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

Cover page

Title: Unique peripheral antennas in the photosystems of the streptophyte alga *Mesostigma viride*.

Running Title: Unique Photosystems in *Mesostigma viride*

Corresponding Author: Atsushi Takabayashi

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

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

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

Subject Area: (5) photosynthesis

Number of black and white figures: 0

Number of color figures: 6

Number of tables: 0

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

Title: Unique peripheral antennas **in the photosystems of the streptophyte alga** ***Mesostigma viride.***

Running Title: The Unique Photosystems in *Mesostigma viride*

Michiki Aso[1], Renon Matsumae[1], Ayumi Tanaka[1], Ryouichi Tanaka[1], Atsushi Takabayashi[1]

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

Abbreviations: CN, clear-native; α-DDM, α-dodecyl maltoside; Isoseq, isoform sequencing; LHC, light-harvesting complex; PAGE, polyacrylamide gel electrophoresis

[Abstract]

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

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

[Introduction]

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

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

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

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

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

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

[Results]

Separation of photosystems by ClearNative (CN)-PAGE

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

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

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

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

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

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

Phylogenetic analysis of LHC proteins

A BLAST query identified 14 putative *M. viride* LHC proteins (Table S1). The annotation of the LHC proteins derived from the BLAST query, however, may not be reliable as considerable sequence similarity has been reported among different LHC proteins. Therefore, a phylogenetic tree of LHC sequences was constructed based on the alignment shown in Fig. S1 and each LHC sequence of *M. viride* was annotated based on the nomenclature of *A. thaliana* LHC proteins, except for LHCBMs (corresponding 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*

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

The distribution of LHC proteins in PSI and PSII

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

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

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

LHCB5, and LHCP proteins, whose peaks overlapped with the peak of PSII-LHCII (Fig. 3C). These data strongly suggest that PSI-LHCI includes the former LHCs, whereas PSII-LHCII includes the latter LHCs. Notably, the migration profile of LHCBM proteins appeared to be different from the profiles of other LHCs. Its highest peak corresponded with LHCII trimers, suggesting that LHCBM is the major component of the LHCII trimer.

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

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

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

[Discussion]

Unique peripheral antenna systems of *M. viride*

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

M. viride PSII possesses LHCP (Koziol et al., 2007), the main peripheral antenna protein in the model prasinophyte alga, *O. tauri* (Six et al., 2005; Swingley et al., 2010), whereas LHCP has not been found in land plants. The peaks in the migration profile of LHCP corresponded with the PSI-PSII megacomplex and the PSII-LHCII (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

is possible to hypothesize that LHCP in streptophyte PSII-LHCII have been replaced by LHCB6 in land plants. Further investigation will be required, however, to confirm this hypothesis.

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

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

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

We constructed a hypothetical model of the antenna structure of PSI-LHCI-LHCII in *M. viride* using the reported structure of PSI-LHCI in chlorophyte algae and land plants as references (Fig. 6). Based on their sequence similarities, the binding manner of LHCA1, two LHCA2s, and LHCA3 proteins to the PSI core in *M. viride* may be similar to the binding of LHCA1, LHCA2, LHCA3, and LHCA4 proteins to the PSI core in vascular plants. Likewise, the binding manner of algae-type LHCA2 and LHCA2 to PSI-LHCI may be similar to the binding mechanism that occurs in PSI-LHCI in *C. reinhardtii* (Su et al., 2019; Suga et al., 2019) and *B. corticulans* (Qin et al., 2019). The binding site of the second LHCII trimer to PSI-LHCI, however, is unknown, as a PSI-LHCI with two LHCII trimers has not, to our knowledge, been reported. Similarly, the binding site of LHCP to PSII-LHCII is also unknown as no structural studies on LHCP-bound PSI-LHCI have been reported. Given the phylogenetic position of *M. viride* and its unique antenna system in PSI and PSII, further structural analyses of *M. 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 μmol photons m^{-2} 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 21,600×g at 4 °C for 5 min. Supernatants from the

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

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

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

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

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

Thylakoid membrane preparation

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

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

Clear-Native (CN)-PAGE

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

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

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

Immunoblot analysis

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

Sucrose density gradient

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

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

LC-MS/MS analysis

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

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

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

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

[Acknowledgement]

Funding

This work was supported by JSPS (Japan Society for the Promotion of Science) KAKENHI Grant numbers 16H06554 to R. Tanaka.

Data availability

Raw data of the Iso-seq analysis have been deposited to the DDBJ Sequence Read Archive (DRA) (https://www.ddbj.nig.ac.jp/dra/index.html) under the accession number PRJDB10366. In addition, the assembled full-length cDNA sequences of PSI, PSII, and

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

Disclosures

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

[References]

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

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

Correa-Galvis, V., Redekop, P., Guan, K., Griess, A., Truong, T.B., Wakao, S., et al. (2016) Photosystem II Subunit PsbS Is Involved in the Induction of LHCSR Protein-dependent Energy Dissipation in *Chlamydomonas reinhardtii*. *J Biological Chem*. 291: 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 bative-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., Valkenborg, 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 Zybailov, B., Mosley, A.L., Sardiu, 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

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

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

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

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

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

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

Figure 5. Separation of two PSI-LHCI-LHCII bands and a PSI-LHCI band by 2D-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

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

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.

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 were subjected to 2D-SDS-PAGE. Protein bands were visualized by silver-staining.

Supplementary Table Legends:

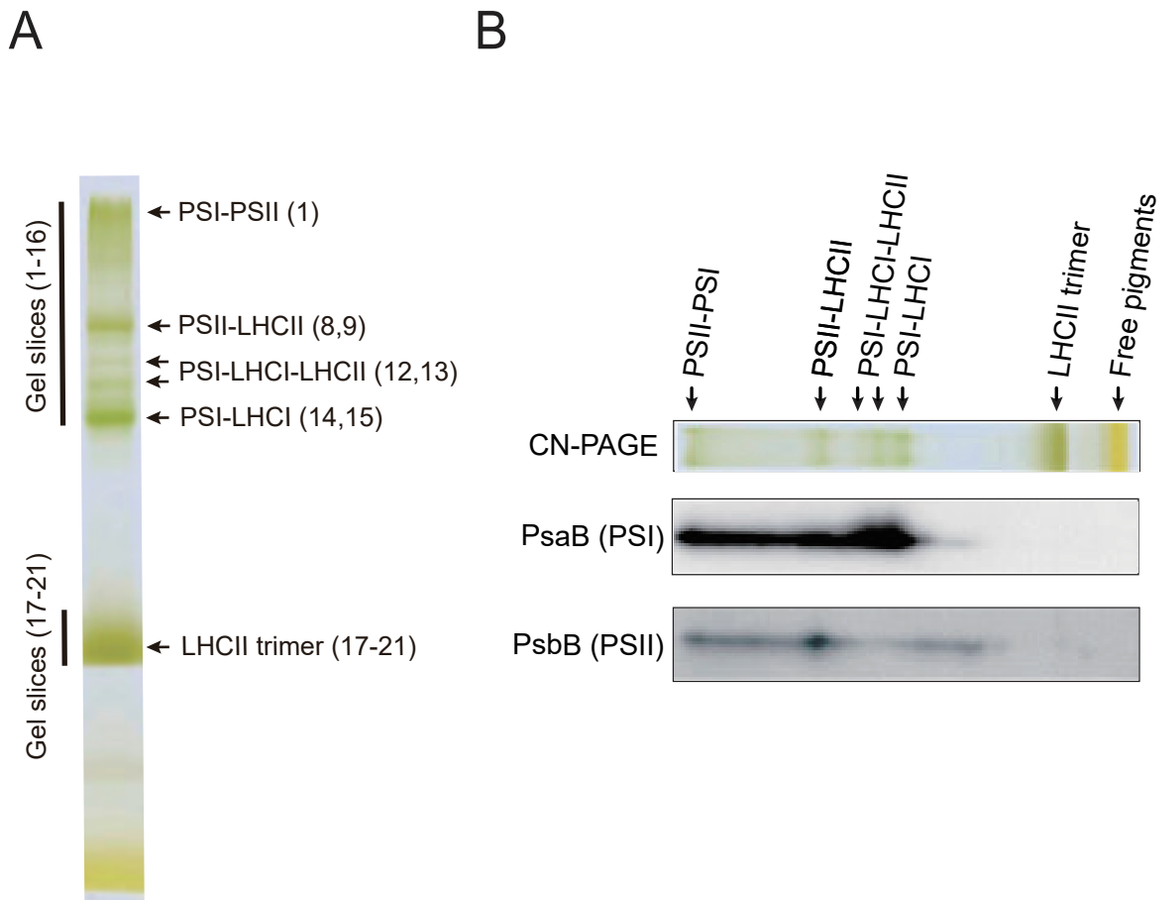
Table S1. List of *M. viride* LHC proteins predicted in the Iso-seq analysis.

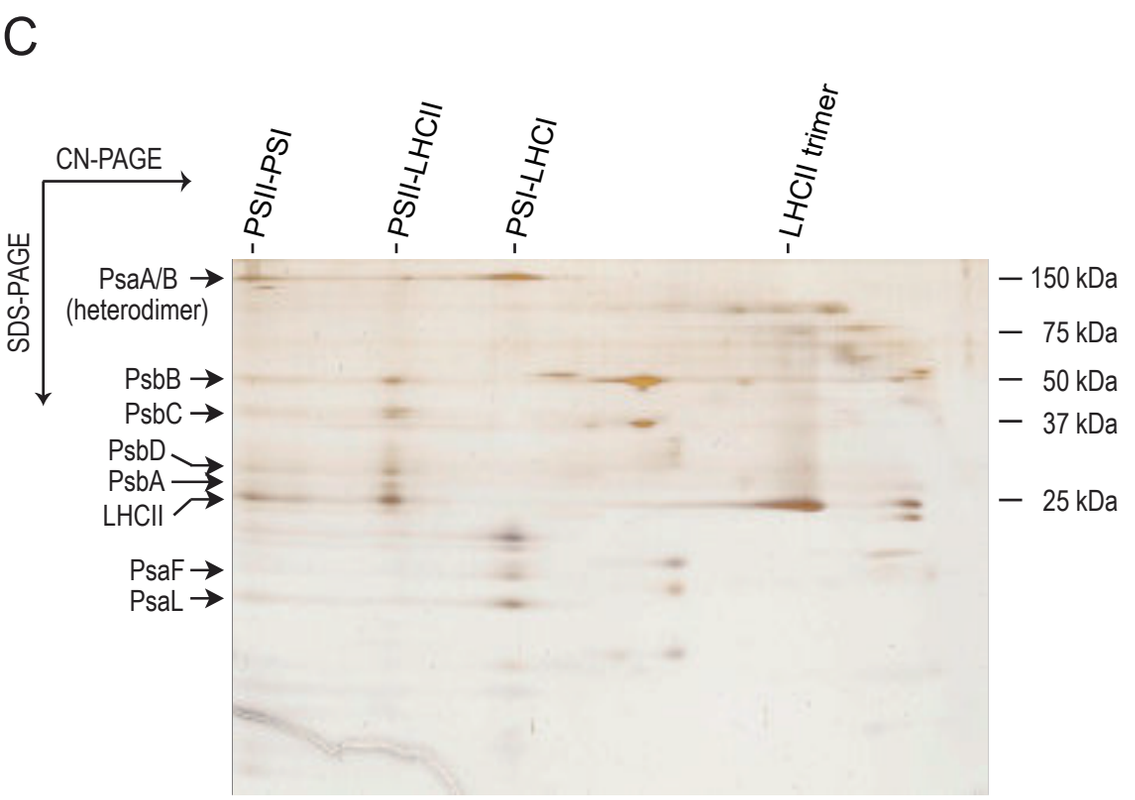
Table S2. Proteins identified in the *M. viride* PSI-PSII band by MS.

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

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

Fig. 1





Aso et al.

Fig. 2

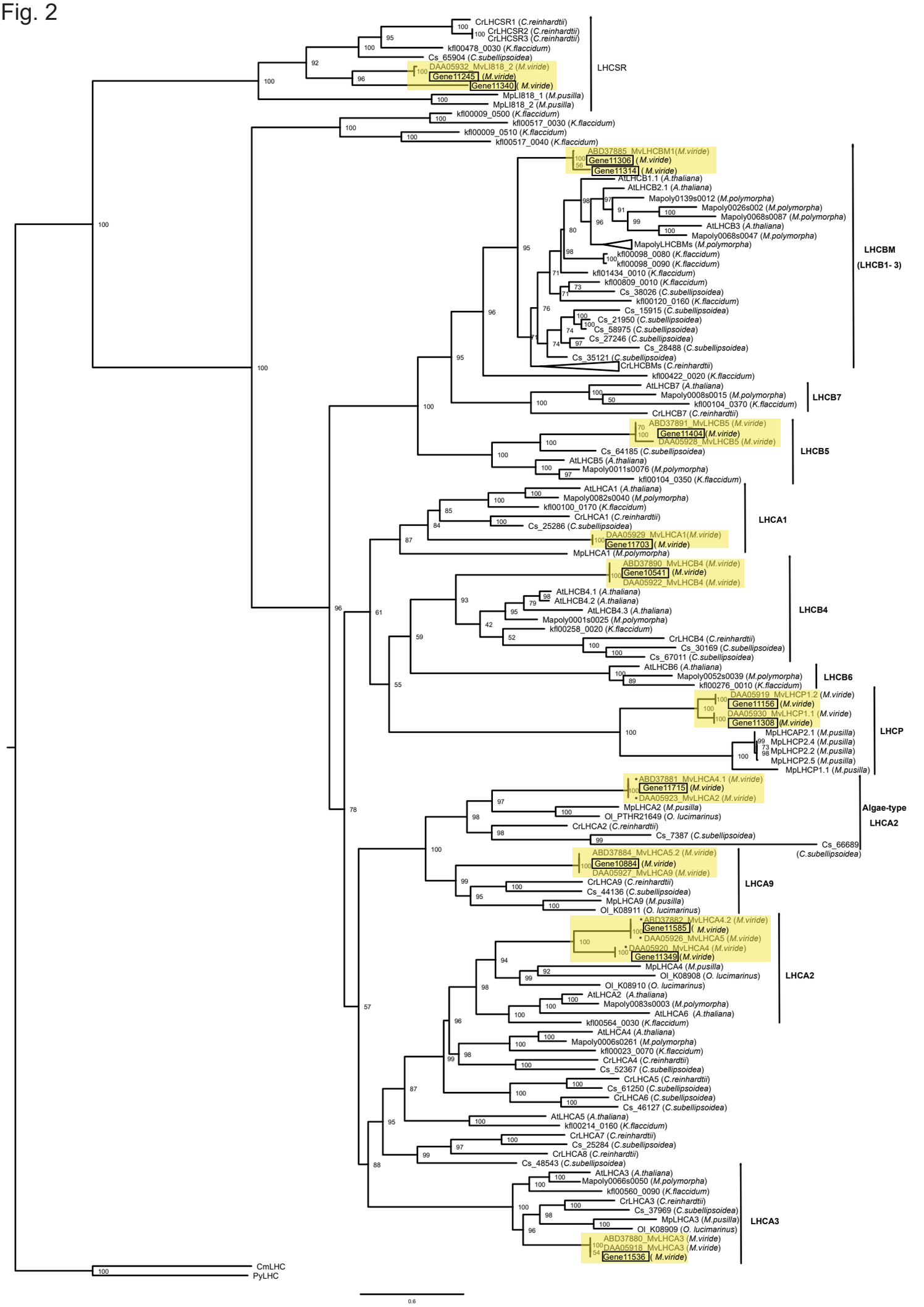
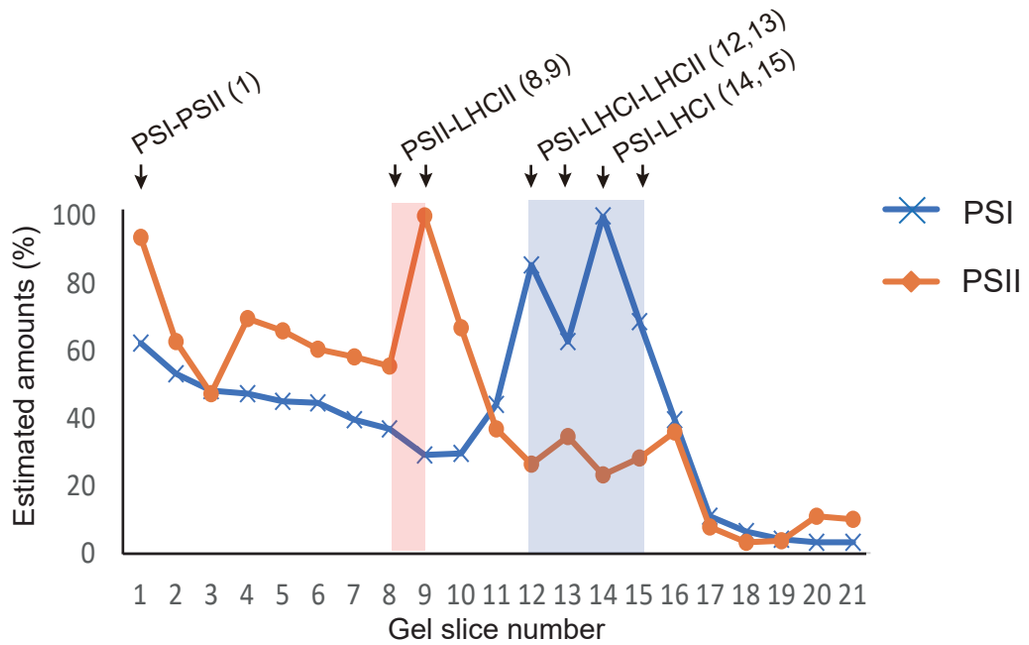
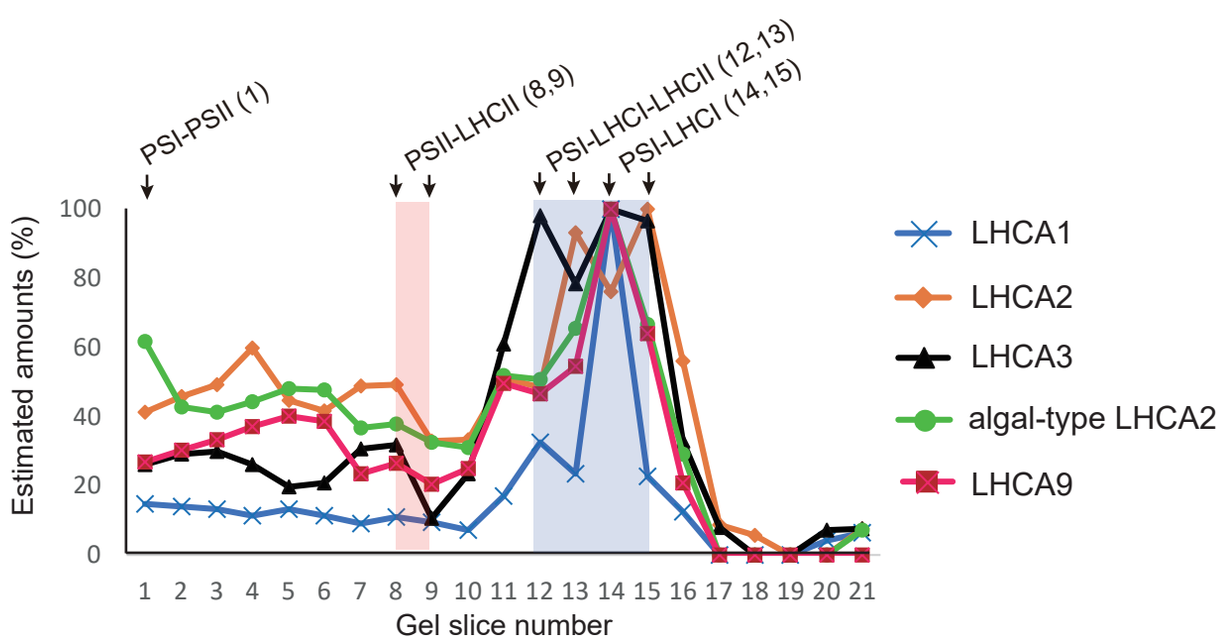

CrLHCSR1 (C.reinhardtii)
CrLHCSR2 (C.reinhardtii)
CrLHCSR3 (C.reinhardtii)
kfl00478_0030 (K.flaccidum)
Cs_65904 (C.subellipsoidea)
DAA05932_MvLI818_2 (M.viride)
Gene11245 (M.viride)
Gene11340 (M.viride)
MpLI818_1 (M.pusilla)
MpLI818_2 (M.pusilla)
LHCSR
kfl00009_0500 (K.flaccidum)
kfl00517_0030 (K.flaccidum)
kfl00009_0510 (K.flaccidum)
kfl00517_0040 (K.flaccidum)
ABD37885_MvLHCBM1(M.viride)
Gene11306 (M.viride)
Gene11314 (M.viride)
AtLHCB1.1 (A.thaliana)
AtLHCB2.1 (A.thaliana)
Mapoly0139s0012 (M.polymorpha)
Mapoly0026s002 (M.polymorpha)
Mapoly0068s0087 (M.polymorpha)
AtLHCB3 (A.thaliana)
Mapoly0068s0047 (M.polymorpha)
MapolyLHCBMs (M.polymorpha)
kfl00098_0080 (K.flaccidum)
kfl00098_0090 (K.flaccidum)
kfl01434_0010 (K.flaccidum)
kfl00809_0010 (K.flaccidum)
Cs_38026 (C.subellipsoidea)
kfl00120_0160 (K.flaccidum)
Cs_15915 (C.subellipsoidea)
Cs_21950 (C.subellipsoidea)
Cs_58975 (C.subellipsoidea)
Cs_27246 (C.subellipsoidea)
Cs_28488 (C.subellipsoidea)
Cs_35121 (C.subellipsoidea)
CrLHCBMs (C.reinhardtii)
kfl00422_0020 (K.flaccidum)
LHCBM
(LHCB1- 3)
AtLHCB7 (A.thaliana)
Mapoly0008s0015 (M.polymorpha)
kfl00104_0370 (K.flaccidum)
CrLHCB7 (C.reinhardtii)
LHCB7
ABD37891_MvLHCB5 (M.viride)
Gene11404 (M.viride)
DAA05928_MvLHCB5 (M.viride)
Cs_64185 (C.subellipsoidea)
AtLHCB5 (A.thaliana)
Mapoly0011s0076 (M.polymorpha)
kfl00104_0350 (K.flaccidum)
LHCB5
AtLHCA1 (A.thaliana)
Mapoly0082s0040 (M.polymorpha)
kfl00100_0170 (K.flaccidum)
CrLHCA1 (C.reinhardtii)
Cs_25286 (C.subellipsoidea)
DAA05929_MvLHCA1(M.viride)
Gene11703 (M.viride)
MpLHCA1 (M.polymorpha)
LHCA1
ABD37890_MvLHCB4 (M.viride)
Gene10541 (M.viride)
DAA05922_MvLHCB4 (M.viride)
AtLHCB4.1 (A.thaliana)
AtLHCB4.2 (A.thaliana)
AtLHCB4.3 (A.thaliana)
Mapoly0001s0025 (M.polymorpha)
kfl00258_0020 (K.flaccidum)
CrLHCB4 (C.reinhardtii)
Cs_30169 (C.subellipsoidea)
Cs_67011 (C.subellipsoidea)
LHCB4
AtLHCB6 (A.thaliana)
Mapoly0052s0039 (M.polymorpha)
kfl00276_0010 (K.flaccidum)
LHCB6
DAA05919_MvLHCP1.2 (M.viride)
Gene11156 (M.viride)
DAA05930_MvLHCP1.1 (M.viride)
Gene11308 (M.viride)
MpLHCAP2.1 (M.pusilla)
MpLHCP2.4 (M.pusilla)
MpLHCP2.2 (M.pusilla)
MpLHCP2.5 (M.pusilla)
MpLHCP1.1 (M.pusilla)
LHCP
*ABD37881_MvLHCA4.1 (M.viride)
Gene11715 (M.viride)
*DAA05923_MvLHCA2 (M.viride)
MpLHCA2 (M.pusilla)
Ol_PTHR21649 (O. lucimarinus)
CrLHCA2 (C.reinhardtii)
Cs_7387 (C.subellipsoidea)
Cs_66689
(C.subellipsoidea)
Algae-type
LHCA2
ABD37884_MvLHCA5.2 (M.viride)
Gene10884 (M.viride)
DAA05927_MvLHCA9 (M.viride)
CrLHCA9 (C.reinhardtii)
Cs_44136 (C.subellipsoidea)
MpLHCA9 (M.pusilla)
Ol_K08911 (O. lucimarinus)
LHCA9
*ABD37882_MvLHCA4.2 (M.viride)
Gene11585 (M.viride)
*DAA05926_MvLHCA5 (M.viride)
*DAA05920_MvLHCA4 (M.viride)
Gene11349 (M.viride)
MpLHCA4 (M.pusilla)
Ol_K08908 (O. lucimarinus)
Ol_K08910 (O. lucimarinus)
AtLHCA2 (A.thaliana)
Mapoly0083s0003 (M.polymorpha)
AtLHCA6 (A.thaliana)
kfl00564_0030 (K.flaccidum)
AtLHCA4 (A.thaliana)
Mapoly0006s0261 (M.polymorpha)
kfl00023_0070 (K.flaccidum)
CrLHCA4 (C.reinhardtii)
Cs_52367 (C.subellipsoidea)
CrLHCA5 (C.reinhardtii)
Cs_61250 (C.subellipsoidea)
CrLHCA6 (C.subellipsoidea)
Cs_46127 (C.subellipsoidea)
AtLHCA5 (A.thaliana)
kfl00214_0160 (K.flaccidum)
CrLHCA7 (C.reinhardtii)
Cs_25284 (C.subellipsoidea)
CrLHCA8 (C.reinhardtii)
Cs_48543 (C.subellipsoidea)
LHCA2
AtLHCA3 (A.thaliana)
Mapoly0066s0050 (M.polymorpha)
kfl00560_0090 (K.flaccidum)
CrLHCA3 (C.reinhardtii)
Cs_37969 (C.subellipsoidea)
MpLHCA3 (M.pusilla)
Ol_K08909 (O. lucimarinus)
ABD37880_MvLHCA3 (M.viride)
DAA05918_MvLHCA3 (M.viride)
Gene11536 (M.viride)
LHCA3
CmLHC
PyLHC
0.6

Fig. 3

A


PSI-PSII (1)
PSII-LHCII (8,9)
PSI-LHCI-LHCII (12,13)
PSI-LHCI (14,15)
Estimated amounts (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Gel slice number
PSI
PSII

B


PSI-PSII (1)
PSII-LHCII (8,9)
PSI-LHCI-LHCII (12,13)
PSI-LHCI (14,15)
Estimated amounts (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Gel slice number
LHCA1
LHCA2
LHCA3
algal-type LHCA2
LHCA9

C

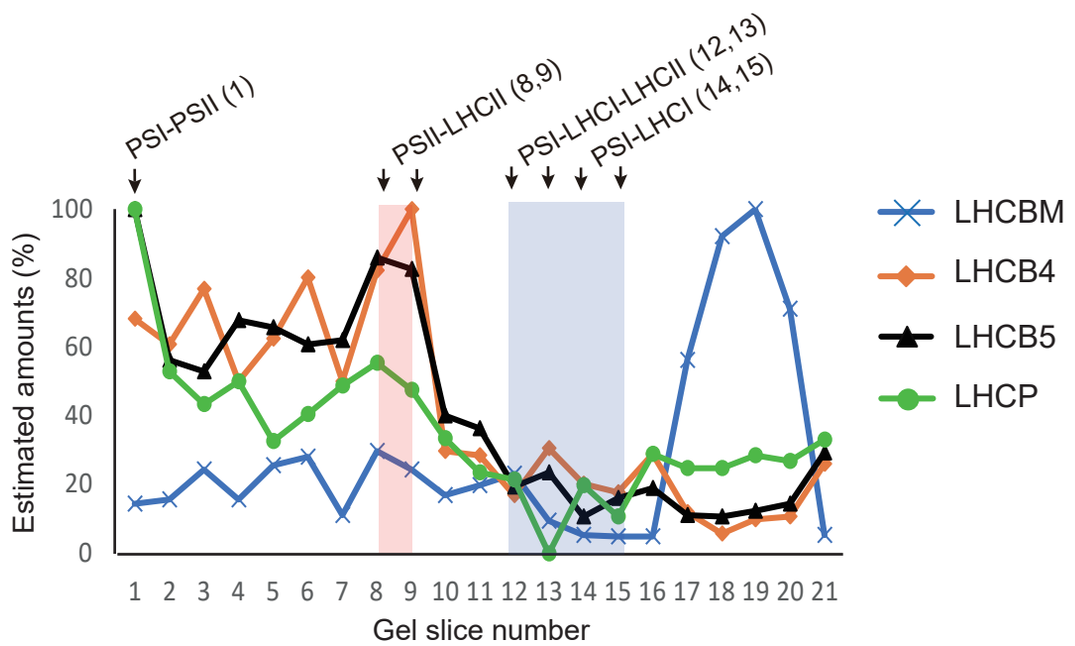

PSI-PSII (1)
PSII-LHCII (8,9)
PSI-LHCI-LHCII (12,13)
PSI-LHCI (14,15)
Estimated amounts (%)
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Gel slice number
LHCBM
LHCB4
LHCB5
LHCP

Fig. 4

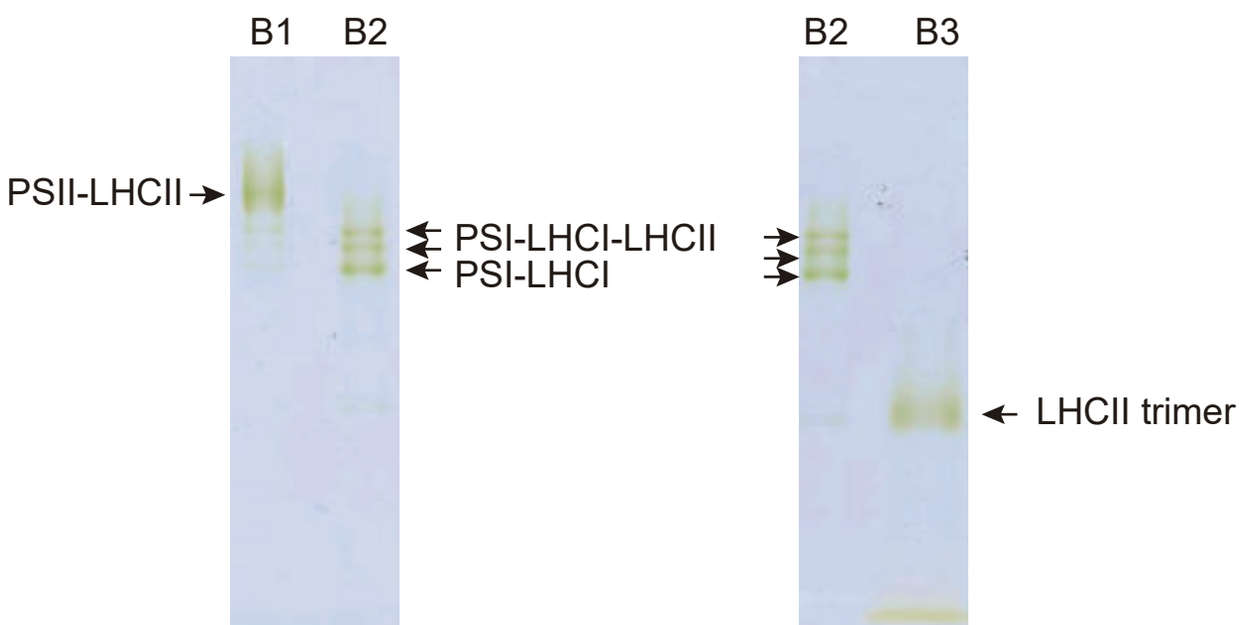
A

B3 (LHCII trimer and LHC monomer)

B2 (PSI-LHCI(-LHCII))

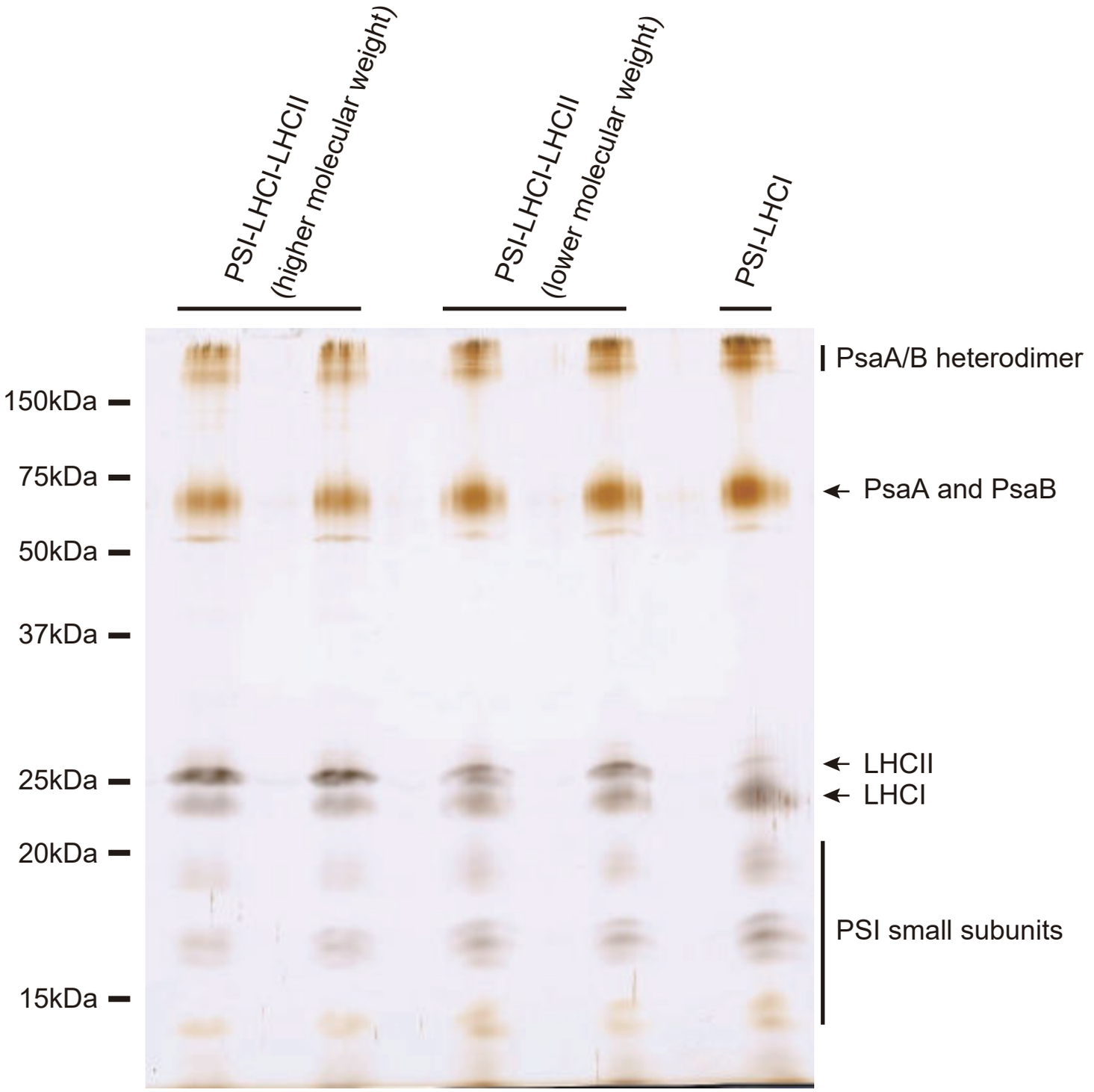
B1 (PSII-LHCII)

B



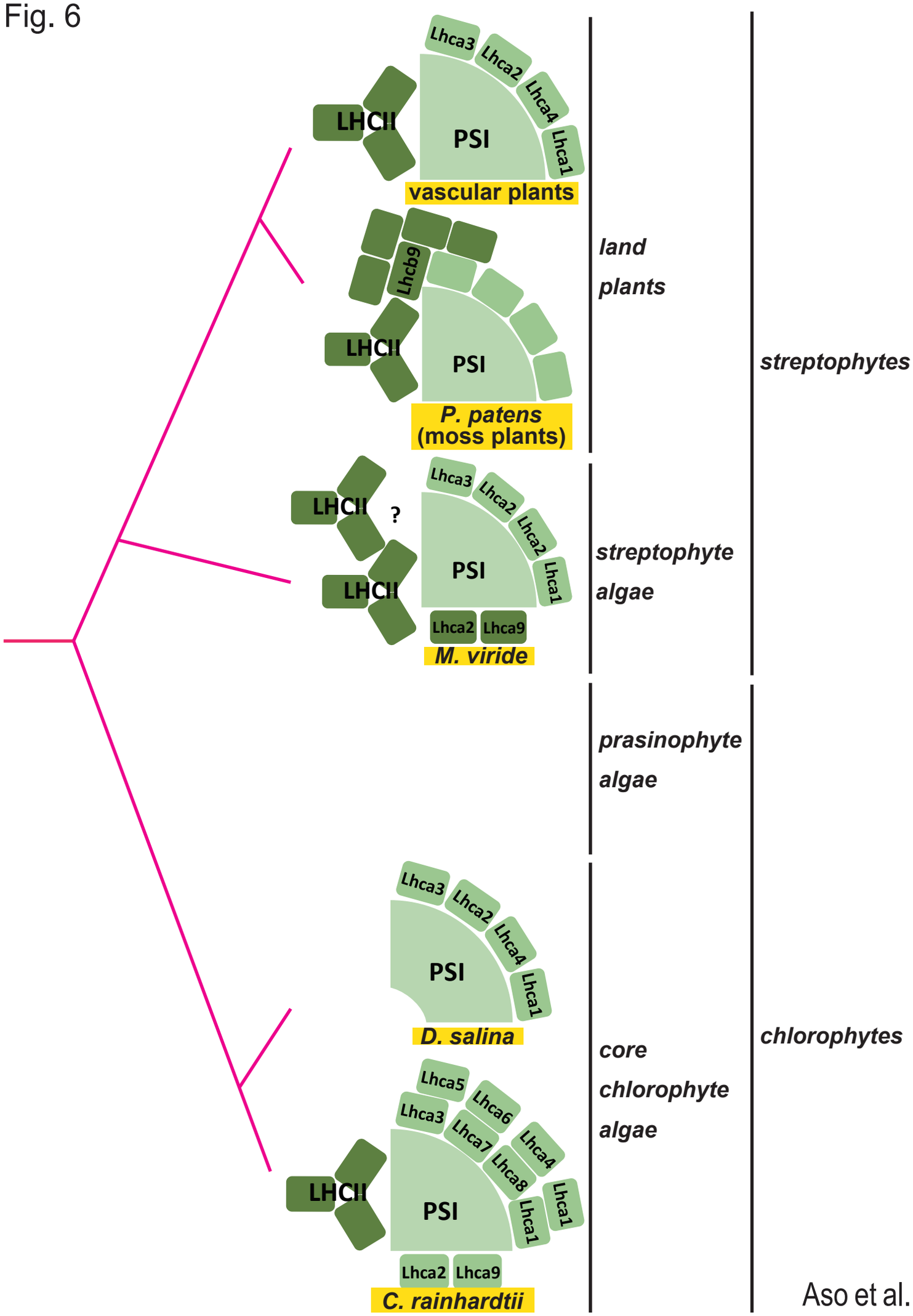
Aso et al.

Fig. 5



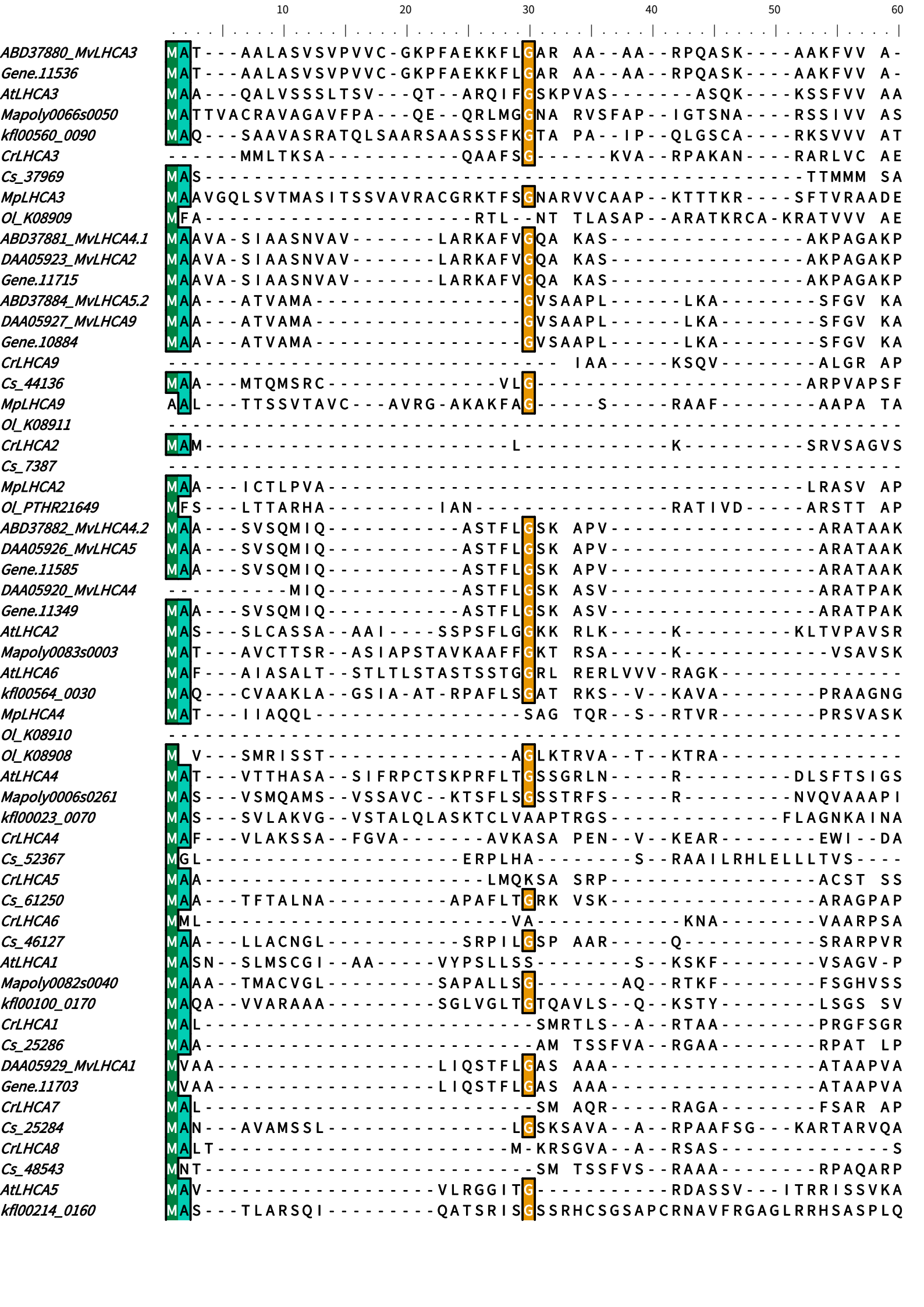
Aso et al.

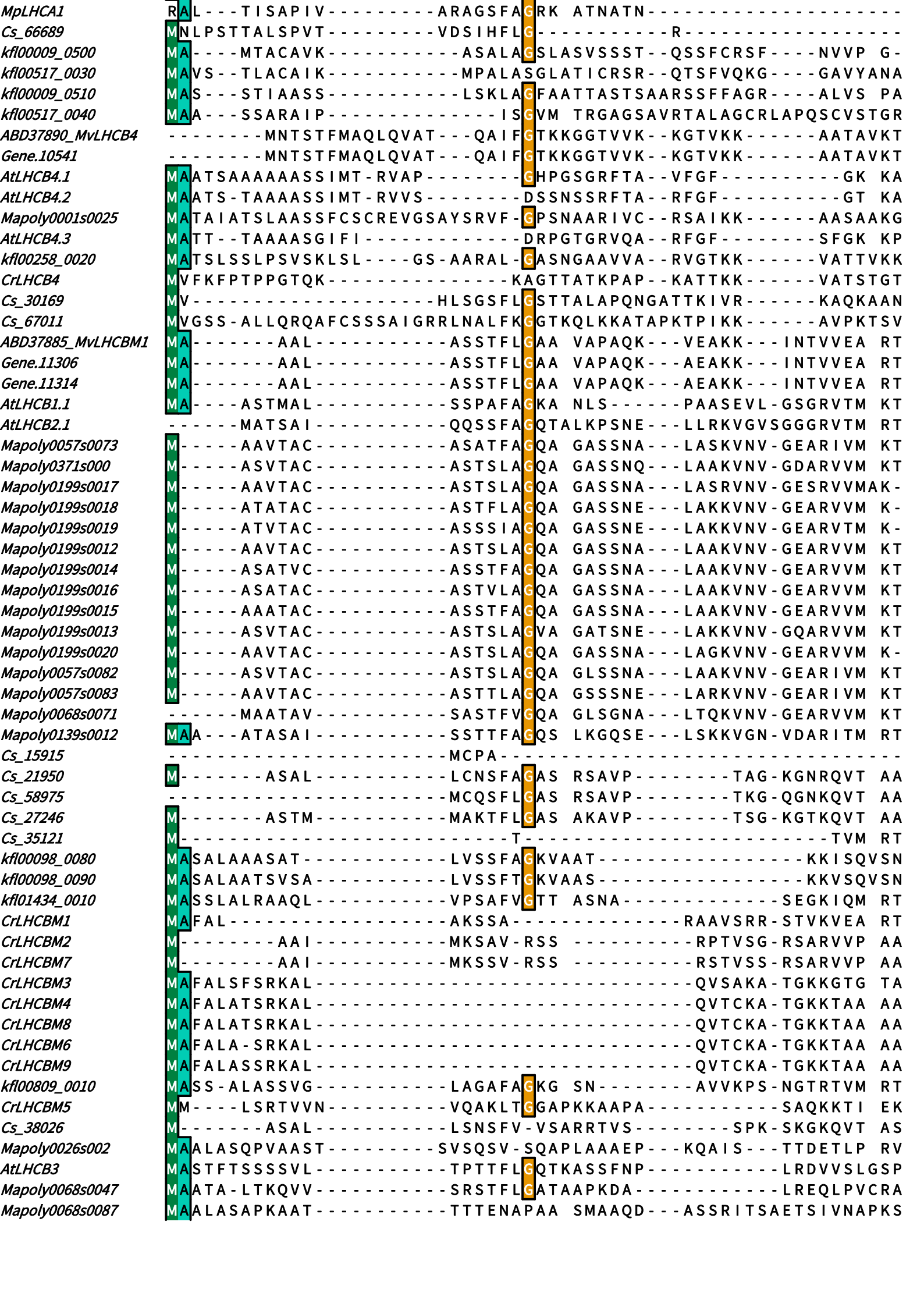
Fig. 6

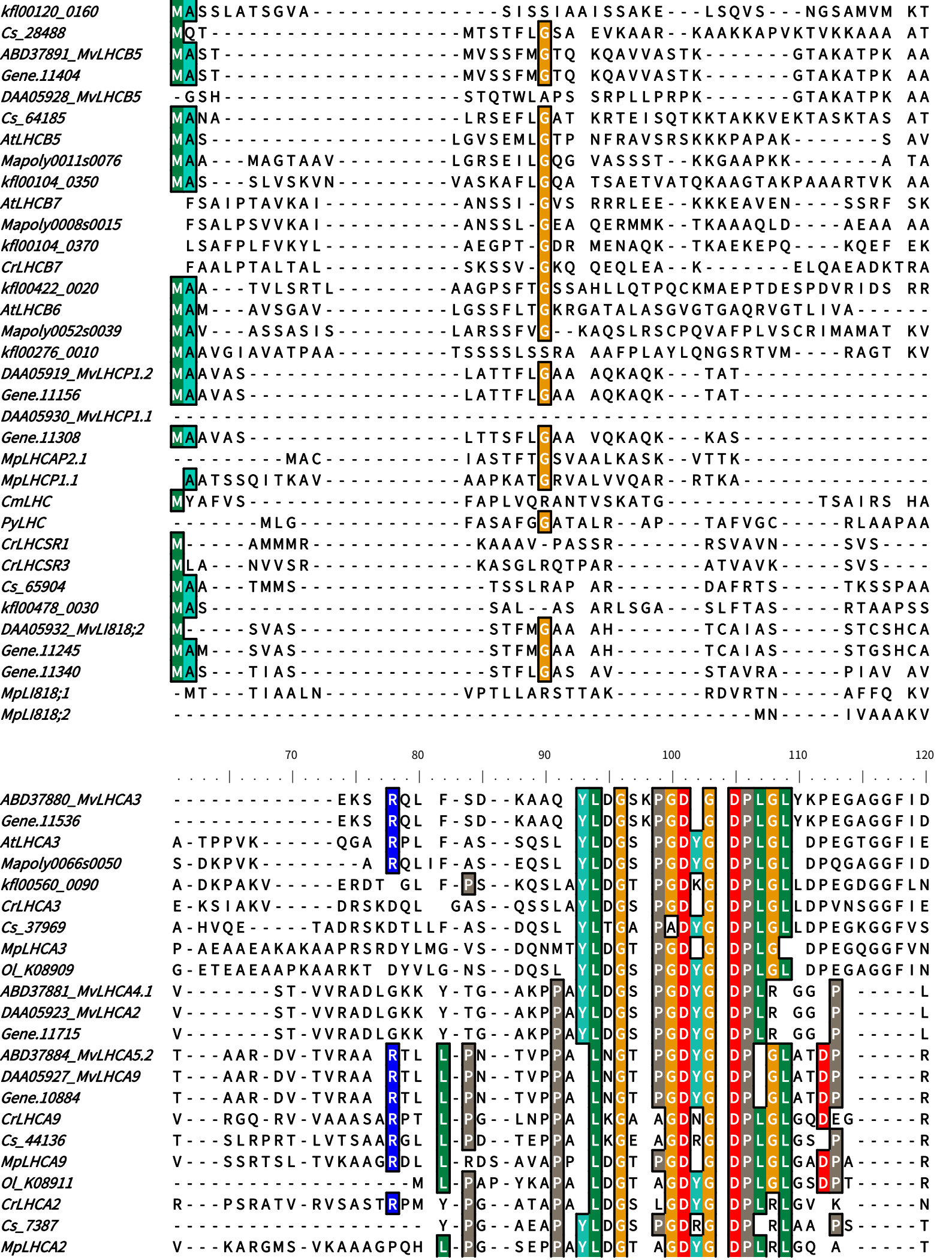


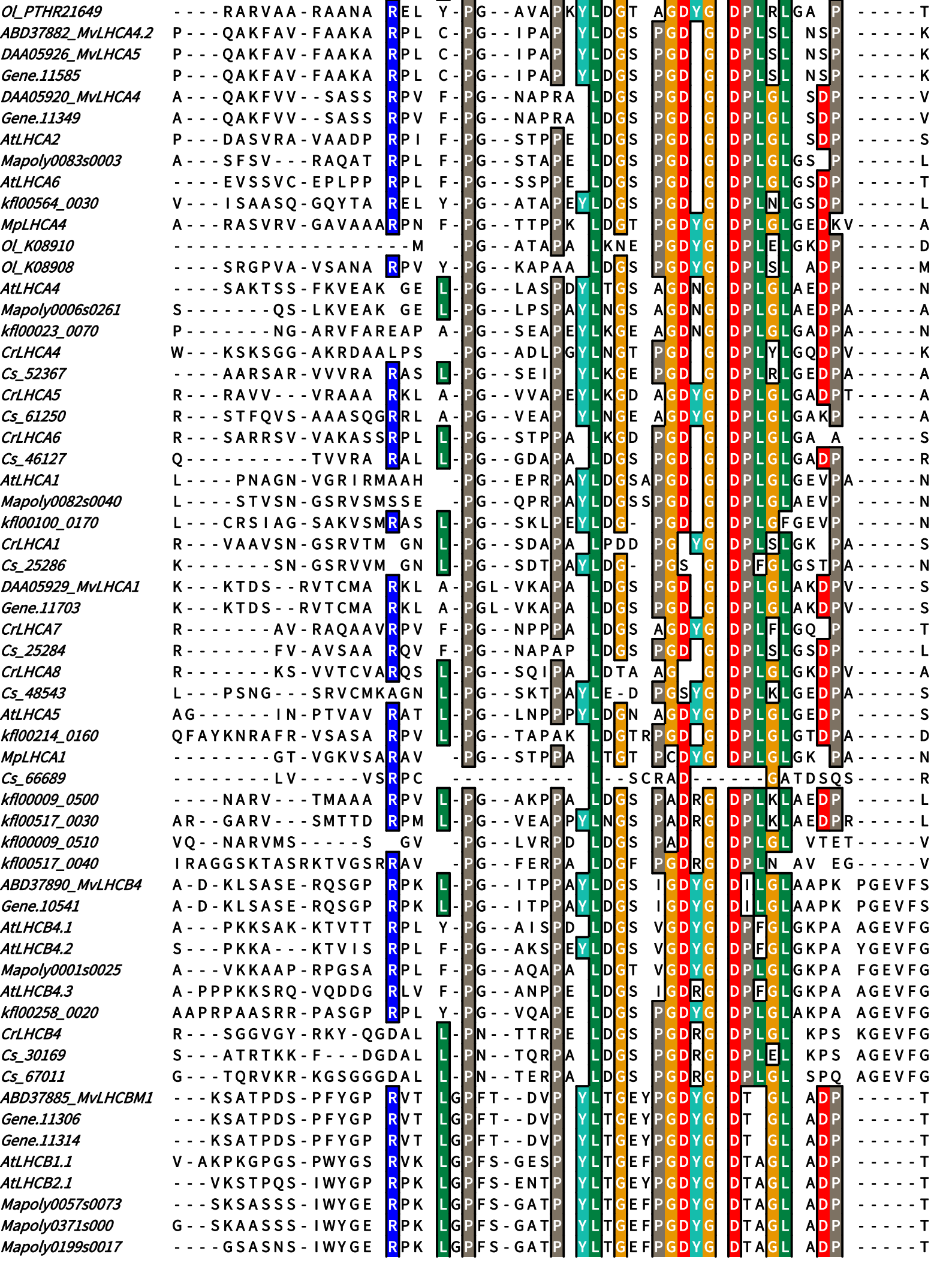
Aso et al.

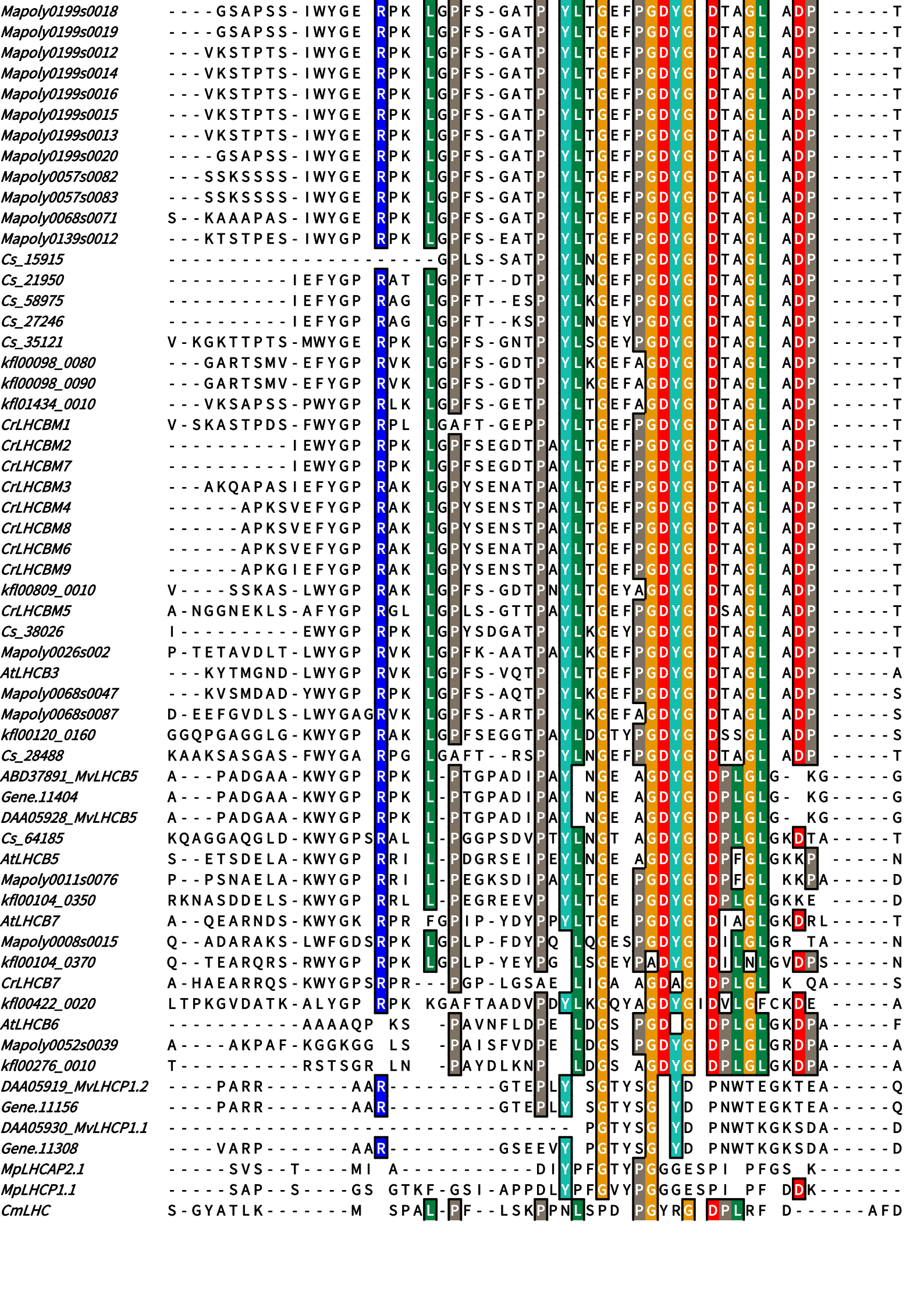
Fig. S1 A multiple sequence alignment of LHC sequences. Below is the multiple sequence alignment used to construct a phylogenetic tree in Fig. 2. A MAFFT program (https://mafft.cbrc.jp/alignment/software/) was used to align LHC sequences and a ClipKIT program was used to do trimming of the alignment.

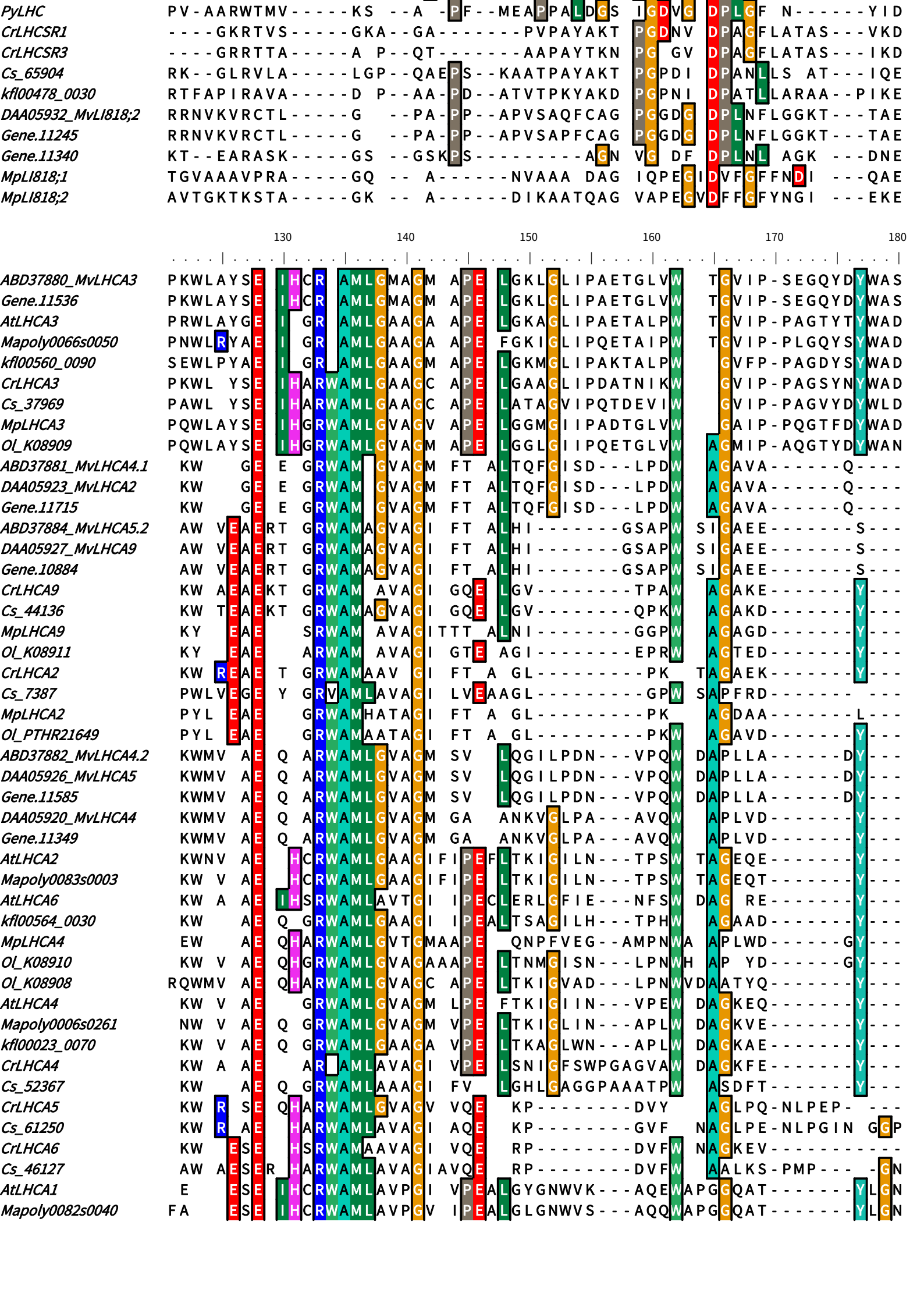

10
20
30
40
50
60
ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2
Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2
Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908
AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040
kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160

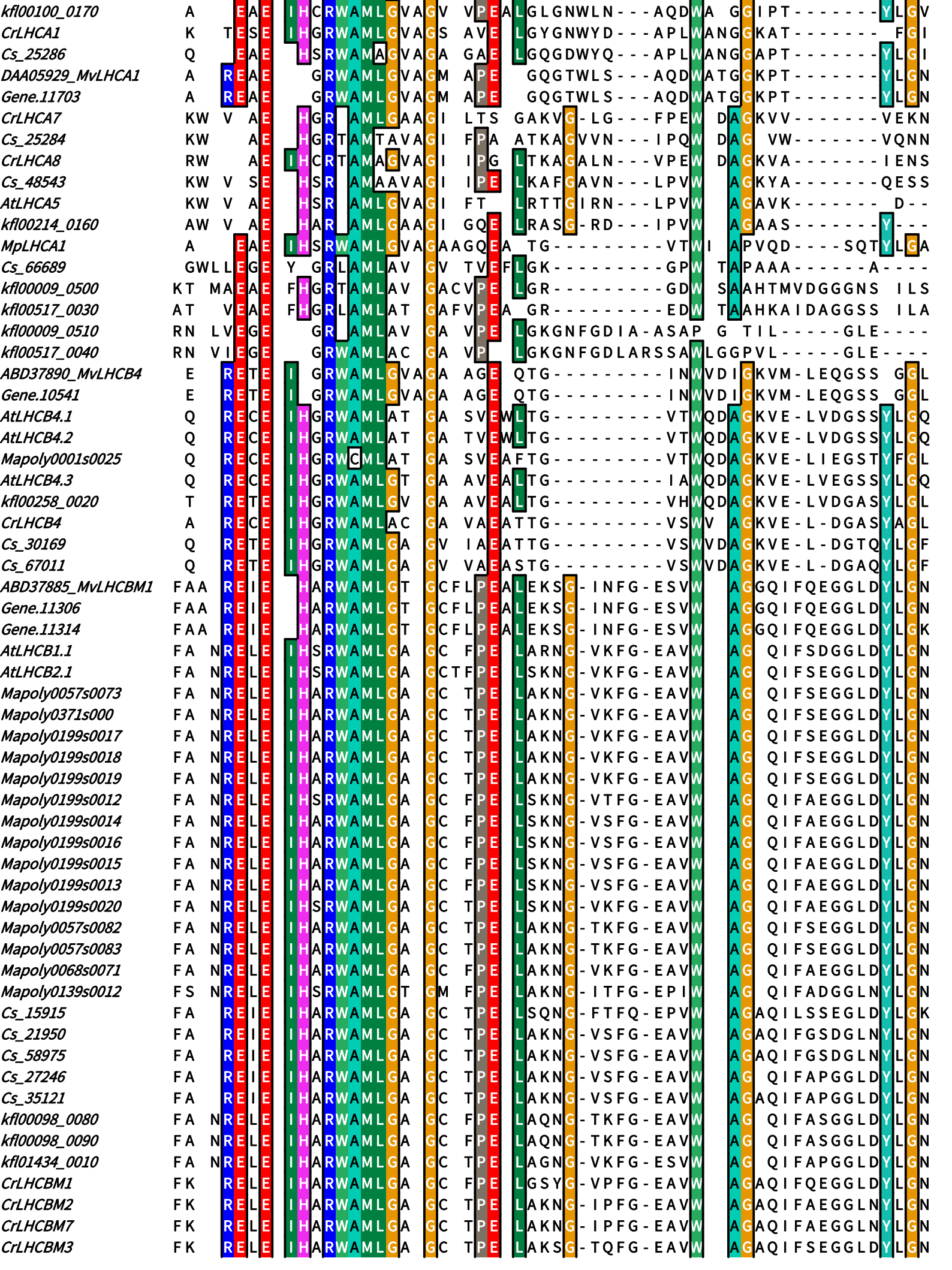

MpLHCA1
Cs_66689
kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020
CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017
Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915
Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3
CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087

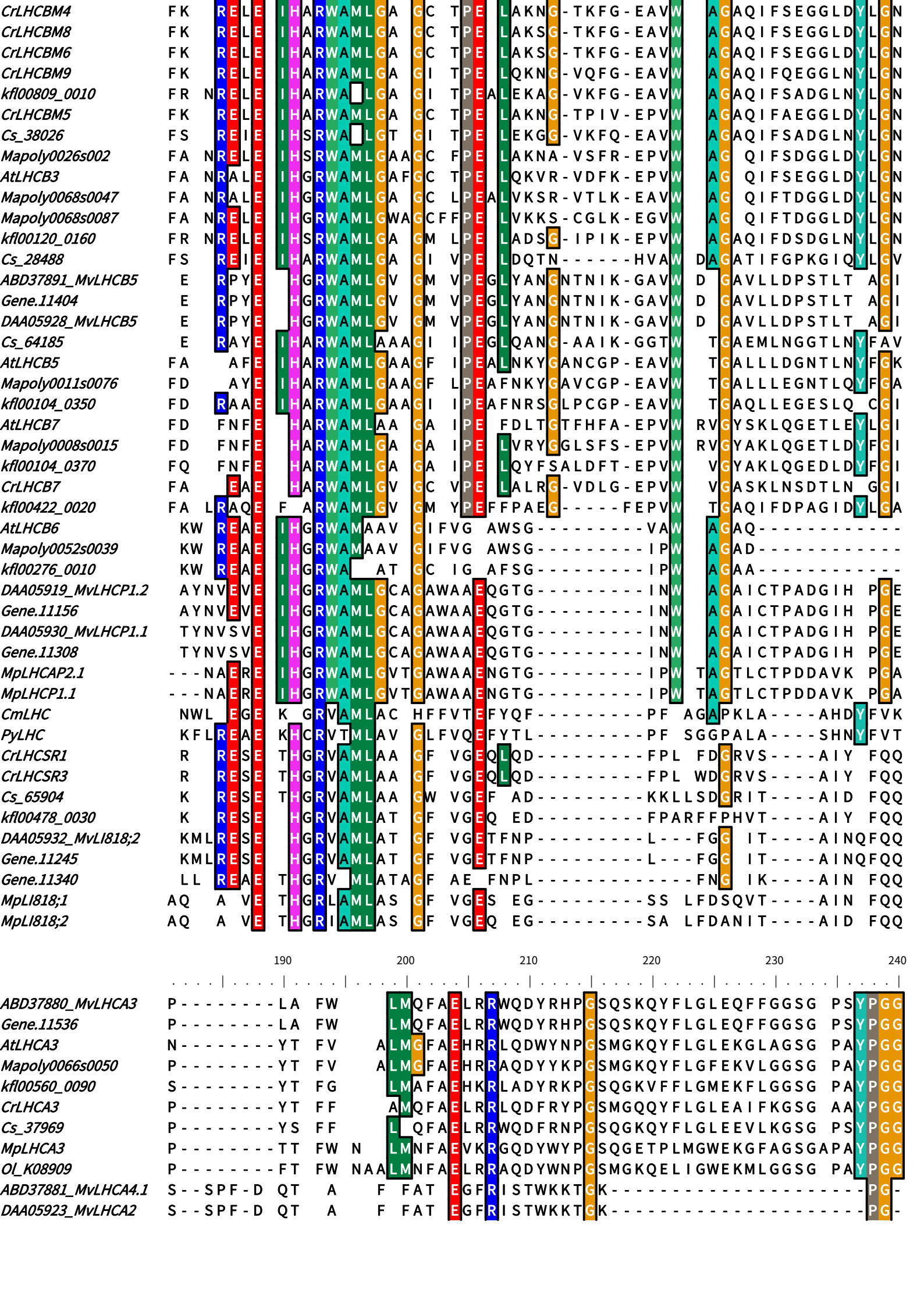

kfl00120_0160
Cs_28488
ABD37891_MvLHCB5
Gene.11404
DAA05928_MvLHCB5
Cs_64185
AtLHCB5
Mapoly0011s0076
kfl00104_0350
AtLHCB7
Mapoly0008s0015
kfl00104_0370
CrLHCB7
kfl00422_0020
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2
Gene.11156
DAA05930_MvLHCP1.1
Gene.11308
MpLHCAP2.1
MpLHCP1.1
CmLHC
PyLHC
CrLHCSR1
CrLHCSR3
Cs_65904
kfl00478_0030
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1
MpLI818;2
70
80
90
100
110
120
ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2
Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2

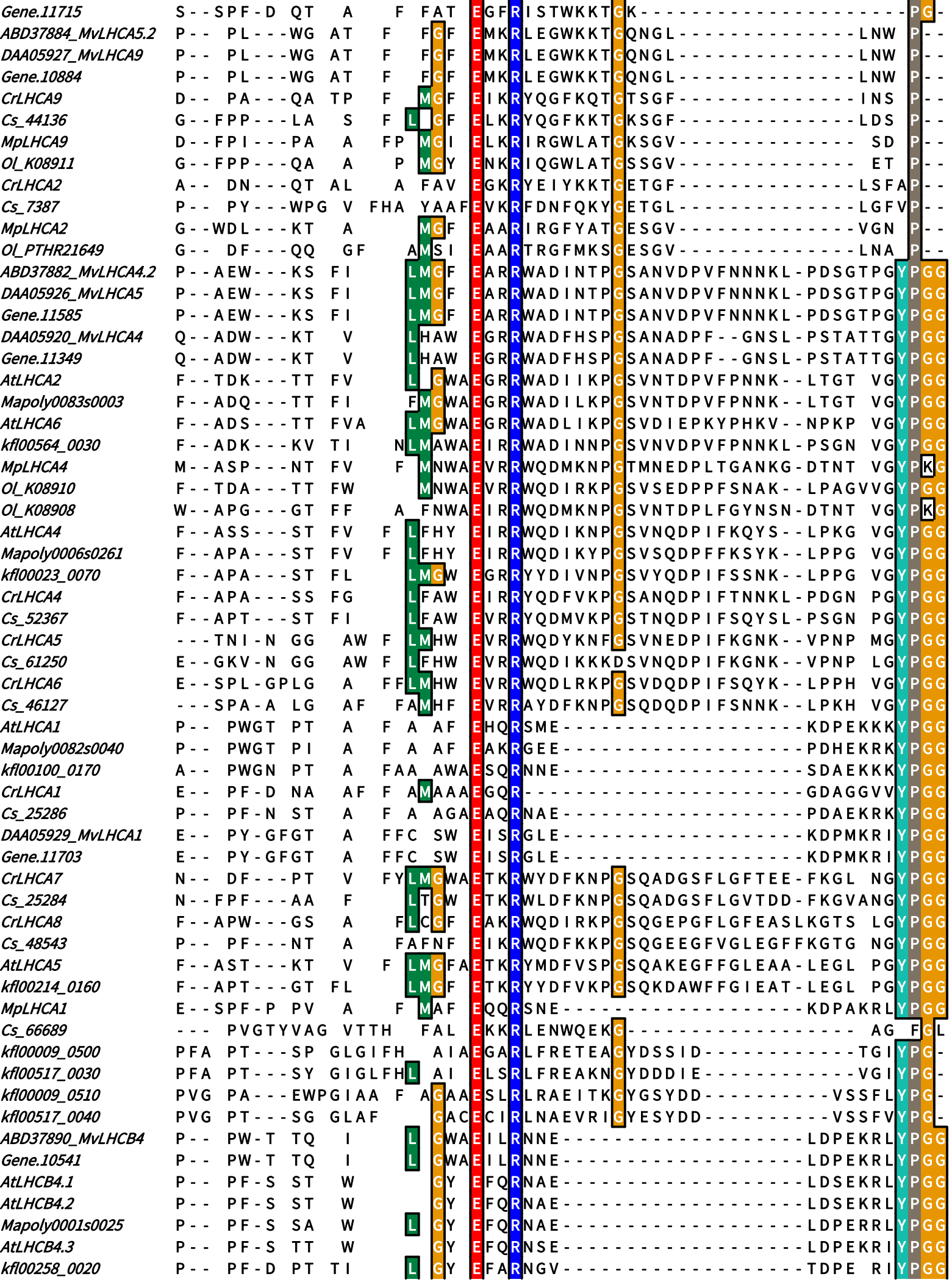

Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908
AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040
kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160
MpLHCA1
Cs_66689
kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020
CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017

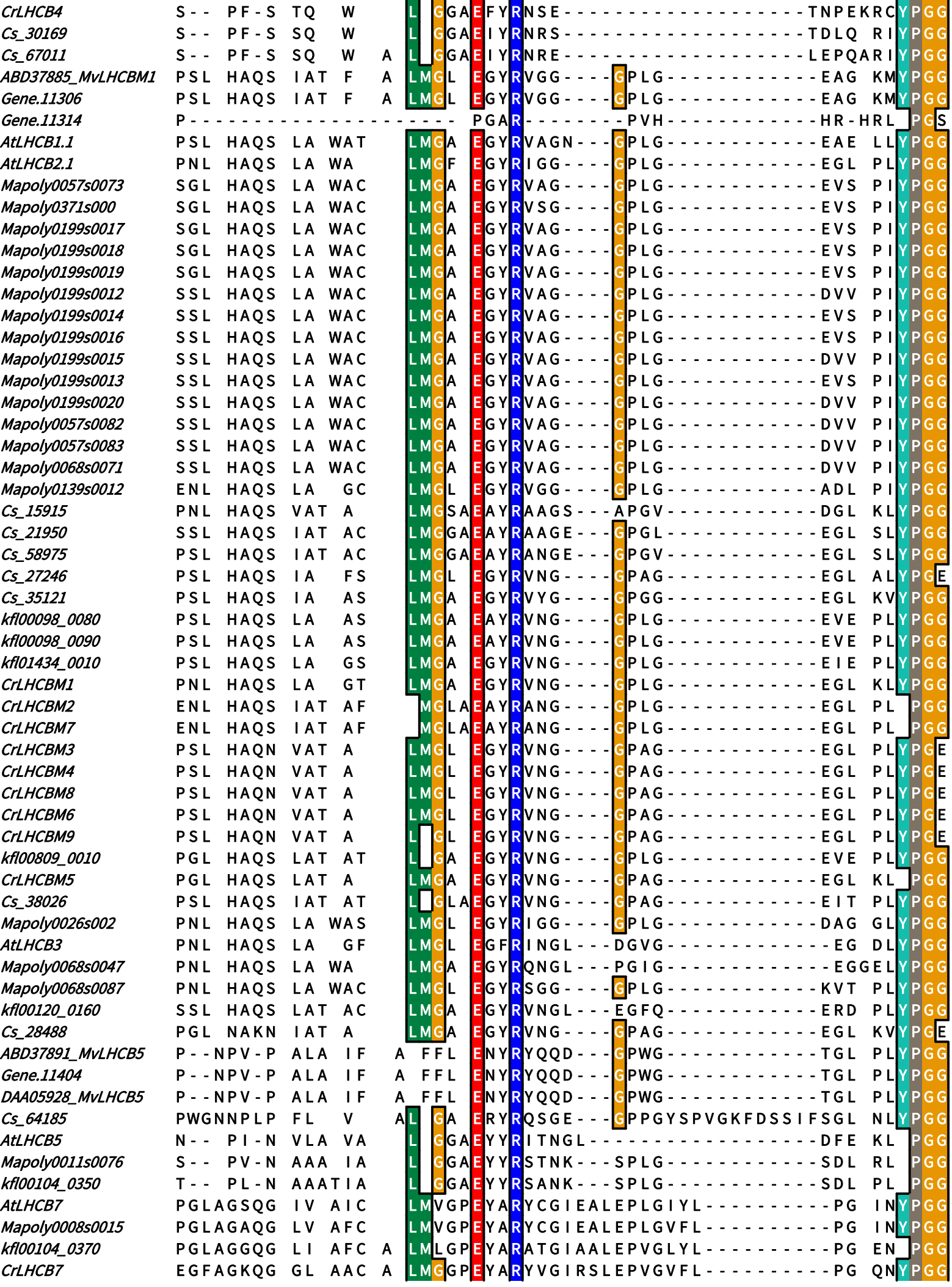

Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915
Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3
CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087
kfl00120_0160
Cs_28488
ABD37891_MvLHCB5
Gene.11404
DAA05928_MvLHCB5
Cs_64185
AtLHCB5
Mapoly0011s0076
kfl00104_0350
AtLHCB7
Mapoly0008s0015
kfl00104_0370
CrLHCB7
kfl00422_0020
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2
Gene.11156
DAA05930_MvLHCP1.1
Gene.11308
MpLHCAP2.1
MpLHCP1.1
CmLHC

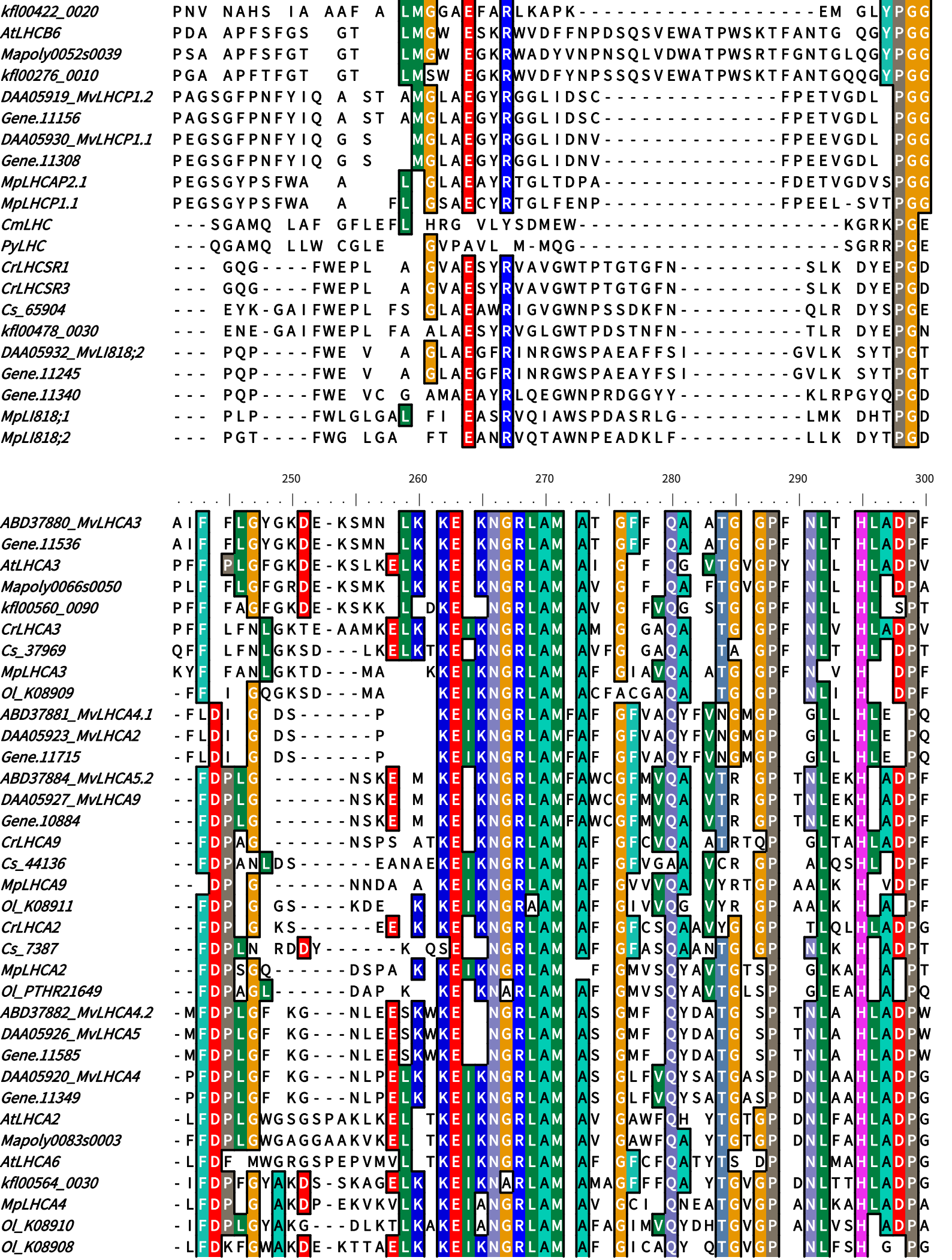

PyLHC
CrLHCSR1
CrLHCSR3
Cs_65904
kfl00478_0030
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1
MpLI818;2
130
140
150
160
170
180
ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2
Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2
Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908
AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040

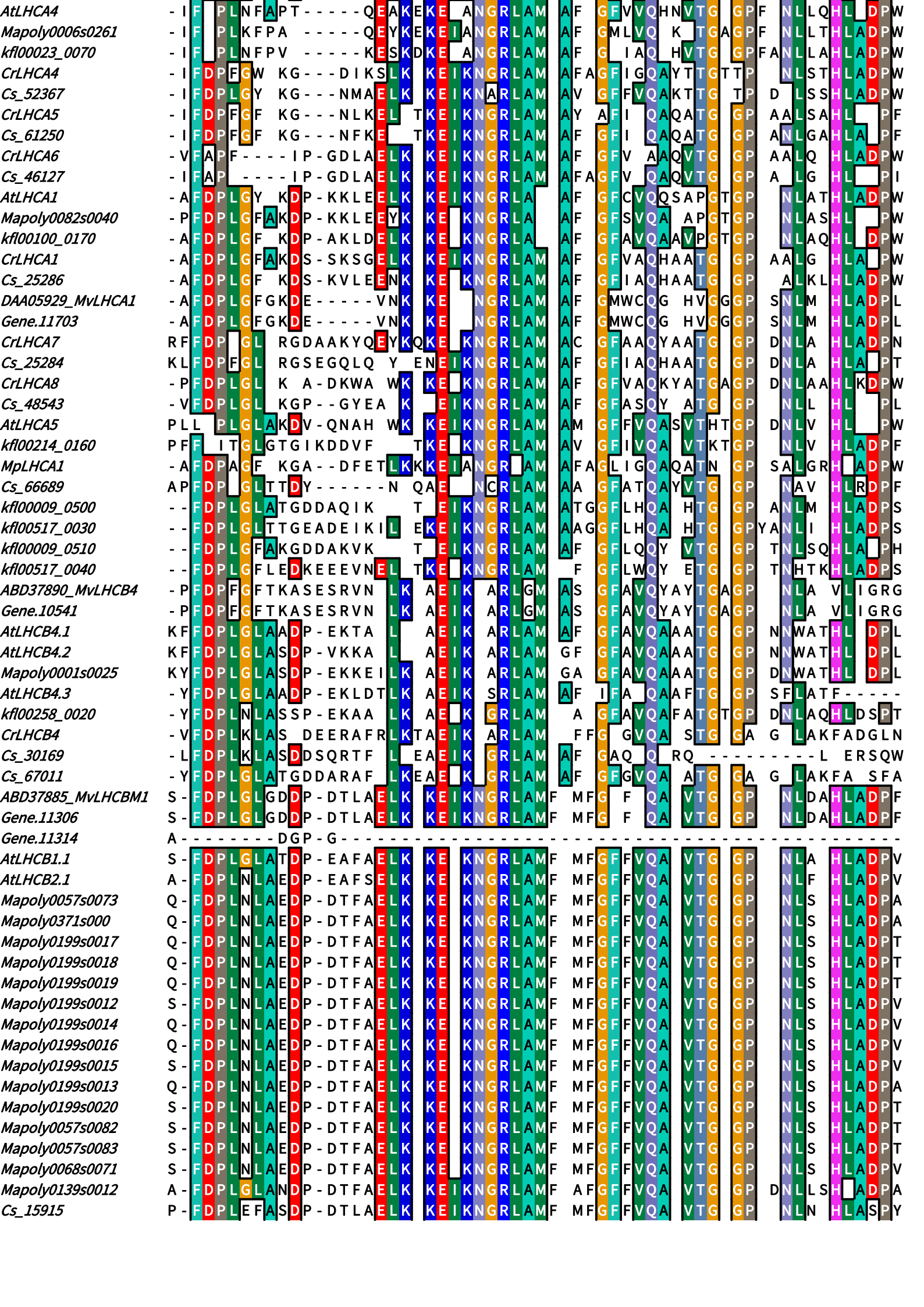

kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160
MpLHCA1
Cs_66689
kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020
CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017
Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915
Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3

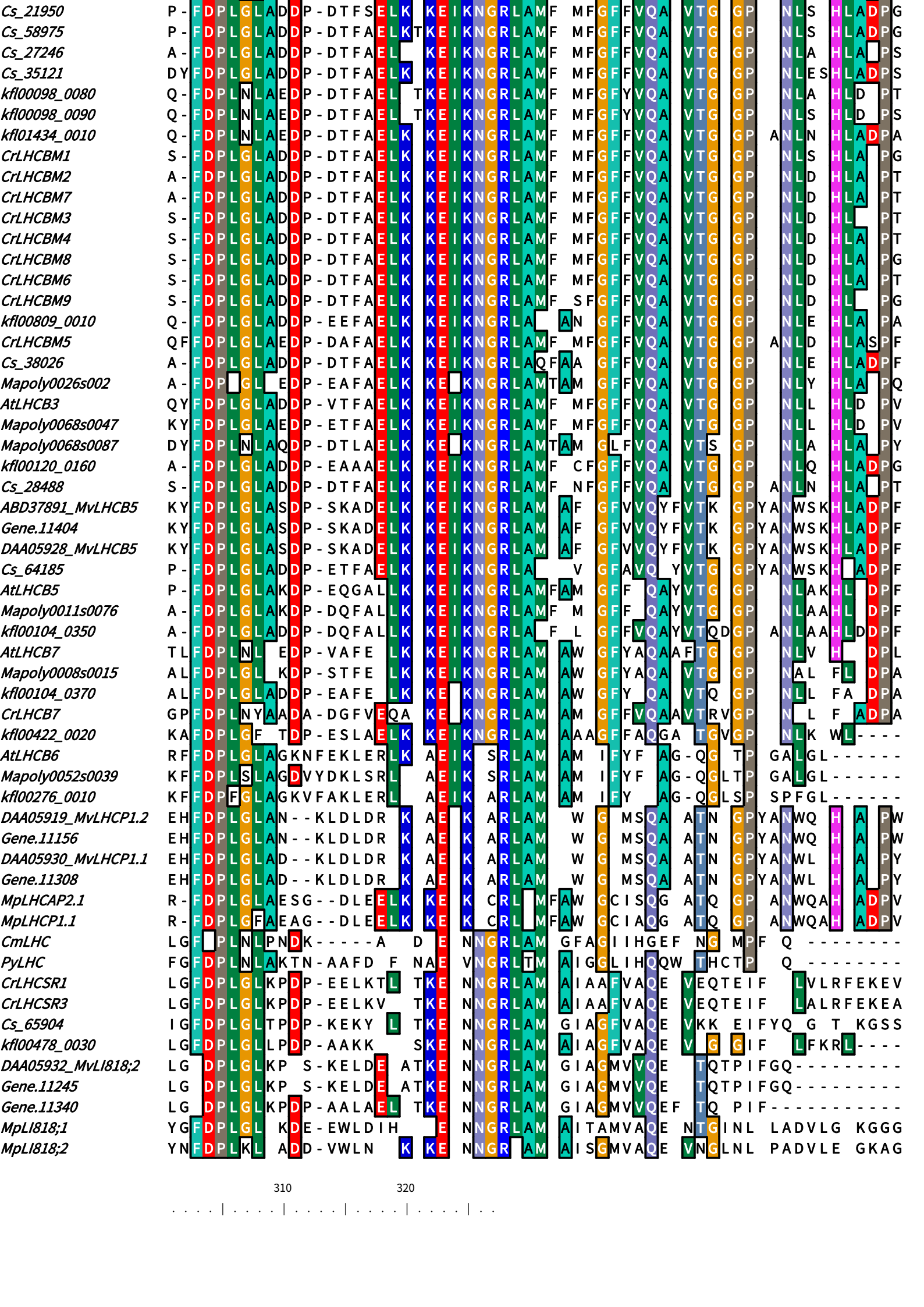

CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087
kfl00120_0160
Cs_28488
ABD37891_MvLHCB5
Gene.11404
DAA05928_MvLHCB5
Cs_64185
AtLHCB5
Mapoly0011s0076
kfl00104_0350
AtLHCB7
Mapoly0008s0015
kfl00104_0370
CrLHCB7
kfl00422_0020
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2
Gene.11156
DAA05930_MvLHCP1.1
Gene.11308
MpLHCAP2.1
MpLHCP1.1
CmLHC
PyLHC
CrLHCSR1
CrLHCSR3
Cs_65904
kfl00478_0030
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1
MpLI818;2
190
200
210
220
230
240
ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2

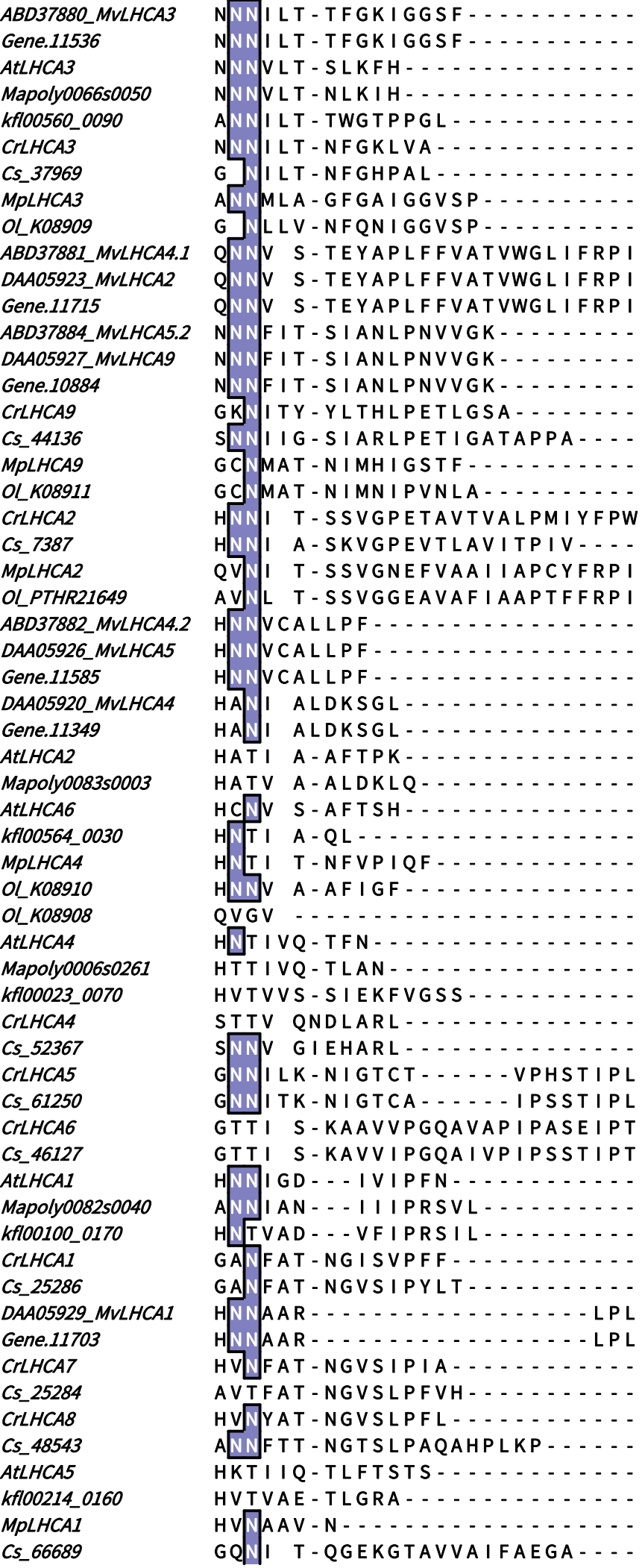

Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2
Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908
AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040
kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160
MpLHCA1
Cs_66689
kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020

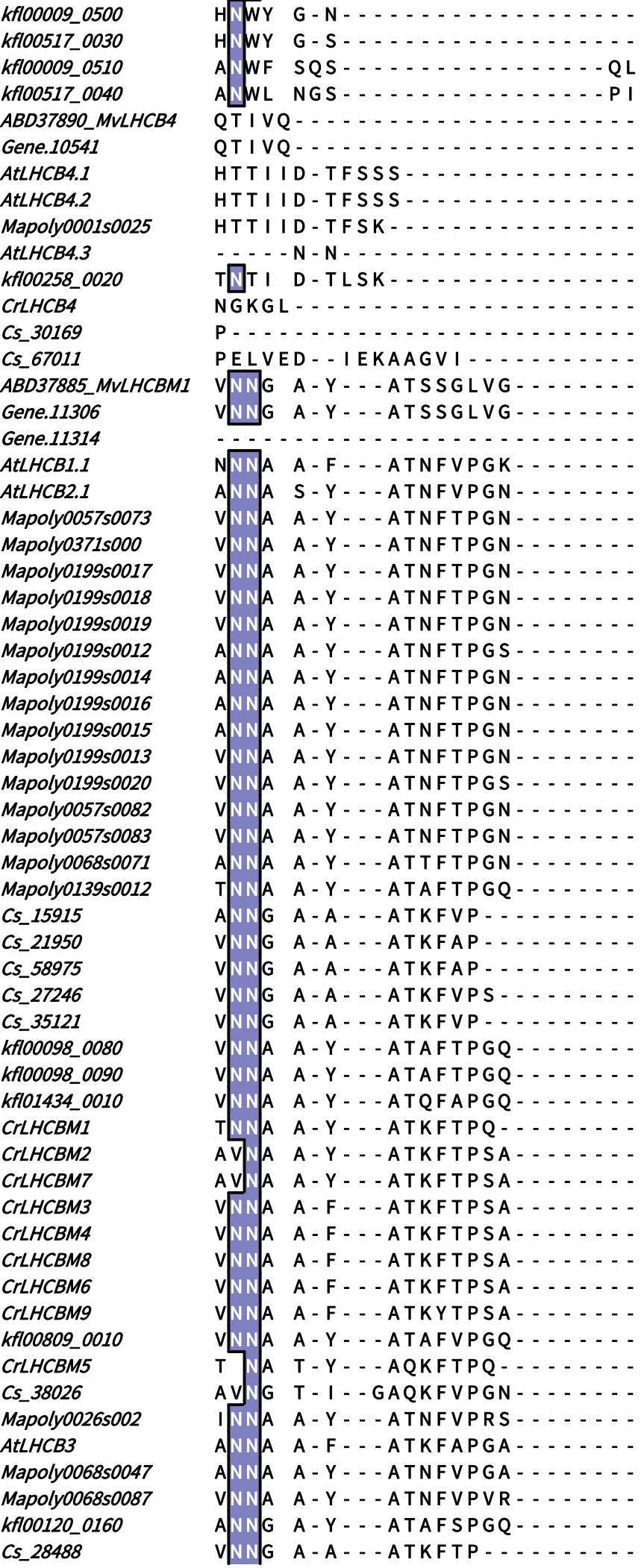

CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017
Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915
Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3
CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087
kfl00120_0160
Cs_28488
ABD37891_MvLHCB5
Gene.11404
DAA05928_MvLHCB5
Cs_64185
AtLHCB5
Mapoly0011s0076
kfl00104_0350
AtLHCB7
Mapoly0008s0015
kfl00104_0370
CrLHCB7


kfl00422_0020
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2
Gene.11156
DAA05930_MvLHCP1.1
Gene.11308
MpLHCAP2.1
MpLHCP1.1
CmLHC
PyLHC
CrLHCSR1
CrLHCSR3
Cs_65904
kfl00478_0030
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1
MpLI818;2
250
260
270
280
290
300
ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2
Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2
Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908


AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040
kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160
MpLHCA1
Cs_66689
kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020
CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017
Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915


Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3
CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087
kfl00120_0160
Cs_28488
ABD37891_MvLHCB5
Gene.11404
DAA05928_MvLHCB5
Cs_64185
AtLHCB5
Mapoly0011s0076
kfl00104_0350
AtLHCB7
Mapoly0008s0015
kfl00104_0370
CrLHCB7
kfl00422_0020
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2
Gene.11156
DAA05930_MvLHCP1.1
Gene.11308
MpLHCAP2.1
MpLHCP1.1
CmLHC
PyLHC
CrLHCSR1
CrLHCSR3
Cs_65904
kfl00478_0030
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1
MpLI818;2
310
320


ABD37880_MvLHCA3
Gene.11536
AtLHCA3
Mapoly0066s0050
kfl00560_0090
CrLHCA3
Cs_37969
MpLHCA3
Ol_K08909
ABD37881_MvLHCA4.1
DAA05923_MvLHCA2
Gene.11715
ABD37884_MvLHCA5.2
DAA05927_MvLHCA9
Gene.10884
CrLHCA9
Cs_44136
MpLHCA9
Ol_K08911
CrLHCA2
Cs_7387
MpLHCA2
Ol_PTHR21649
ABD37882_MvLHCA4.2
DAA05926_MvLHCA5
Gene.11585
DAA05920_MvLHCA4
Gene.11349
AtLHCA2
Mapoly0083s0003
AtLHCA6
kfl00564_0030
MpLHCA4
Ol_K08910
Ol_K08908
AtLHCA4
Mapoly0006s0261
kfl00023_0070
CrLHCA4
Cs_52367
CrLHCA5
Cs_61250
CrLHCA6
Cs_46127
AtLHCA1
Mapoly0082s0040
kfl00100_0170
CrLHCA1
Cs_25286
DAA05929_MvLHCA1
Gene.11703
CrLHCA7
Cs_25284
CrLHCA8
Cs_48543
AtLHCA5
kfl00214_0160
MpLHCA1
Cs_66689


kfl00009_0500
kfl00517_0030
kfl00009_0510
kfl00517_0040
ABD37890_MvLHCB4
Gene.10541
AtLHCB4.1
AtLHCB4.2
Mapoly0001s0025
AtLHCB4.3
kfl00258_0020
CrLHCB4
Cs_30169
Cs_67011
ABD37885_MvLHCBM1
Gene.11306
Gene.11314
AtLHCB1.1
AtLHCB2.1
Mapoly0057s0073
Mapoly0371s000
Mapoly0199s0017
Mapoly0199s0018
Mapoly0199s0019
Mapoly0199s0012
Mapoly0199s0014
Mapoly0199s0016
Mapoly0199s0015
Mapoly0199s0013
Mapoly0199s0020
Mapoly0057s0082
Mapoly0057s0083
Mapoly0068s0071
Mapoly0139s0012
Cs_15915
Cs_21950
Cs_58975
Cs_27246
Cs_35121
kfl00098_0080
kfl00098_0090
kfl01434_0010
CrLHCBM1
CrLHCBM2
CrLHCBM7
CrLHCBM3
CrLHCBM4
CrLHCBM8
CrLHCBM6
CrLHCBM9
kfl00809_0010
CrLHCBM5
Cs_38026
Mapoly0026s002
AtLHCB3
Mapoly0068s0047
Mapoly0068s0087
kfl00120_0160
Cs_28488

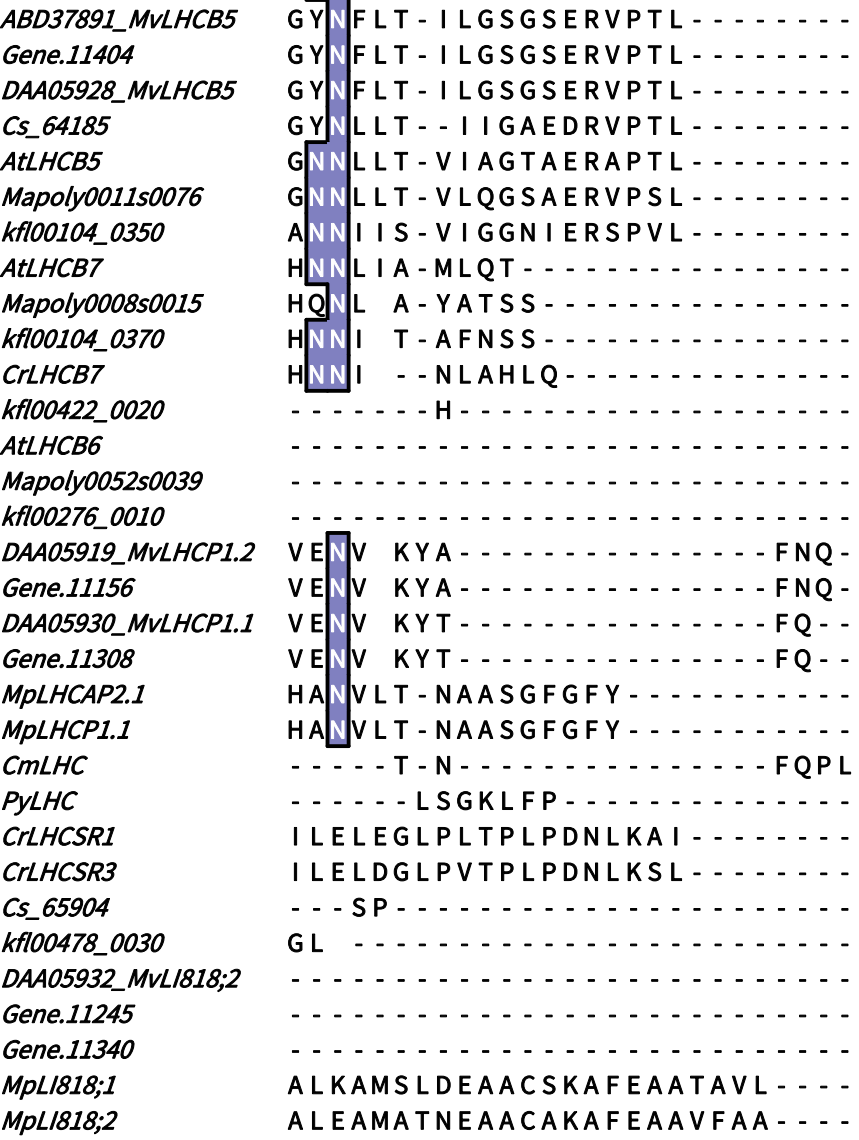

ABD37891_MvLHCB5 GYNFLT-ILGSGSERVPTL
Gene.11404 GYNFLT-ILGSGSERVPTL
DAA05928_MvLHCB5 GYNFLT-ILGSGSERVPTL
Cs_64185 GYNLLT--IIGAEDRVPTL
AtLHCB5 GNNLLT-VIAGTAERAPTL
Mapoly0011s0076 GNNLLT-VLQGSAERVPSL
kfl00104_0350 ANNIIS-VIGGNIERSPVL
AtLHCB7 HNNLIA-MLQT
Mapoly0008s0015 HQNL A-YATSS
kfl00104_0370 HNNI T-AFNSS
CrLHCB7 HNNI --NLAHLQ
kfl00422_0020 -------H
AtLHCB6
Mapoly0052s0039
kfl00276_0010
DAA05919_MvLHCP1.2 VENV KYA----------FNQ-
Gene.11156 VENV KYA----------FNQ-
DAA05930_MvLHCP1.1 VENV KYT----------FQ--
Gene.11308 VENV KYT----------FQ--
MpLHCAP2.1 HANVLT-NAASGFGFY----------
MpLHCP1.1 HANVLT-NAASGFGFY----------
CmLHC -----T-N----------FQPL
PyLHC ------LSGKLFP----------
CrLHCSR1 ILELEGLPLTPLPDNLKAI--------
CrLHCSR3 ILELDGLPVTPLPDNLKSL--------
Cs_65904 ---SP----------
kfl00478_0030 GL ----------
DAA05932_MvLI818;2
Gene.11245
Gene.11340
MpLI818;1 ALKAMSLDEAACSKAFEAATAVL----
MpLI818;2 ALEAMATNEAACAKAFEAAVFAA----

Fig. S2

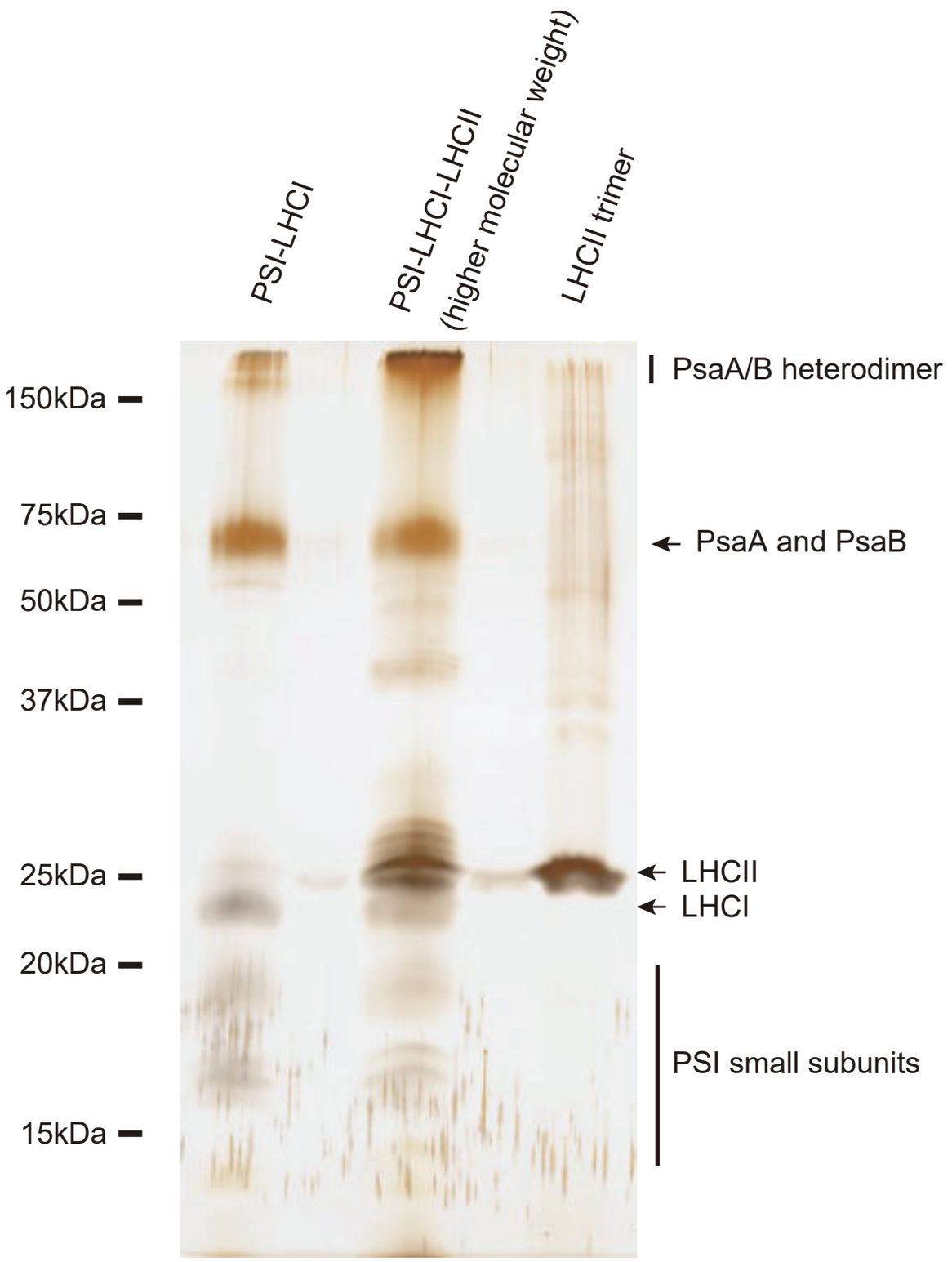


Figure S2. 2D-SDS-PAGE of the high-molecular-weight PSI-LHCI-LHCII, PSI-LHCI, and LHCII trimer.

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	Accesson No.	The best-hit gene in *A.thliana*			The best-hit gene in *C. reinhardtii*		
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 l	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	PSAO1	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	Accesson No.	The best-hit gene in *A.thliana*			The best-hit gene in *C. reinhardtii*		
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	Accesson No.	The best-hit gene in *A.thliana*			The best-hit gene in *C. reinhardtii*		
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

					Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
Gene ID	NSAF (fmol)	Category	Gene annotation in this study	Accesson No.	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	HFO18
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.g234550.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

					Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
Gene ID	NSAF(fmol)	Category	Gene annotation in this study	Accesson No.	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	HFO18
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

					Best-Hit Gene in Arabidopsis		Best-Hit Gene in Chlamydomonas	
Gene ID	NSAF(fmol)	Category	Gene annotation in this study	Accesson No.	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		