W Q D F R N P G S Q G K Q Y F L G L E E V L K G S G P S Y P G G
MpLHCA3 P - - - - - T T F W N L M N F A E V K R G Q D Y W Y P G S Q G E T P L M G W E K G F A G S G A P A Y P G G
OL_K08909 P - - - - - F T F W N A A L M N F A E L R R A Q D Y W N P G S M G K Q E L I G W E K M L G G S G P A Y P G G
ABD37881_MvLHCA4.1 S - - S P F - D Q T A F F A T E G F R I S T W K K T G K - - - - - P G -
DAA05923_MvLHCA2 S - - S P F - D Q T A F F A T E G F R I S T W K K T G K - - - - - P G -

kfl00422_0020 PNV NAHS IA AAF A LMGGA E FAR LKAPK - - - - - EM GLY PGG
AtLHCB6 PDA APFSFGS GT LMGW ESKR WVDFFNPDSQSVEWATPWSKTFANTG QGY PGG
Mapoly0052s0039 PSA APFSFGT GT LMGW ESKR WADYVNPNSQLVDWATPWSRTFGNTGLQGY PGG
kfl00276_0010 PGA APFTFGT GT LMSW ESKR WVDFFNPSSQSVEWATPWSKTFANTGQQGY PGG
DAA05919_MvLHCP1.2 PAGSGFPNFYIQ A ST AMGLA EGYRGG LIDSC - - - - - FPETVGD L PGG
Gene.11156 PAGSGFPNFYIQ A ST AMGLA EGYRGG LIDSC - - - - - FPETVGD L PGG
DAA05930_MvLHCP1.1 PEGSGFPNFYIQ G S MGLA EGYRGG LIDNV - - - - - FPEEVGD L PGG
Gene.11308 PEGSGFPNFYIQ G S MGLA EGYRGG LIDNV - - - - - FPEEVGD L PGG
MpLHCA2.1 PEGSGYPSFWA A L GLA EAYRTGLT DPA - - - - - FDETVDG DVS PGG
MpLHCP1.1 PEGSGYPSFWA A FL GSA E CYRTGL FENP - - - - - FPEEL - SVT PGG
CmLHC - - - SGAMQ LAF GFLE FL HRG VLYSDMEW - - - - - KGRK PGE
PyLHC - - - QGAMQ LLW CGLE G VPAVL M-MQG - - - - - SGRR PGE
CrLHCSR1 - - - GQG - - - FWEPL A GVA ESYR VAVGWTP TGTGFN - - - - - SLK DYE P GD
CrLHCSR3 - - - GQG - - - FWEPL A GVA ESYR VAVGWTP TGTGFN - - - - - SLK DYE P GD
Cs_65904 - - - EYK - GAIFWEPL FS GLA EAWR IGVGNPSSDKFN - - - - - QLR DYS PGE
kfl00478_0030 - - - ENE - GAIFWEPL FA ALA ESYR VGLGWTP DSTNFN - - - - - TLR DYE PGN
DAA05932_MvLI818;2 - - - PQP - - - FWE V A GLA EGFRI NRGWSPAEAFS I - - - - - GVLK SYT PGT
Gene.11245 - - - PQP - - - FWE V A GLA EGFRI NRGWSPAEAYFS I - - - - - GVLK SYT PGT
Gene.11340 - - - PQP - - - FWE VC G AMA EAYR LQEGWNPRDGGY - - - - - KLRPGYQ P GD
MpLI818;1 - - - PLP - - - FWLGLGAL L FI EASRVQ IAWSPDASRLG - - - - - LMK DHT P GD
MpLI818;2 - - - PGT - - - FWG LGA FT EANRVQTAWNPEADKLF - - - - - LLK DYT P GD

250 260 270 280 290 300
ABD37880_MvLHCA3 AIF FLGYGKDE - KSMN LK KE KNGRLAM AT GFF QA ATG GPF NLT HLADPF
Gene.11536 AIF FLGYGKDE - KSMN LK KE KNGRLAM AT GFF QA ATG GPF NLT HLADPF
AtLHCA3 PFF PLGFGKDE - KSLK ELK KE KNGRLAM AI G F QG VTGVGPY NLL HLADPV
Mapoly0066s0050 PLF FLGFGRDE - KSMK LK KE KNGRLAM AV G F QA FTGVGPF NLL HLADPA
kfl00560_0090 PFF FAGFGKDE - KSKK L D KE KNGRLAM AV G F VQG STG GPF NLL HLSPT
CrLHCA3 PFF LFN LGKTE - AAMK ELK KE KNGRLAM AM G GAQA TG GPF NLV HLADPV
Cs_37969 QFF LFN LGKSD - - LK ELK T KE KNGRLAM AVFG GAQA TA GPF NLT HLDPT
MpLHCA3 KYF FAN LGKTD - - MA K KEI KNGRLAM AF GIAVQA ATG GPF N V H DPF
OL_K08909 - FF I GQGKSD - - MA KEI KNGRLAM ACFACGAQA TG GP NLI H DPF
ABD37881_MvLHCA4.1 - FLDI G DS - - - - P KEI KNGRLAMFAF GFVAQYFVNGMGP GLL HLE PQ
DAA05923_MvLHCA2 - FLDI G DS - - - - P KEI KNGRLAMFAF GFVAQYFVNGMGP GLL HLE PQ
Gene.11715 - FLDI G DS - - - - P KEI KNGRLAMFAF GFVAQYFVNGMGP GLL HLE PQ
ABD37884_MvLHCA5.2 - - FDP LG - - - - - NSK E M KE KNGRLAMFAWC GFMVQA VTR GP TNLEKH ADPF
DAA05927_MvLHCA9 - - FDP LG - - - - - NSK E M KE KNGRLAMFAWC GFMVQA VTR GP TNLEKH ADPF
Gene.10884 - - FDP LG - - - - - NSK E M KE KNGRLAMFAWC GFMVQA VTR GP TNLEKH ADPF
CrLHCA9 - - FDPAG - - - - - NSPS AT KE KNGRLAM AF GFCVQA ATRTQP GLTAHLADPF
Cs_44136 - - FDPANL DS - - - - - EANA E KEI KNGRLAM AF GFVGA VCR GP ALQSHLDPF
MpLHCA9 - - DP G - - - - - NNDA A KEI KNGRLAM AF GVVVQA VYRTGP AALK H VDPF
OL_K08911 - - FDP G GS - - - - KDE K KEI KNGRAM AF GIVVQG VYR GP AALK H A PF
CrLHCA2 - - FDP G KS - - - - - EE K KE KNGRLAM AF GFC SQA AVY GP TLQLHLADPG
Cs_7387 - - FDP LN RDY - - - - - K QSE KNGRLAM AF GFASQAANTG GP NLK H ADPT
MpLHCA2 - - FDP SGQ - - - - - DSPA K KEI KNGRLAM F GMVSQYAVTGTSP GLKAH A PT
OL_PTHR21649 - - FDPAGL - - - - - DAP K KEI KNGRLAM AF GMVSQYAVTGLSP GLEAH A PQ
ABD37882_MvLHCA4.2 - M FDP LGF KG - - - NLE ESKWKE KNGRLAM AS GMF QYDATG SP TNLAHLADPW
DAA05926_MvLHCA5 - M FDP LGF KG - - - NLE ESKWKE KNGRLAM AS GMF QYDATG SP TNLAHLADPW
Gene.11585 - M FDP LGF KG - - - NLE ESKWKE KNGRLAM AS GMF QYDATG SP TNLAHLADPW
DAA05920_MvLHCA4 - P FDP LGF KG - - - NLP ELK KEI KNGRLAM AS GLFVQYSATGASP DNLA AHLADPG
Gene.11349 - P FDP LGF KG - - - NLP ELK KEI KNGRLAM AS GLFVQYSATGASP DNLA AHLADPG
AtLHCA2 - L FDP LGW GSGSPA K LK EL T KEI KNGRLAM AV GAWFQH YTGTP DNLF AHLADPG
Mapoly0083s0003 - F FDP LGW GAGGA AKVK EL T KEI KNGRLAM AV GAWFQA YTGTP DNLF AHLADPG
AtLHCA6 - L FDP MWGRGSPEPVMV L T KEI KNGRLAM AF GFCFQATYTS DP NLMAHLADPG
kfl00564_0030 - I FDP FGYAKD S - SKAG ELK KEI KNGRAM AMA GFFQA YTGVP DNLTTHLADPG
MpLHCA4 - L FDP G AKD P - EKVKV LK KEI ANGR LAM AV GCI QNEATGVGP ANLKAHLADPA
OL_K08910 - I FDP LGYAKG - - - DLKTLKAKEI ANGR LAM AFAGIMVQYDHTGVGP ANLVSH ADPA
OL_K08908 - L FDP KFGMAKDE - KTTA ELK KEI KNGRLAM AF GICAQY QTGVGP NLFSH G PG

AtLHCA4 - I F P L N F A P T - - - - Q E A K E K E I A N G R L A M A F G F V V Q H N V T G G P F N L L Q H L D P W
Mapoly0006s0261 - I F P L K F P A - - - - Q E Y K E K E I A N G R L A M A F G M L V Q K T G A G P F N L L T H L A D P W
kfl00023_0070 - I F P L N F P V - - - - K E S K D K E I A N G R L A M A F G I A Q H V T G G P F A N L L A H L A D P W
CrLHCA4 - I F D P F G W K G - - - D I K S L K K E I K N G R L A M A F A G F I G Q A Y T T G T T P N L S T H L A D P W
Cs_52367 - I F D P L G Y K G - - - N M A E L K K E I K N A R L A M A V G F F V Q A K T T G T P D L S S H L A D P W
CrLHCA5 - I F D P F G F K G - - - N L K E L T K E I K N G R L A M A Y A F I Q A Q A T G G P A A L S A H L A P F
Cs_61250 - I F D P F G F K G - - - N F K E T K E I K N G R L A M A F G F I Q A Q A T G G P A N L G A H L A P F
CrLHCA6 - V F A P F - - - - I P - G D L A E L K K E I K N G R L A M A F G F V A A Q V T G G P A A L Q H L A D P W
Cs_46127 - I F A P - - - - I P - G D L A E L K K E I K N G R L A M A F A G F V Q A Q V T G G P A L G H L A P I
AtLHCA1 - A F D P L G Y K D P - K K L E E L K K E I K N G R L A M A F G F C V Q Q S A P G T G P N L A T H L A D P W
Mapoly0082s0040 - P F D P L G F A K D P - K K L E E Y K K E I K N G R L A M A F G F S V Q A A P G T G P N L A S H L A P W
kfl00100_0170 - A F D P L G F K D P - A K L D E L K K E I K N G R L A M A F G F A V Q A A V P G T G P N L A Q H L A D P W
CrLHCA1 - A F D P L G F A K D S - S K S G E L K K E I K N G R L A M A F G F V A Q H A A T G G P A A L G H L A P W
Cs_25286 - A F D P L G F K D S - K V L E E N K K E I K N G R L A M A F G F I A Q H A A T G G P A L K L H L A D P W
DAA05929_MvLHCA1 - A F D P L G F G K D E - - - - V N K K E I K N G R L A M A F G M W C Q G H V G G G P S N L M H L A D P L
Gene.11703 - A F D P L G F G K D E - - - - V N K K E I K N G R L A M A F G M W C Q G H V G G G P S N L M H L A D P L
CrLHCA7 R F F D P G L R G D A A K Y Q E Y K Q K E I K N G R L A M A C G F A A Q Y A A T G G P D N L A H L A D P N
Cs_25284 K L F D P F G L R G S E G Q L Q Y E N E I K N G R L A M A F G F I A Q H A A T G G P D N L A H L A P T
CrLHCA8 - P F D P L G L K A - D K W A W K K E I K N G R L A M A F G F V A Q K Y A T G A G P D N L A A H L K D P W
Cs_48543 - V F D P L G L K G P - - G Y E A K K E I K N G R L A M A F G F A S Q Y A T G G P N L L H L A P L
AtLHCA5 P L L P L G L A K D V - Q N A H W K K E I K N G R L A M A M G F F V Q A S V T H T G P D N L V H L A P W
kfl00214_0160 P F F I T G L G T G I K D D V F T K E I K N G R L A M A V G F I V Q A V T K T G P D N L V H L A D P F
MpLHCA1 - A F D P A G F K G A - - D F E T L K K E I A N G R L A M A F A G L I G Q A Q A T N G P S A L G R H A D P W
Cs_66689 A P F D P G L T T D Y - - - - - N Q A E I K N C R L A M A A G F A T Q A Y V T G G P N A V H L R D P F
kfl00009_0500 - - F D P L G L A T G D D A Q I K T E I K N G R L A M A T G G F L H Q A H T G G P A N L M H L A D P S
kfl00517_0030 - - F D P L G L T T G E A D E I K I L E K E I K N G R L A M A A G G F L H Q A H T G G P Y A N L I H L A D P S
kfl00009_0510 - - F D P L G F A K G D D A K V K T E I K N G R L A M A F G F L Q Q Y V T G G P T N L S Q H L A P H
kfl00517_0040 - - F D P L G F L E D K E E E V N E L T K E I K N G R L A M F G F L W Q Y E T G G P T N H T K H L A D P S
ABD37890_MvLHCB4 - P F D P F G F T K A S E S R V N L K A E I K A R L G M A S G F A V Q Y A Y T G A G P N L A V L I G R G
Gene.10541 - P F D P F G F T K A S E S R V N L K A E I K A R L G M A S G F A V Q Y A Y T G A G P N L A V L I G R G
AtLHCB4.1 K F F D P L G L A A D P - E K T A L A E I K A R L A M A F G F A V Q A A A T G G P N N W A T H L D P L
AtLHCB4.2 K F F D P L G L A S D P - V K K A L A E I K A R L A M G F G F A V Q A A A T G G P N N W A T H L D P L
Mapoly0001s0025 K Y F D P L G L A S D P - E K K E I L K A E I K A R L A M G A G F A V Q A A A T G G P D N W A T H L D P L
AtLHCB4.3 - Y F D P L G L A A D P - E K L D T L K A E I K S R L A M A F I F A Q A A F T G G P S F L A T F - - - -
kfl00258_0020 - Y F D P L N L A S S P - E K A A L K A E I K G R L A M A G F A V Q A F A T G T G P D N L A Q H L D S P T
CrLHCB4 - V F D P L K L A S D E E R A F R L K T A E I K A R L A M F F G G V Q A S T G G A G L A K F A D G L N
Cs_30169 - L F D P L K L A S D D S Q R T F L E A E I K G R L A M A F G A Q Q R Q - - - - - L E R S Q W
Cs_67011 - Y F D P L G L A T G D D A R A F L K E A E I K G R L A M A F G F G V Q A A T G G A G L A K F A S F A
ABD37885_MvLHCBM1 S - F D P L G L G D D P - D T L A E L K K E I K N G R L A M F M F G F Q A V T G G P N L D A H L A D P F
Gene.11306 S - F D P L G L G D D P - D T L A E L K K E I K N G R L A M F M F G F Q A V T G G P N L D A H L A D P F
Gene.11314 A - - - - - D G P - G - - - - -
AtLHCB1.1 S - F D P L G L A T D P - E A F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L A H L A D P V
AtLHCB2.1 A - F D P L N L A E D P - E A F S E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L F H L A D P V
Mapoly0057s0073 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P A
Mapoly0371s000 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P A
Mapoly0199s0017 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0199s0018 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0199s0019 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0199s0012 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P V
Mapoly0199s0014 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P V
Mapoly0199s0016 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P V
Mapoly0199s0015 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P V
Mapoly0199s0013 Q - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P A
Mapoly0199s0020 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0057s0082 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0057s0083 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P T
Mapoly0068s0071 S - F D P L N L A E D P - D T F A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L S H L A D P V
Mapoly0139s0012 A - F D P L G L A N D P - D T F A E L K K E I K N G R L A M F A F G F F V Q A V T G G P D N L L S H A D P A
Cs_15915 P - F D P L E F A S D P - D T L A E L K K E I K N G R L A M F M F G F F V Q A V T G G P N L N H L A S P Y

Cs_21950 P - FDPLGLADDP-DTFS ELK KEIKNGRLAMF MFGFFVQA VTG GP NLS HLA DPG
Cs_58975 P - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLS HLA DPG
Cs_27246 A - FDPLGLADDP-DTFA EL KEIKNGRLAMF MFGFFVQA VTG GP NLA HLA PS
Cs_35121 DY FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLES HLA DPS
kfl00098_0080 Q - FDP LN LAEDP-DTFA EL T KEIKNGRLAMF MFGFFVQA VTG GP NLA HLD PT
kfl00098_0090 Q - FDP LN LAEDP-DTFA EL T KEIKNGRLAMF MFGFFVQA VTG GP NLS HLD PS
kfl01434_0010 Q - FDP LN LAEDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP ANLN HLA DPA
CrLHCBM1 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLS HLA PG
CrLHCBM2 A - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HLA PT
CrLHCBM7 A - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HLA PT
CrLHCBM3 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HL PT
CrLHCBM4 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HLA PT
CrLHCBM8 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HLA PG
CrLHCBM6 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLD HLA PT
CrLHCBM9 S - FDPLGLADDP-DTFA ELK KEIKNGRLAMF SFGFFVQA VTG GP NLD HL PG
kfl00809_0010 Q - FDP LN LAEDP-EEFA ELK KEIKNGRLA AN GFFVQA VTG GP NLE HLA PA
CrLHCBM5 QF FDP LN LAEDP-DAFA ELK KEIKNGRLAMF MFGFFVQA VTG GP ANLD HLAS PF
Cs_38026 A - FDP LN LAEDP-DTFA ELK KEIKNGRLA QFAA GFFVQA VTG GP NLE HLA DPF
Mapoly0026s002 A - FDP LN LAEDP-EAFA ELK KEIKNGRLAMT AM GFFVQA VTG GP NLY HLA PQ
AtLHCB3 QY FDP LN LAEDP-VTFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLL HLD PV
Mapoly0068s0047 KY FDP LN LAEDP-ETFA ELK KEIKNGRLAMF MFGFFVQA VTG GP NLL HLD PV
Mapoly0068s0087 DY FDP LN LAQDP-DTLA ELK KEIKNGRLAMT AM GLFVQA VTS GP NLA HLA PY
kfl00120_0160 A - FDP LN LAEDP-EAAA ELK KEIKNGRLAMF CF GFFVQA VTG GP NLQ HLA DPG
Cs_28488 S - FDP LN LAEDP-DTFA ELK KEIKNGRLAMF NF GFFVQA VTG GP ANLN HLA PT
ABD37891_MvLHCB5 KY FDP LN LAEDP-SKAD ELK KEIKNGRLAM AF GFV VQYFVTK GPYANWSK HLA DPF
Gene.11404 KY FDP LN LAEDP-SKAD ELK KEIKNGRLAM AF GFV VQYFVTK GPYANWSK HLA DPF
DAA05928_MvLHCB5 KY FDP LN LAEDP-SKAD ELK KEIKNGRLAM AF GFV VQYFVTK GPYANWSK HLA DPF
Cs_64185 P - FDP LN LAEDP-ETFA ELK KEIKNGRLA V GFAVQ YVTG GPYANWSK HLA DPF
AtLHCB5 P - FDP LN LAEDP-EQGALLK KEIKNGRLAMF AM GFF QAYVTG GP NLA HL DPF
Mapoly0011s0076 A - FDP LN LAEDP-DQFALLK KEIKNGRLAMF M GFF QAYVTG GP NLA HL DPF
kfl00104_0350 A - FDP LN LAEDP-DQFALLK KEIKNGRLA FL GFFVQAYVTQD GP ANLA HL DPF
AtLHCB7 TL FDP LN LAEDP-VAFE LK KEIKNGRLAM AW GFYAQA AFTG GP NLV H DPL
Mapoly0008s0015 AL FDP LN LAEDP-STFE LK KEIKNGRLAM AW GFYAQA VTG GP NAL FL DPA
kfl00104_0370 AL FDP LN LAEDP-EAFE LK KEIKNGRLAM AW GFY QA VTQ GP NLL FA DPA
CrLHCB7 GP FDP LN LAADA-DGFVEQA KEIKNGRLAM AM GFFVQAAVTRV GP NLF ADPA
kfl00422_0020 KA FDP LN LAEDP-ESLA ELK KEIKNGRLAM AAA GFFAQA TGV GP NLK WL ----
AtLHCB6 RF FDP LN LAEDP-KNFEKLERLK AEIK SRLAM AM IFYF AG-QG TP GALGL ----
Mapoly0052s0039 KF FDP LN LAEDP-VYDKLSRL AEIK SRLAM AM IFYF AG-QGLTP GALGL ----
kfl00276_0010 KF FDP LN LAEDP-GKVFALKERL AEIK ARLAM AM IFY AG-QGLSP SPFGL ----
DAA05919_MvLHCP1.2 EH FDP LN LAEDP-KLDLDR K AEK ARLAM W G MSQA ATN GPYANWQ H A PW
Gene.11156 EH FDP LN LAEDP-KLDLDR K AEK ARLAM W G MSQA ATN GPYANWQ H A PW
DAA05930_MvLHCP1.1 EH FDP LN LAEDP-KLDLDR K AEK ARLAM W G MSQA ATN GPYANWL H A PY
Gene.11308 EH FDP LN LAEDP-KLDLDR K AEK ARLAM W G MSQA ATN GPYANWL H A PY
MpLHCAP2.1 R - FDP LN LAESG--DLE ELK KE K CRL MFAW GCISQG ATQ GP ANWQA H ADPV
MpLHCP1.1 R - FDP LN LAEAG--DLE ELK KE K CRL MFAW GCIAQG ATQ GP ANWQA H ADPV
CmLHC LG F PLNLPNDK-----A D E NNGRLAM GFAGI IHGEF NG MPF Q ----
PyLHC FG FDP LN LAEDP-AAFD F NAE VNGRLTM AIGGLIHQW THCTP Q ----
CrLHCSR1 LG FDP LN LAEDP-EELKT L TKE NNGRLAM AIAAFVQA VEQTEIF LVLRFEKEV
CrLHCSR3 LG FDP LN LAEDP-EELKV TKE NNGRLAM AIAAFVQA VEQTEIF LALRFEKEA
Cs_65904 IG FDP LN LAEDP-KEY L TKE NNGRLAM GIAGFVQA VKK EIFYQ G T KGSS
kfl00478_0030 LG FDP LN LAEDP-AAKK SKE NNGRLAM AIAGFVQA V G G IF LFKRL ----
DAA05932_MvLI818;2 LG DPLGLKP S-KELDE ATKE NNGRLAM GIAGMVQE TQTPIFGQ ----
Gene.11245 LG DPLGLKP S-KELDE ATKE NNGRLAM GIAGMVQE TQTPIFGQ ----
Gene.11340 LG DPLGLKP DP-AALA EL TKE NNGRLAM GIAGMVQEF TQ PIF ----
MpLI818;1 YG FDP LN LAEDP-EWLDIH E NNGRLAM AITAMVAQE NTGINL LADV LG KGGG
MpLI818;2 YN FDP LN LAEDP-VWLN K KE NNGRLAM AISGMVAQE VNG LNL PADVLE GKAG



ABD37880_MvLHCA3 NNNIILT-TFGKIGGSF-----
 Gene.11536 NNNIILT-TFGKIGGSF-----
 AtLHCA3 NNNVLT-SLKFH-----
 Mapoly0066s0050 NNNVLT-NLKIH-----
 kfl00560_0090 ANNIILT-TWGTTPPGL-----
 CrLHCA3 NNNIILT-NFGKLVA-----
 Cs_37969 GNILT-NFGHPAL-----
 MpLHCA3 ANNMLA-GFGAIGGVSP-----
 Ol_K08909 GNLLV-NFQNIIGGVSP-----
 ABD37881_MvLHCA4.1 QNNV S-TEYAPLFFVATVWGLIFRPI
 DAA05923_MvLHCA2 QNNV S-TEYAPLFFVATVWGLIFRPI
 Gene.11715 QNNV S-TEYAPLFFVATVWGLIFRPI
 ABD37884_MvLHCA5.2 NNNFIT-SIANLPNVVGK-----
 DAA05927_MvLHCA9 NNNFIT-SIANLPNVVGK-----
 Gene.10884 NNNFIT-SIANLPNVVGK-----
 CrLHCA9 GKNIITY-YLTHLPETLGSA-----
 Cs_44136 SNNIIG-SIARLPETIGATAPPA----
 MpLHCA9 GCNMAT-NIMHIGSTF-----
 Ol_K08911 GCNMAT-NIMNIPVNLA-----
 CrLHCA2 HNNI T-SSVGPETA VTVALPMIYFPW
 Cs_7387 HNNI A-SKVGPEVT LAVITPIV----
 MpLHCA2 QVNI T-SSVGN EFVAAI IAPCYFRPI
 Ol_PTHR21649 AVNL T-SSVGGEAVAFIAAPTFFRPI
 ABD37882_MvLHCA4.2 HNNVCALLPF-----
 DAA05926_MvLHCA5 HNNVCALLPF-----
 Gene.11585 HNNVCALLPF-----
 DAA05920_MvLHCA4 HANI ALDKSGL-----
 Gene.11349 HANI ALDKSGL-----
 AtLHCA2 HATI A-AFTP K-----
 Mapoly0083s0003 HATV A-ALDKLQ-----
 AtLHCA6 HCNV S-AFTSH-----
 kfl00564_0030 HNTI A-QL-----
 MpLHCA4 HNTI T-NFVPIQF-----
 Ol_K08910 HNNV A-AFIGF-----
 Ol_K08908 QVGV-----
 AtLHCA4 HNTIVQ-TFN-----
 Mapoly0006s0261 HTTIVQ-TLAN-----
 kfl00023_0070 HVTVVS-SIEKFVGS S-----
 CrLHCA4 STTV QNDLARL-----
 Cs_52367 SNNV GIEHARL-----
 CrLHCA5 GNNILK-NIGTCT-----VPHSTIPL
 Cs_61250 GNNITK-NIGTCA-----IPSSTIPL
 CrLHCA6 GTTI S-KAAVVP GQAVAPIPASEIPT
 Cs_46127 GTTI S-KAVVIP GQAIVIPSPSTIPT
 AtLHCA1 HNNIGD---IVIPFN-----
 Mapoly0082s0040 ANNIAN---IIPRSVL-----
 kfl00100_0170 HNTVAD---VFIPRSIL-----
 CrLHCA1 GANFAT-NGISV PFF-----
 Cs_25286 GANFAT-NGVSIPYLT-----
 DAA05929_MvLHCA1 HNNAAAR-----LPL
 Gene.11703 HNNAAAR-----LPL
 CrLHCA7 HVNFAT-NGVSIPIA-----
 Cs_25284 AVTFAT-NGVSLPFVH-----
 CrLHCA8 HVNYAT-NGVSLPFL-----
 Cs_48543 ANNFTT-NGTSLPAQAHPLKP-----
 AtLHCA5 HKTI IQ-TLFTSTS-----
 kfl00214_0160 HVTVAE-TLGRA-----
 MpLHCA1 HVNAAV-N-----
 Cs_66689 GQNI T-QGEKGTAVVAIFAEGA----

<i>kfl00009_0500</i>	HNWY G-N-----
<i>kfl00517_0030</i>	HNWY G-S-----
<i>kfl00009_0510</i>	ANWF SQS-----QL
<i>kfl00517_0040</i>	ANWL NGS-----PI
<i>ABD37890_MvLHCB4</i>	QTIVQ-----
<i>Gene.10541</i>	QTIVQ-----
<i>AtLHCB4.1</i>	HTTIID-TFSSS-----
<i>AtLHCB4.2</i>	HTTIID-TFSSS-----
<i>Mapoly0001s0025</i>	HTTIID-TFSK-----
<i>AtLHCB4.3</i>	-----N-N-----
<i>kfl00258_0020</i>	TNTI D-TLSK-----
<i>CrLHCB4</i>	NGKGL-----
<i>Cs_30169</i>	P-----
<i>Cs_67011</i>	PELVED--IEKAAGVI-----
<i>ABD37885_MvLHCBM1</i>	VNNG A-Y---ATSSGLVG-----
<i>Gene.11306</i>	VNNG A-Y---ATSSGLVG-----
<i>Gene.11314</i>	-----
<i>AtLHCB1.1</i>	NNNA A-F---ATNFVPGK-----
<i>AtLHCB2.1</i>	ANNA S-Y---ATNFVPGN-----
<i>Mapoly0057s0073</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0371s000</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0017</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0018</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0019</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0012</i>	ANNA A-Y---ATNFTP GS-----
<i>Mapoly0199s0014</i>	ANNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0016</i>	ANNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0015</i>	ANNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0013</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0199s0020</i>	VNNA A-Y---ATNFTP GS-----
<i>Mapoly0057s0082</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0057s0083</i>	VNNA A-Y---ATNFTP GN-----
<i>Mapoly0068s0071</i>	ANNA A-Y---ATTFTP GN-----
<i>Mapoly0139s0012</i>	TNNA A-Y---ATAFTP GQ-----
<i>Cs_15915</i>	ANNG A-A---ATKFVP-----
<i>Cs_21950</i>	VNNG A-A---ATKFAP-----
<i>Cs_58975</i>	VNNG A-A---ATKFAP-----
<i>Cs_27246</i>	VNNG A-A---ATKFVPS-----
<i>Cs_35121</i>	VNNG A-A---ATKFVP-----
<i>kfl00098_0080</i>	VNNA A-Y---ATAFTP GQ-----
<i>kfl00098_0090</i>	VNNA A-Y---ATAFTP GQ-----
<i>kfl01434_0010</i>	VNNA A-Y---ATQFAP GQ-----
<i>CrLHCBM1</i>	TNNA A-Y---ATKFTP Q-----
<i>CrLHCBM2</i>	AVNA A-Y---ATKFTP SA-----
<i>CrLHCBM7</i>	AVNA A-Y---ATKFTP SA-----
<i>CrLHCBM3</i>	VNNA A-F---ATKFTP SA-----
<i>CrLHCBM4</i>	VNNA A-F---ATKFTP SA-----
<i>CrLHCBM8</i>	VNNA A-F---ATKFTP SA-----
<i>CrLHCBM6</i>	VNNA A-F---ATKFTP SA-----
<i>CrLHCBM9</i>	VNNA A-F---ATKYTP SA-----
<i>kfl00809_0010</i>	VNNA A-Y---ATAFVPGQ-----
<i>CrLHCBM5</i>	TNA T-Y---AQKFTP Q-----
<i>Cs_38026</i>	AVNG T-I--GAQKFVPGN-----
<i>Mapoly0026s002</i>	INNA A-Y---ATNFVPRS-----
<i>AtLHCB3</i>	ANNA A-F---ATKFAP GA-----
<i>Mapoly0068s0047</i>	ANNA A-Y---ATNFVPGA-----
<i>Mapoly0068s0087</i>	VNNA A-Y---ATNFVVR-----
<i>kfl00120_0160</i>	ANNG A-Y---ATAFSP GQ-----
<i>Cs_28488</i>	VNNG A-A---ATKFTP-----

ABD37891_MvLHCB5 GY N FLT - I L G S G S E R V P T L - - - - -
Gene.11404 GY N FLT - I L G S G S E R V P T L - - - - -
DAA05928_MvLHCB5 GY N FLT - I L G S G S E R V P T L - - - - -
Cs_64185 GY N L L T - - I I G A E D R V P T L - - - - -
AtLHCB5 G N N L L T - V I A G T A E R A P T L - - - - -
Mapoly0011s0076 G N N L L T - V L Q G S A E R V P S L - - - - -
kfl00104_0350 A N N I I S - V I G G N I E R S P V L - - - - -
AtLHCB7 H N N L I A - M L Q T - - - - -
Mapoly0008s0015 H Q N L A - Y A T S S - - - - -
kfl00104_0370 H N N I T - A F N S S - - - - -
CrLHCB7 H N N I - - N L A H L Q - - - - -
kfl00422_0020 - - - - - H - - - - -
AtLHCB6 - - - - -
Mapoly0052s0039 - - - - -
kfl00276_0010 - - - - -
DAA05919_MvLHCP1.2 V E N V K Y A - - - - - F N Q -
Gene.11156 V E N V K Y A - - - - - F N Q -
DAA05930_MvLHCP1.1 V E N V K Y T - - - - - F Q - -
Gene.11308 V E N V K Y T - - - - - F Q - -
MpLHCAP2.1 H A N V L T - N A A S G F G F Y - - - - -
MpLHCP1.1 H A N V L T - N A A S G F G F Y - - - - -
CmLHC - - - - - T - N - - - - - F Q P L
PyLHC - - - - - L S G K L F P - - - - -
CrLHCSR1 I L E L E G L P L T P L P D N L K A I - - - - -
CrLHCSR3 I L E L D G L P V T P L P D N L K S L - - - - -
Cs_65904 - - - S P - - - - -
kfl00478_0030 G L - - - - -
DAA05932_MvLI818;2 - - - - -
Gene.11245 - - - - -
Gene.11340 - - - - -
MpLI818;1 A L K A M S L D E A A C S K A F E A A T A V L - - - -
MpLI818;2 A L E A M A T N E A A C A K A F E A A V F A A - - - -

Fig. S2

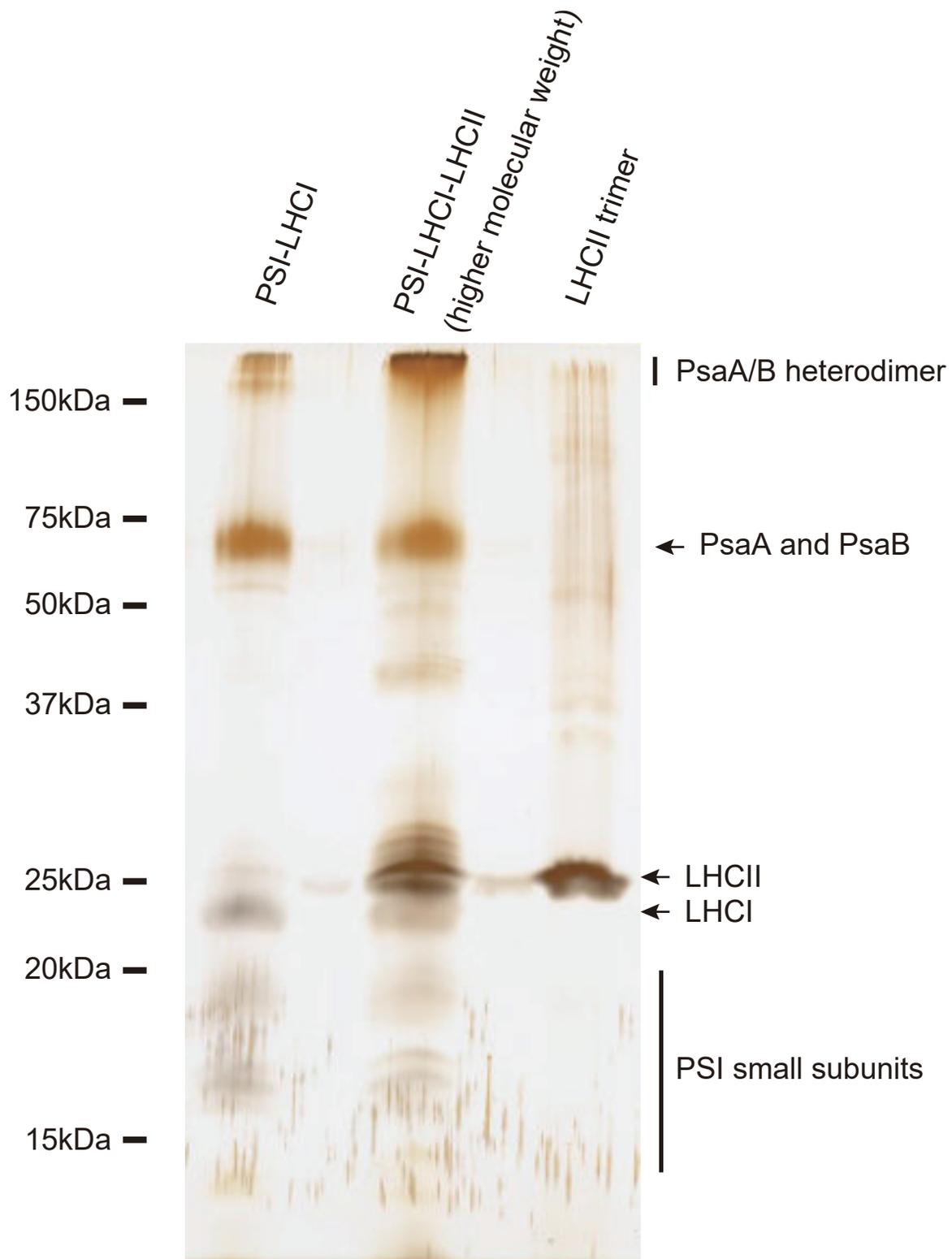


Figure S2. 2D-SDS-PAGE of the high-molecular-weight PSI-LHCI-LHCII, PSI-LHCI, and LHCII trimer.

The high-molecular-weight PSI-LHCI-LHCII, PSI-LHCI, and LHCII trimer bands separated by CN-PAGE after sucrose density gradient centrifugation (Fig. 7) were subjected to 2D-SDS-PAGE. Protein bands were visualized by silver-staining.

Table S1-1. The list of the PSI proteins predicted by the Iso-seq analysis in this study

	Gene annotation in this study	Accession No.	The best-hit gene in <i>A.thaliana</i>			The best-hit gene in <i>C.reinhardtii</i>		
Gene3244	PSAA	ICQU01000001	ATCG00350.1	PSAA	Photosystem I, PsaA/PsaB protein	NP_958375.1	PSAA	Photosystem I P700 chlorophyll a apoprotein A1
Gene3504	PSAB	ICQU01000002	ATCG00340.1	PSAB	Photosystem I, PsaA/PsaB protein	NP_958404.1	PSAB	Photosystem I P700 chlorophyll a apoprotein A2
Gene9022	PSAD	ICQU01000006	AT1G03130.1	PSAD-2	photosystem I subunit D-2	Cre05.g238332.t1.1	PSAD	Photosystem I reaction center subunit II, 20 kDa
Gene11867	PSAE	ICQU01000029	AT2G20260.1	PSAE-2	photosystem I subunit E-2	Cre10.g420350.t1.2	PSAE	Photosystem I 8.1 kDa reaction center subunit IV
Gene11501	PSAF	ICQU01000021	AT1G31330.1	PSAF	photosystem I subunit F	Cre09.g412100.t1.2	PSAF	Photosystem I reaction center subunit III
Gene11916	PSAG	ICQU01000031	AT1G55670.1	PSAG	photosystem I subunit G	Cre12.g560950.t1.2	PSAG	Photosystem I reaction center subunit V
Gene11865	PSAH	ICQU01000028	AT1G52230.1	PSAH-2	photosystem I subunit H2	Cre07.g330250.t1.2	PSAH	Subunit H of photosystem I
Gene11650	PSAL	ICQU01000024	AT4G12800.1	PSAL	photosystem I subunit I	Cre12.g486300.t1.2	PSAL	Photosystem I reaction center subunit XI
Gene11942	PSAO	ICQU01000034	AT1G08380.1	PSAO	photosystem I subunit O	Cre07.g334550.t1.2	PSA01	Photosystem I subunit O

Table S1-2. The list of the PSII proteins predicted by the Iso-seq analysis in this study

	Gene annotation in this study	Accession No.	The best-hit gene in <i>A.thaliana</i>			The best-hit gene in <i>C.reinhardtii</i>		
Gene10292	PSBA	ICQU01000008	ATCG00020.1	PSBA	photosystem II reaction center protein A	NP_958377.1	PSBA	photosystem II protein D1
Gene4825	PSBB	ICQU01000003	ATCG00680.1	PSBB	photosystem II reaction center protein B	NP_958388.1	PSBB	photosystem II 47 kDa protein
Gene6781	PSBC	ICQU01000005	ATCG00280.1	PSBC	photosystem II reaction center protein C	NP_958422.1	PSBC	photosystem II 44 kDa protein
Gene4841	PSBD	ICQU01000004	ATCG00270.1	PSBD	photosystem II reaction center protein D	NP_958420.1	PSBD	photosystem II protein D2
Gene10723	PSBO	ICQU01000010	AT3G50820.1	PSBO-2	photosystem II subunit O-2	Cre09.g396213.t1.1	PSBO	Oxygen-evolving enhancer protein 1 of photosystem II
Gene11499	PSBP	ICQU01000020	AT1G06680.1	PSBP-1	photosystem II subunit P-1	Cre12.g550850.t1.2	PSBP1	Oxygen-evolving enhancer protein 2 of photosystem II
Gene11808	PSBQ	ICQU01000027	AT4G05180.1	PSBQ-2	photosystem II subunit Q-2	Cre08.g372450.t1.2	PSBQ	Oxygen evolving enhancer protein 3
Gene11938	PSBR	ICQU01000033	AT1G79040.1	PSBR	photosystem II subunit R	Cre06.g261000.t1.2	PSBR	10 kDa photosystem II polypeptide
Gene11934	PSBW	ICQU01000032	AT2G30570.1	PSBW	photosystem II reaction center W			
Gene11880	PSBX	ICQU01000030	AT2G06520.1	PSBX	photosystem II subunit X			
Gene9164	PSBY	ICQU01000007	AT1G67740.1	PSBY	photosystem II BY	Cre10.g452100.t1.1	PSBY	Ycf32-related polyprotein of photosystem II

Table S1-3. The list of the LHC proteins predicted by the Iso-seq analysis in this study

	Gene annotation in this study	Accession No.	The best-hit gene in <i>A.thaliana</i>			The best-hit gene in <i>C.reinhardtii</i>		
Gene11245	LHCSR	ICQU01000013	AT1G15820.1	LHCB6	light harvesting complex photosystem II subunit 6	Cre08.g367500.t1.1	LHCSR3.1	Stress-related chlorophyll a/b binding protein 2
Gene11340	LHCSR	ICQU01000017	AT3G61470.1	LHCA2	photosystem I light harvesting complex gene 2	Cre08.g367500.t1.1	LHCSR3.1	Stress-related chlorophyll a/b binding protein 2
Gene11306	LHCBM	ICQU01000014	AT2G05100.1	LHCB2.1	photosystem II light harvesting complex gene 2.1	Cre01.g066917.t1.1	LHCBM1	Chlorophyll a/b binding protein of LHCII
Gene11314	LHCBM	ICQU01000016	AT2G05070.1	LHCB2.2	photosystem II light harvesting complex gene 2.2	Cre01.g066917.t1.1	LHCBM1	Chlorophyll a/b binding protein of LHCII
Gene10541	LHCB4	ICQU01000009	AT2G40100.1	LHCB4.3	light harvesting complex photosystem II	Cre17.g720250.t1.2	LHCB4	Chlorophyll a/b binding protein of photosystem II
Gene11404	LHCB5	ICQU01000019	AT4G10340.1	LHCB5	light harvesting complex of photosystem II 5	Cre16.g673650.t1.1	LHCB5	Minor chlorophyll a/b binding protein of photosystem II
Gene11156	LHCP	ICQU01000012	AT5G54270.1	LHCB3	light-harvesting chlorophyll B-binding protein 3	Cre03.g156900.t1.2	LHCBM5	Chlorophyll a/b binding protein of LHCII
Gene11308	LHCP	ICQU01000015	AT5G54270.1	LHCB3	light-harvesting chlorophyll B-binding protein 3	Cre04.g232104.t1.1	LHCBM3	Light-harvesting complex II chlorophyll a/b binding protein M3
Gene11703	LHCA1	ICQU01000025	AT3G54890.1	LHCA1	photosystem I light harvesting complex gene 1	Cre06.g283050.t1.2	LHCA1	Light-harvesting protein of photosystem I
Gene11349	LHCA2	ICQU01000018	AT3G61470.1	LHCA2	photosystem I light harvesting complex gene 2	Cre16.g687900.t1.2	LHCA7	Light-harvesting protein of photosystem I
Gene11585	LHCA2	ICQU01000023	AT3G47470.1	LHCA4	light-harvesting chlorophyll-protein complex I subunit A4	Cre10.g452050.t1.2	LHCA4	Light-harvesting protein of photosystem I
Gene11536	LHCA3	ICQU01000022	AT1G61520.1	LHCA3	photosystem I light harvesting complex gene 3	Cre11.g467573.t1.1	LHCA3	Chlorophyll a/b binding protein of photosystem I, type III
Gene11715	algae-type LHCA2	ICQU01000026	AT1G45474.1	LHCA5	photosystem I light harvesting complex gene 5	Cre12.g508750.t1.2	LHCA2	Light-harvesting protein of photosystem I
Gene10884	LHCA9	ICQU01000011	AT1G45474.1	LHCA5	photosystem I light harvesting complex gene 5	Cre07.g344950.t1.2	LHCA9	Light-harvesting protein of photosystem I

Table S2. The identified proteins in the PSI-PSII band by MS

Gene ID	NSAF (fmol)	Category	Gene annotation in this study	Accession No.	Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
					Gene ID	Gene Name	Gene ID	Gene Name
Gene11306	4526	LHC	LHCBM	ICQU01000014	AT2G05100.1	LHCB2.1	Cre01.g066917.t1.1	LHCBM1
Gene10292	4034	PSII	PSBA	ICQU01000008	ATCG00020.1	PSBA	NP_958377.1	PSBA
Gene11501	3411	PSI	PSAF	ICQU01000021	AT1G31330.1	PSAF	Cre09.g412100.t1.2	PSAF
Gene11156	3322	LHC	LHCP	ICQU01000012	AT5G54270.1	LHCB3	Cre03.g156900.t1.2	LHCBM5
Gene11585	2942	LHC	LHCA2	ICQU01000023	AT3G47470.1	LHCA4	Cre10.g452050.t1.2	LHCA4
Gene4825	2883	PSII	PSBB	ICQU01000003	ATCG00680.1	PSBB	NP_958388.1	PSBB
Gene11867	2773	PSI	PSAE	ICQU01000029	AT2G20260.1	PSAE-2	Cre10.g420350.t1.2	PSAE
Gene11865	2654	PSI	PSAH	ICQU01000028	AT1G52230.1	PSAH-2	Cre07.g330250.t1.2	PSAH
Gene4841	2605	PSII	PSBD	ICQU01000004	ATCG00270.1	PSBD	NP_958420.1	PSBD
Gene3547	2468	Others			AT5G59970.1		Cre12.g506350.t1.2	HF018
Gene11916	2444	PSI	PSAG	ICQU01000031	AT1G55670.1	PSAG	Cre12.g560950.t1.2	PSAG
Gene9022	2440	PSI	PSAD	ICQU01000006	AT1G03130.1	PSAD-2	Cre05.g238332.t1.1	PSAD
Gene11404	2412	LHC	LHCB5	ICQU01000019	AT4G10340.1	LHCB5	Cre16.g673650.t1.1	LHCB5
Gene10541	2301	LHC	LHCB4	ICQU01000009	AT2G40100.1	LHCB4.3	Cre17.g720250.t1.2	LHCB4
Gene11715	2235	LHC	algae-type LHCA2	ICQU01000026	AT1G45474.1	LHCA5	Cre12.g508750.t1.2	LHCA2
Gene6781	1782	PSII	PSBC	ICQU01000005	ATCG00280.1	PSBC	NP_958422.1	PSBC
Gene11340	1707	LHC	LHCSR	ICQU01000017	AT3G61470.1	LHCA2	Cre08.g367500.t1.1	LHCSR3.1
Gene6617	1577	Others			AT3G61320.1		Cre06.g261750.t1.2	
Gene3504	1569	PSII	PSAB	ICQU01000002	ATCG00340.1	PSAB	NP_958404.1	PSAB
Gene7944	1382	Others			AT3G61320.1		Cre06.g261750.t1.2	
Gene11703	1355	LHC	LHCA1	ICQU01000025	AT3G54890.1	LHCA1	Cre06.g283050.t1.2	LHCA1
Gene10723	1232	PSII	PSBO	ICQU01000010	AT3G50820.1	PSBO-2	Cre09.g396213.t1.1	PSBO
Gene11536	1228	LHC	LHCA3	ICQU01000022	AT1G61520.1	LHCA3	Cre11.g467573.t1.1	LHCA3
Gene10884	1219	LHC	LHCA9	ICQU01000011	AT1G45474.1	LHCA5	Cre07.g344950.t1.2	LHCA9
Gene7515	1169	Others			AT5G04180.1	ATACA3	Cre09.g415700.t1.2	CAH3
Gene11308	1066	LHC	LHCP	ICQU01000015	AT5G54270.1	LHCB3	Cre04.g232104.t1.1	LHCBM3
Gene7423	858	Others						
Gene8340	839	Others			AT5G17170.1	ENH1	Cre12.g510400.t1.1	CPLD30
Gene9362	780	Others			AT4G38970.1	FBA2	Cre05.g234450.t1.2	FBA3
Gene8224	529	Others			AT5G23860.1	TUB8	Cre12.g549550.t1.2	TUB2
Gene8089	512	Others			ATCG00490.1	RBCL	NP_958405.1	RBCL

Table S3. The identified proteins in the PSII-LHCII band by MS

Gene ID	NSAF(fmol)	Category	Gene annotation in this study	Accession No.	Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
					Gene ID	Gene Name	Gene ID	Gene Name
Gene11306	7753	LHC	LHCBM	ICQU01000014	AT2G05100.1	LHCB2.1	Cre01.g066917.t1.1	LHCBM1
Gene10292	6458	PSII	PSBA	ICQU01000008	ATCG00020.1	PSBA	NP_958377.1	PSBA
Gene10541	3371	LHC	LHCB4	ICQU01000009	AT2G40100.1	LHCB4.3	Cre17.g720250.t1.2	LHCB4
Gene4825	3021	PSII	PSBB	ICQU01000003	ATCG00680.1	PSBB	NP_958388.1	PSBB
Gene11404	1991	LHC	LHCB5	ICQU01000019	AT4G10340.1	LHCB5	Cre16.g673650.t1.1	LHCB5
Gene11865	1757	PSI	PSAH	ICQU01000028	AT1G52230.1	PSAH-2	Cre07.g330250.t1.2	PSAH
Gene6781	1495	PSII	PSBC	ICQU01000005	ATCG00280.1	PSBC	NP_958422.1	PSBC
Gene11938	1341	PSII	PSBR	ICQU01000033	AT1G79040.1	PSBR	Cre06.g261000.t1.2	PSBR
Gene3547	1317	Others			AT5G59970.1		Cre12.g506350.t1.2	HF018
Gene11585	1159	LHCI	LHCA2	ICQU01000023	AT3G47470.1	LHCA4	Cre10.g452050.t1.2	LHCA4
Gene11867	1142	PSI	PSAE	ICQU01000029	AT2G20260.1	PSAE-2	Cre10.g420350.t1.2	PSAE
Gene11916	1105	PSI	PSAG	ICQU01000031	AT1G55670.1	PSAG	Cre12.g560950.t1.2	PSAG
Gene4841	1095	PSII	PSBD	ICQU01000004	ATCG00270.1	PSBD	NP_958420.1	PSBD
Gene7779	1091	Others			ATCG00430.1	PSBG	Cre12.g492300.t1.2	NUO10
Gene11308	1084	LHC	LHCP	ICQU01000015	AT5G54270.1	LHCB3	Cre04.g232104.t1.1	LHCBM3
Gene9022	1080	PSI	PSAD	ICQU01000006	AT1G03130.1	PSAD-2	Cre05.g238332.t1.1	PSAD
Gene11715	976	LHC	algae-type LHCA2	ICQU01000026	AT1G45474.1	LHCA5	Cre12.g508750.t1.2	LHCA2
Gene10884	937	LHC	LHCA9	ICQU01000011	AT1G45474.1	LHCA5	Cre07.g344950.t1.2	LHCA9
Gene11501	926	PSI	PSAF	ICQU01000021	AT1G31330.1	PSAF	Cre09.g412100.t1.2	PSAF
Gene11349	902	LHC	LHCA2	ICQU01000018	AT3G61470.1	LHCA2	Cre16.g687900.t1.2	LHCA7
Gene11703	858	LHC	LHCA1	ICQU01000025	AT3G54890.1	LHCA1	Cre06.g283050.t1.2	LHCA1
Gene7515	822	Others			AT5G04180.1	ATACA3	Cre09.g415700.t1.2	CAH3
Gene11156	789	LHC	LHCP	ICQU01000012	AT5G54270.1	LHCB3	Cre03.g156900.t1.2	LHCBM5
Gene2597	781	Others			ATCG01100.1	NDHA		
Gene11934	741	PSII	PSBW	ICQU01000032	AT2G30570.1	PSBW		
Gene4851	714	Others			ATCG01110.1	NDHH	Cre09.g405850.t1.1	NUO7
Gene11718	690	Others			ATCG00420.1	NDHJ	Cre07.g327400.t1.1	NUO9
Gene9021	683	Others			ATCG01070.1	NDHE	Cre09.g402552.t1.1	NUO11
Gene11761	676	Others						
Gene6617	618	Others			AT3G61320.1		Cre06.g261750.t1.2	
Gene11650	605	PSI	PSAL	ICQU01000024	AT4G12800.1	PSAL	Cre12.g486300.t1.2	PSAL
Gene8813	572	Others			AT1G20020.1	ATLFNR2	Cre11.g476750.t1.2	FNR1
Gene11139	565	Others			AT2G28720.1		Cre13.g590750.t1.2	HTB21
Gene10723	565	PSII	PSBO	ICQU01000010	AT3G50820.1	PSBO-2	Cre09.g396213.t1.1	PSBO
Gene11808	539	PSII	PSBQ	ICQU01000027	AT4G05180.1	PSBQ-2	Cre08.g372450.t1.2	PSBQ
Gene8340	531	Others			AT5G17170.1	ENH1	Cre12.g510400.t1.1	CPLD30
Gene10856	527	Others			AT5G58260.1			
Gene11340	526	LHC	LHCSR	ICQU01000017	AT3G61470.1	LHCA2	Cre08.g367500.t1.1	LHCSR3.1
Gene5104	522	Others			AT1G15980.1	NDF1		

Table S4. The identified proteins in the PSI-LHCI band by MS

Gene ID	NSAF(fmol)	Category	Gene annotation in this study	Accession No.	Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
					Gene ID	Gene Name	Gene ID	Gene Name
Gene11703	9063	LHC	LHCA1	ICQU01000025	AT3G54890.1	LHCA1	Cre06.g283050.t1.2	LHCA1
Gene11916	6111	PSI	PSAG	ICQU01000031	AT1G55670.1	PSAG	Cre12.g560950.t1.2	PSAG
Gene11867	5182	PSI	PSAE	ICQU01000029	AT2G20260.1	PSAE-2	Cre10.g420350.t1.2	PSAE
Gene11501	5101	PSI	PSAF	ICQU01000021	AT1G31330.1	PSAF	Cre09.g412100.t1.2	PSAF
Gene11536	4709	LHC	LHCA3	ICQU01000022	AT1G61520.1	LHCA3	Cre11.g467573.t1.1	LHCA3
Gene10884	4541	LHC	LHCA9	ICQU01000011	AT1G45474.1	LHCA5	Cre07.g344950.t1.2	LHCA9
Gene11865	2669	PSI	PSAH	ICQU01000028	AT1G52230.1	PSAH-2	Cre07.g330250.t1.2	PSAH
Gene11349	2593	LHC	LHCA2	ICQU01000018	AT3G61470.1	LHCA2	Cre16.g687900.t1.2	LHCA7
Gene9022	2205	PSI	PSAD	ICQU01000006	AT1G03130.1	PSAD-2	Cre05.g238332.t1.1	PSAD
Gene11585	1980	LHC	LHCA2	ICQU01000023	AT3G47470.1	LHCA4	Cre10.g452050.t1.2	LHCA4
Gene11715	1861	LHC	algae-type LHCA2	ICQU01000026	AT1G45474.1	LHCA5	Cre12.g508750.t1.2	LHCA2
Gene11306	1695	LHC	LHCBM	ICQU01000014	AT2G05100.1	LHCB2.1	Cre01.g066917.t1.1	LHCBM1
Gene3547	1333	Others			AT5G59970.1		Cre12.g506350.t1.2	HFO18
Gene3504	1286	PSI	PSAB	ICQU01000002	ATCG00340.1	PSAB	NP_958404.1	PSAB
Gene11650	1115	PSI	PSAL	ICQU01000024	AT4G12800.1	PSAL	Cre12.g486300.t1.2	PSAL
Gene11761	914	Others						
Gene11139	868	Others			AT2G28720.1		Cre13.g590750.t1.2	HTB21
Gene4841	822	PSII	PSBD	ICQU01000004	ATCG00270.1	PSBD	NP_958420.1	PSBD
Gene3244	820	PSI	PSAA	ICQU01000001	ATCG00350.1	PSAA	NP_958375.1	PSAA
Gene10292	783	PSII	PSBA	ICQU01000008	ATCG00020.1	PSBA	NP_958377.1	PSBA
Gene9567	782	Others						
Gene6617	765	Others			AT3G61320.1		Cre06.g261750.t1.2	
Gene7515	730	Others			AT5G04180.1	ATACA3	Cre09.g415700.t1.2	CAH3
Gene4825	682	PSII	PSBB	ICQU01000003	ATCG00680.1	PSBB	NP_958388.1	PSBB
Gene10541	675	LHC	LHCB4	ICQU01000009	AT2G40100.1	LHCB4.3	Cre17.g720250.t1.2	LHCB4
Gene5134	575	Others						
Gene9362	538	Others			AT4G38970.1	FBA2	Cre05.g234550.t1.2	FBA3
Gene6781	535	PSII	PSBC	ICQU01000005	ATCG00280.1	PSBC		