Evolution of Ycf54-independent chlorophyll biosynthesis in cyanobacteria

Supporting Information

Supplementary Methods

Plasmid construction. The pPD[*gene*] plasmids were made by cloning the indicated *gene* fragment into the *Nde*I/*Bgl*II sites of pPD-*N*FLAG (1). The pPD-*N*FLAG[slr1916] and pPD-*C*FLAG[slr1916] plasmids were made by cloning slr1916 into the *Not*I/*Bgl*II sites of pPD-*N*FLAG (1) and the *Nde*I/*Nhe*I sites of pPD-*C*FLAG (2), respectively. The pBB[*gene*] plasmids were constructed by cloning the indicated gene fragment into the *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$ (3). The *cycI* and *cycII* genes and the ORF slr1916 were amplified from *Synechocystis* genomic DNA. The *cycI*^{SM} gene was amplified from *Synechocystis* SM1 genomic DNA. The *ycf54*^{MED4} gene was amplified from a synthesized *acsF*^{MED4}*-ycf54*^{MED4} gene fragment (with codon optimization for *Rvi. gelatinosus*; GenScript) (see *SI Appendix*, Table S3 for sequence). The *cycI-ycf54* fragment was amplified from the pK18[*cycI-ycf54*] plasmid (4). The pBB[*cycI*^{SM}*-ycf54*] was made by introducing the D219G substitution into the pBB[*cycI-ycf54*] plasmid using the QuikChange II site-directed mutagenesis kit (Agilent). The CDS of *Arabidopsis* chlorophyllase-1 (CLH1, AT1G19670) was synthesized (IDT) with nucleotide 729 changed from a T to a C to remove an internal *Nde*I site but still encode a His (see *SI Appendix*, Table S3 for sequence). The synthesized gene was PCR amplified with and without a stop codon and cloned into the *Nde*I/*Xho*I sites of pET21a to get the pET21a-*CLH1* and pET21a-*CLH1-His* plasmids, respectively. The *E. coli menH* gene was cloned into the *Nde*I/*Xho*I sites of pET28a to generate the pET28a-*menH* plasmid. Primers used in this study are listed in *SI Appendix*, Table S4.

Construction of *Synechocystis* and *Rvi. gelatinosus* strains. Plasmids based on pPD-*N*FLAG or pPD-*C*FLAG were introduced to *Synechocystis* by natural transformation and transformants were selected on BG11 agar supplemented with 10 μg·mL^{-1} kanamycin. Full segregation was achieved by incrementally doubling the kanamycin concentration to a final concentration of 80 μg·mL^{-1} and

confirmed by colony PCR using the psbAIIflankF/psbAIIflankR primers. We were unable to clone the synthesized *acsF*^{MED4} and *acsF*^{9313} gene fragments (GenScript) (see *SI Appendix*, Table S3 for sequence) into pPD-*N*FLAG, suggesting toxicity of their products in *E. coli*. Instead, overlap extension PCR was used to fuse the *acsF*^{MED4} or *acsF*^{9313} gene with the 430-bp sequence upstream of the *Nde*I site and the 1604-bp sequence downstream of *Bgl*II site of pPD-*N*FLAG vector. The resulting PCR products were verified by sequencing before introduction to *Synechocystis*. Disruption of the *cycI* gene was conducted as described previously (4). Similarly, truncation or deletion of slr1916 was achieved by inserting a GTAA sequence and a chloramphenicol resistance cassette (CmR) between nucleotides 386 and 387, or replacing the whole gene with CmR, respectively. Transformants were selected and segregated on BG11 agar with the chloramphenicol concentration incrementally doubled from 5 to 80 μg·mL^{-1}. Full segregation was confirmed by colony PCR using the 1214UpF/1214DownR and 1214insideF/1214insideR primers for the *cycI* gene, and the 1916UpF/1916DownR primers for slr1916.

The pBB[*gene*] plasmids were conjugated into the *Rvi. gelatinosus* Δ*bchE* Δ*acsF Rif*^{R} mutant via the *E. coli* S17-1 strain. *E. coli* S17-1 cells harbouring the plasmid were grown in LB medium with 30 μg·mL^{-1} kanamycin at 37 °C for 24 h. Thirty microliters of the resulting *E. coli* culture were mixed with *Rvi. gelatinosus* cells, which were harvested from a 30 mL culture and resuspended in 100 μL of LB medium. The mating mixture was spotted onto an LB agar plate, which was incubated at 30 °C overnight prior to streaking onto PYS agar supplemented with 40 μg·mL^{-1} rifampicin (to prevent growth of *E. coli* S17-1) and 50 μg·mL^{-1} kanamycin (to select for transconjugants).

Genome sequencing and variant calling. *Synechocystis* cells grown on BG11 agar plates were harvested and treated with saturated NaI solution to remove extracellular polysaccharides (5) and then used for genomic DNA isolation as described previously (6) with some modifications. Briefly, cells were treated with lysozyme (Sigma-Aldrich) and proteinase K (Sigma-Aldrich), followed by precipitation with cetrimonium bromide. The resulting solution was subjected to phenol-chloroform

extraction twice with phenol:chloroform:isoamyl alcohol (25:24:1, vol/vol/vol, saturated with 10 mM Tris-HCl pH 8.0, 1 mM EDTA; Sigma-Aldrich). Residual protein contamination was removed by precipitation with chloroform. Genomic DNA was recovered by isopropanol precipitation, washed with 70% (vol/vol) ethanol, and dissolved in 10 mM Tris-HCl pH 8.0 and 1 mM EDTA. RNA contamination was eliminated by RNase treatment, followed by genomic DNA recovery as described above. Genomic DNA was fragmented by nebulisation with N_2 gas and used for construction of a DNA library for paired-end sequencing using the NexteraTM DNA Library Preparation Kit (Illumina) with a median insert size of ~300 bp. The constructed library was subjected to 100-bp paired-end sequencing on an Illumina HiSeq 2000 platform according to the manufacturer's instructions. 9.92, 11.21, 7.03, and 6.64 million reads were obtained for WT, Δ*ycf54*, SM1, and SM2, respectively. Considering the *Synechocystis* genome has a size of 3.6 Mb (7), these sequencing data corresponds to 502-, 567-, 178-, and 168-fold coverage of the WT, Δ*ycf54*, SM1, and SM2 genomes, respectively.

Variants were called using the mapping-based method. The chromosomal sequence of the GT-S strain (NC_017277) (8), as well as the sequences of the four large plasmids (pSYSM, NC_005229; pSYSA, NC_005230; pSYSG, NC_005231; pSYSX, NC_005232) (9) and the three small plasmids (pCA2.4, NC_020289; pCB2.4, NC_020298; pCC5.2, NC_020290) (10), was used as reference. Each read was mapped to the references using BWA (11) version 0.7.12 with default options. Duplicates were removed using Picard (http://broadinstitute.github.io/picard/) version 1.139 and indel intervals were locally realigned with GATK (12, 13) version 3.5. Then single-nucleotide polymorphism (SNP) and indel variants were called using the HaplotypeCaller tool from GATK with the parameter ploidy set as 1. To further reduce false-positive errors, variants were filtered according to the following criteria: mapping quality > 0, quality score > 30, approximate read depth > 20, quality by depth > 2, genotype quality > 60 and read support > 50%. The effects of putative genetic variants were predicted using SnpEff (14) version 4.2. The variants found in the suppressor mutants but not in the Δ*ycf54* strain were identified as putative suppressor mutations and listed in *SI Appendix*, Table S2. Both SM1 and SM2

harbour a D219G mutation in the *cycI* gene and truncations in the slr1916 ORF. These variants were validated by sequencing the surrounding region following PCR amplification with the 1214seqF/1214seqR primers for the D219G mutation, and 1916seqF/1916seqR primers for the slr1916 truncations.

Protein electrophoresis and immunodetection. Membrane proteins from cyanobacteria were denatured at room temperature for 30 min with 2% (wt/vol) SDS and 1% (wt/vol) dithiothreitol and analysed by SDS-PAGE. Proteins were stained by Coomassie Brilliant Blue, or with SYPRO Orange if to subsequently be transferred onto a polyvinylidene difluoride (PVDF) membrane. Membranes were incubated with specific primary antibodies (1) and then with secondary antibody conjugated with horseradish peroxidase (Sigma-Aldrich). Chemiluminescent signal was developed using the WESTAR SUN enhanced chemiluminescence substrate (Cyanagen) or Luminata Crescendo Western HRP substrate (Merck Millipore), and detected by an Amersham Imager 600 (GE Healthcare) or an LAS 4000 (Fuji). To assess the reactivity of the antibody (Agrisera) raised against the *Arabidopsis* AcsF homolog, $AcsF^{Rg}$ and CycI were recombinantly produced in *E. coli* and purified from inclusion bodies (15). *E. coli* BL21(DE3) pLysS cells, harboring either the pET3a-$acsF^{Rg}$ or the pET3a-*cycI* plasmid (16), were grown at 37 °C with shaking at 220 rpm for 2 h 30 min, followed by induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside at 30 °C with shaking at 175 rpm for 3 h 30 min. Cells were harvested, resuspended in 50 mM Tris-HCl pH 8.0 and 5 mM EDTA, and disrupted by sonication. The lysate was centrifuged at 16,602 × *g* at 4 °C for 10 min. The pellet containing inclusion bodies was washed once in 50 mM Tris-HCl pH 8.0, 5 mM EDTA and 2% (wt/vol) sodium deoxycholate, and then solubilized in 2% (wt/vol) SDS. Protein concentration was estimated by absorbance at 280 nm using an extinction coefficient of 1 $mg^{-1}{\cdot}mL{\cdot}cm^{-1}$. Protein samples were analyzed by SDS-PAGE and stained with Coomassie Brilliant Blue or blotted onto a PVDF membrane followed by immunodetection as described above.

Phylogenetic analyses. We employed two phylogenetic analyses to compare the evolutionary history of AcsF proteins and their parental organisms using an identical representative set of 103 organisms ranging from Acidobacteria, photosynthetic Proteobacteria, Chloroflexi, and Cyanobacteria to plant and algal plastids. The first tree was based on alignments of AcsF proteins, while the second tree was inferred from 13 universally conserved proteins selected from those used previously for studies of plastid evolution (17), which are AtpA, AtpB, AtpH, Rpl2, Rpl14, Rpl16, RpoB, Rps2, Rps3, Rps4, Rps7, Rps11, and Rps19. First, amino acid sequences of AscFI/II and the 13 conserved proteins were mined from the target set of 103 genomes. Hits for each protein were aligned using MAFFT version 7 (18) and the alignments were manually reviewed to remove ambiguous sites and gap regions. A maximum likelihood tree was calculated using each of the 13 alignments of the universally conserved proteins with 1000 bootstrap pseudoreplications in RaxML version 8 (19). The resulting phylogenies were inspected manually for topological incongruence. As individual protein trees were consistent, all alignments were concatenated. The final AcsF alignment spanned 394 positions while the concatenated conserved multilocus alignment ranged 3182 positions. A Bayesian Inference (BI) tree was obtained for each of the two alignments separately using Mrbayes version 3.2.6 (20). Two independent runs of eight Markov chains were performed for a million generations, sampling every 100th tree, until the likelihood values were stable and the divergence criterion was lower than 0.01. The BI calculation employed a common LG+I+G substitution model, posterior probabilities were estimated from branch frequencies in the sampled trees, discarding the first 25% of the harvested data as burn-in. The ML and BI calculations were run via the CIPRES supercomputing facility (21).

References

1. S. Hollingshead *et al.*, Conserved chloroplast open-reading frame *ycf54* is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in *Synechocystis* PCC 6803. *J. Biol. Chem.* **287**, 27823–27833 (2012).

2. J. W. Chidgey *et al*., A cyanobacterial chlorophyll synthase-HliD complex associates with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* **26**, 1267–1279 (2014).
3. I. B. Tikh, M. Held, C. Schmidt-Dannert, BioBrick^{TM} compatible vector system for protein expression in *Rhodobacter sphaeroides*. *Appl. Microbiol. Biotechnol.* **98**, 3111–3119 (2014).
4. G. E. Chen, D. P. Canniffe, C. N. Hunter, Three classes of oxygen-dependent cyclase involved in chlorophyll and bacteriochlorophyll biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6280–6285 (2017).
5. J. G. K. Williams, Construction of specific mutations in photosystem II photosynthetic reaction center by genetic engineering methods in *Synechocystis* 6803. *Methods Enzymol*. **167**, 766–778 (1988).
6. K. Wilson, Preparation of genomic DNA from bacteria. *Curr. Protoc. Mol. Biol.* **56**, 2.4.1–2.4.5 (2001).
7. T. Kaneko *et al*., Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 185–209 (1996).
8. N. Tajima *et al*., Genomic structure of the cyanobacterium *Synechocystis* sp. PCC 6803 strain GT-S. *DNA Res.* **18**, 393–399 (2011).
9. T. Kaneko *et al*., Structural analysis of four large plasmids harboring in a unicellular cyanobacterium, *Synechocystis* sp. PCC 6803. *DNA Res.* **10**, 221–228 (2003).
10. D. Trautmann, B. Voss, A. Wilde, S. Al-Babili, W. R. Hess, Microevolution in cyanobacteria: re-sequencing a motile substrain of *Synechocystis* sp. PCC 6803. *DNA Res.* **19**, 435–448 (2012).
11. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
12. A. McKenna *et al*., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
13. M. A. DePristo *et al*., A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
14. P. Cingolani *et al*., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}, *iso*-2; *iso*-3. *Fly (Austin)* **6**, 80–92 (2012).
15. A. K. Patra *et al*., Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. *Protein Expr. Purif.* **18**, 182–192 (2000).

16. G. E. Chen *et al*., Complete enzyme set for chlorophyll biosynthesis in *Escherichia coli*. *Sci. Adv.* **4**, eaaq1407 (2018).
17. P. M. Shih *et al*., Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1053–1058 (2013).
18. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
19. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
20. F. Ronquist *et al*., Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
21. M. A. Miller *et al*., A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform.* **11**, 43–48 (2015).

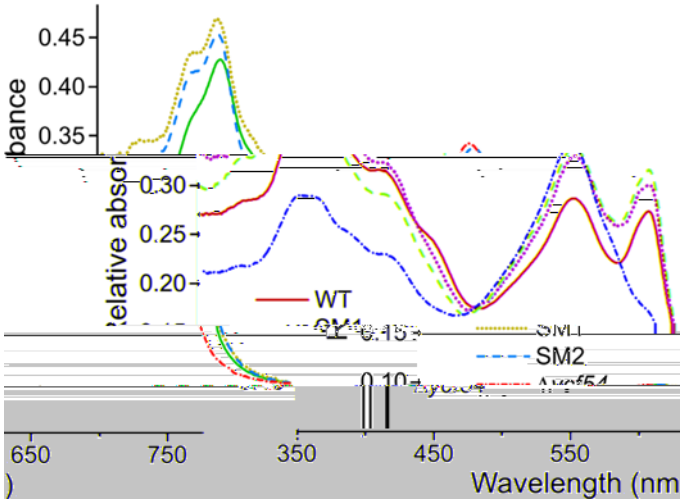

Figure S1. Comparison of the whole-cell absorption spectra of SM1 and SM2 with the WT and Δ*ycf54* strains

Strains were grown autotrophically under SL except for the Δ*ycf54* strain, which was grown mixotrophically under LL.

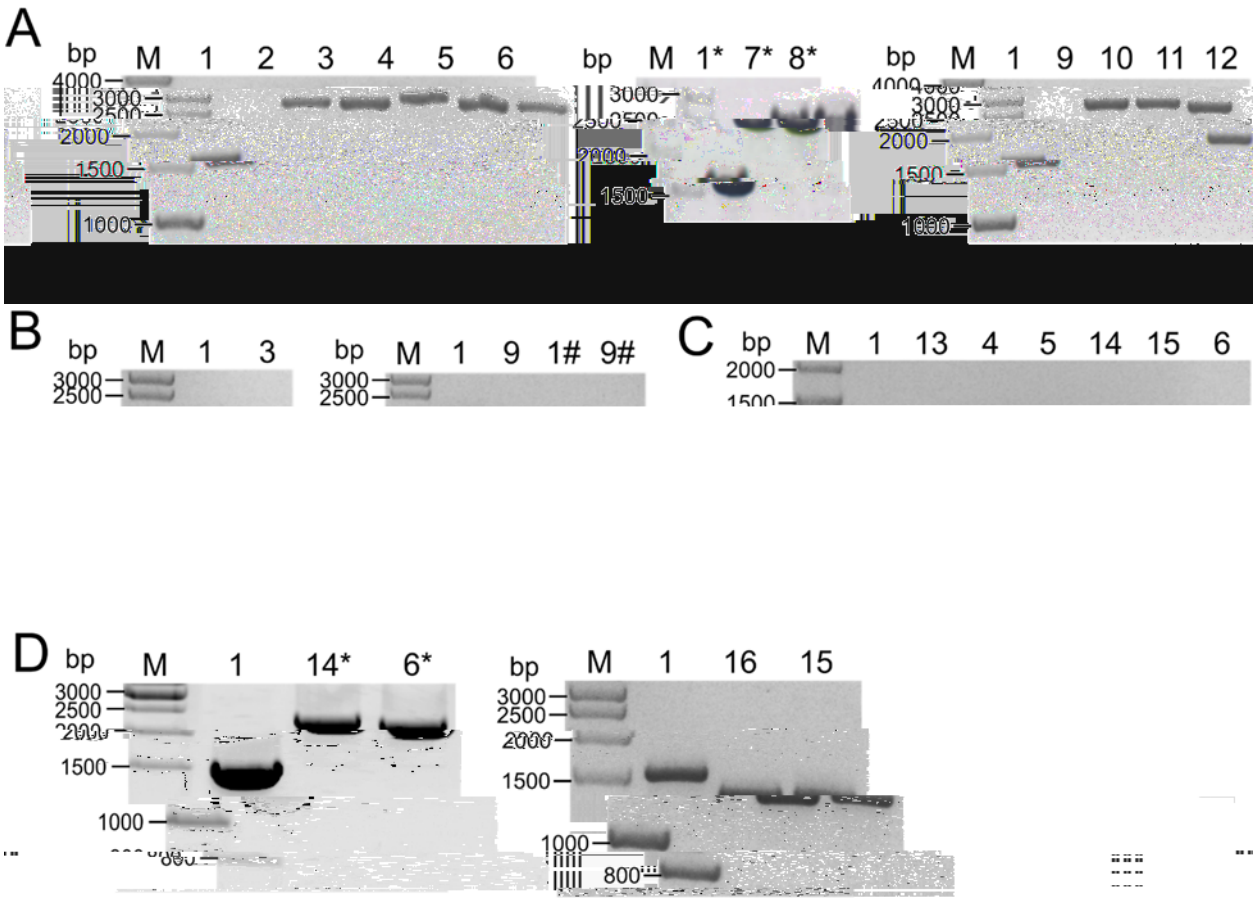

Figure S2. Colony PCR screening of the segregation level of *Synechocystis* strains generated in this study

Colony PCR was conducted using cells streaked from cryo-stocks onto BG11 agar in the presence (unmarked lanes) or absence (lanes marked with an asterisk) of 5 mM glucose. M, DNA marker (HyperLadder 1kb, Bioline). 1, WT; 2, *cycI*$^{SM+}$; 3, *cycI*$^{+}$Δ*cycI*; 4, Δ*ycf54 cycI*$^{+}$; 5, Δ*ycf54 cycI*$^{SM+}$; 6, Δ*ycf54 cycI*$^{SM+}$ slr1916^{SM}; 7, *FLAG*-slr1916^{+}; 8, slr1916-*FLAG*$^{+}$; 9, *acsF*$^{MED4+}$ Δ*cycI*; 10, Δ*ycf54 acsF*$^{MED4+}$; 11, Δ*ycf54 acsF*$^{9313+}$; 12, Δ*ycf54 ycf54*$^{MED4+}$; 13, Δ*ycf54*; 14, Δ*ycf54* slr1916^{SM}; 15, Δ*ycf54* Δslr1916; 16, Δslr1916. (*A*) The *psbAII* locus. (*B*) The *cycI* locus checked with a set of flanking primers and a set of primers internal to the deleted region of *cycI* (indicated by #). (*C*) The *ycf54* locus. (*D*) The slr1916 locus.

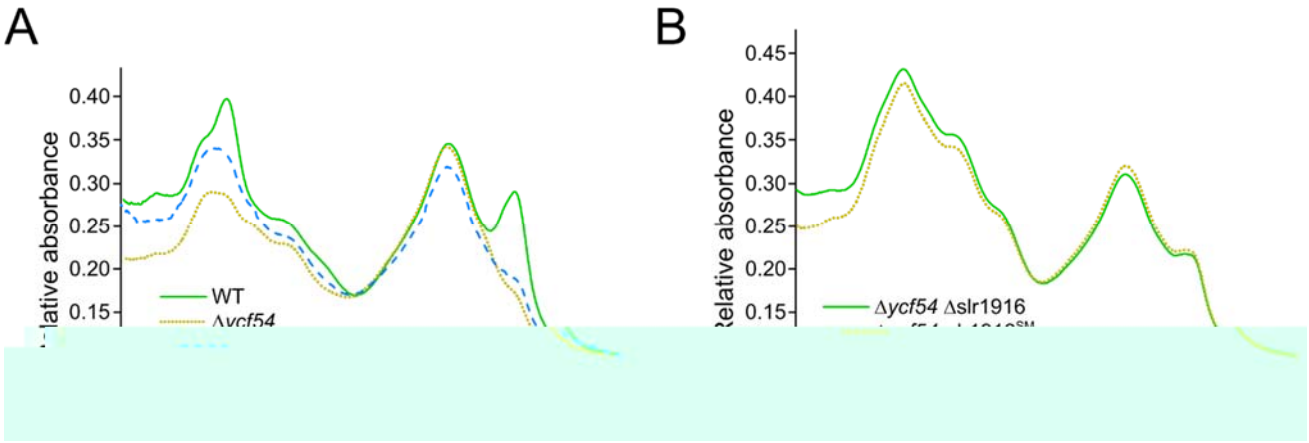

Figure S3. Whole-cell absorption spectra of the indicated *Synechocystis* strains

Strains were grown mixotrophically under LL (*A*) and autotrophically under SL (*B*).

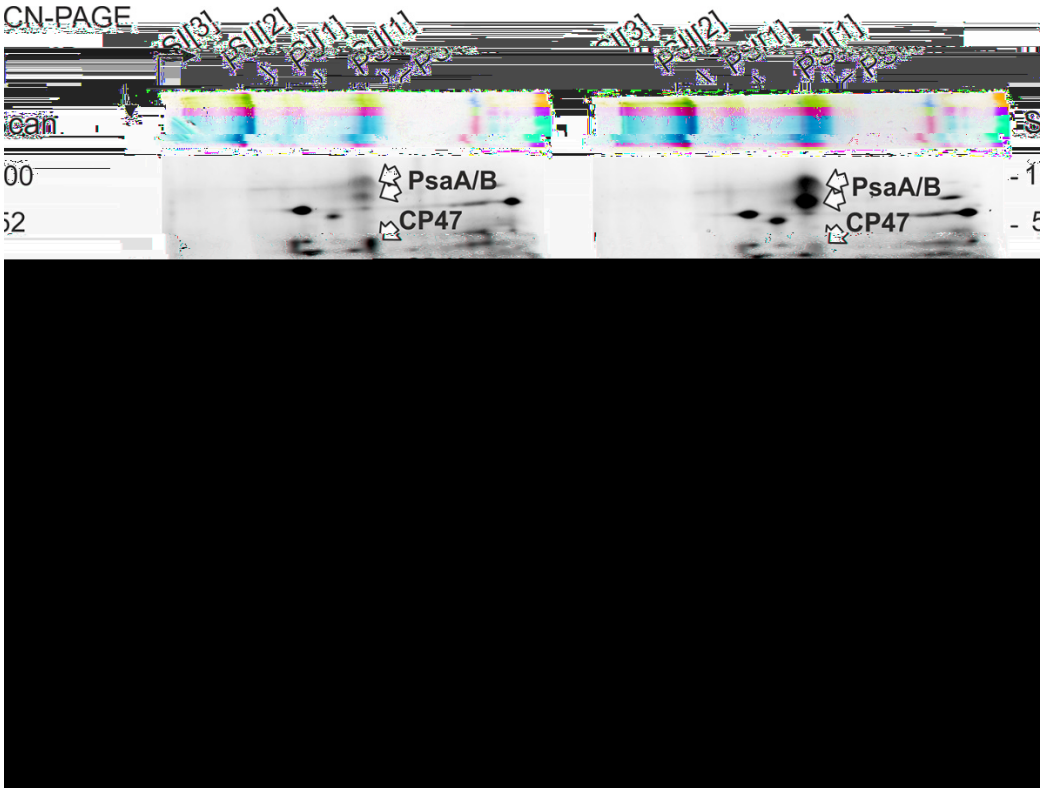

Figure S4. Analysis of membrane proteins isolated from the WT and Δ*ycf54 cycI*$^{SM+}$ slr1916^{SM} strains by 2D CN/SDS-PAGE

Strains were grown autotrophically under SL. Membrane fractions were isolated and solubilized before analysis by CN-PAGE with loading on an equal cell number basis. Pigmented complexes were detected by their colour (Scan); PSI[1] and PSI[3] indicate monomeric and trimeric PSI, respectively; PSII[1] and PSII[2] indicate monomeric and dimeric PSII, respectively. Subsequently, proteins were separated in the second dimension by SDS-PAGE and the resulting gel was stained with SYPRO Orange. PSI and PSII subunits are marked by hollow arrows. Note that PsaA/B subunits from the PSI trimer do not usually migrate into the SDS-PAGE gel.

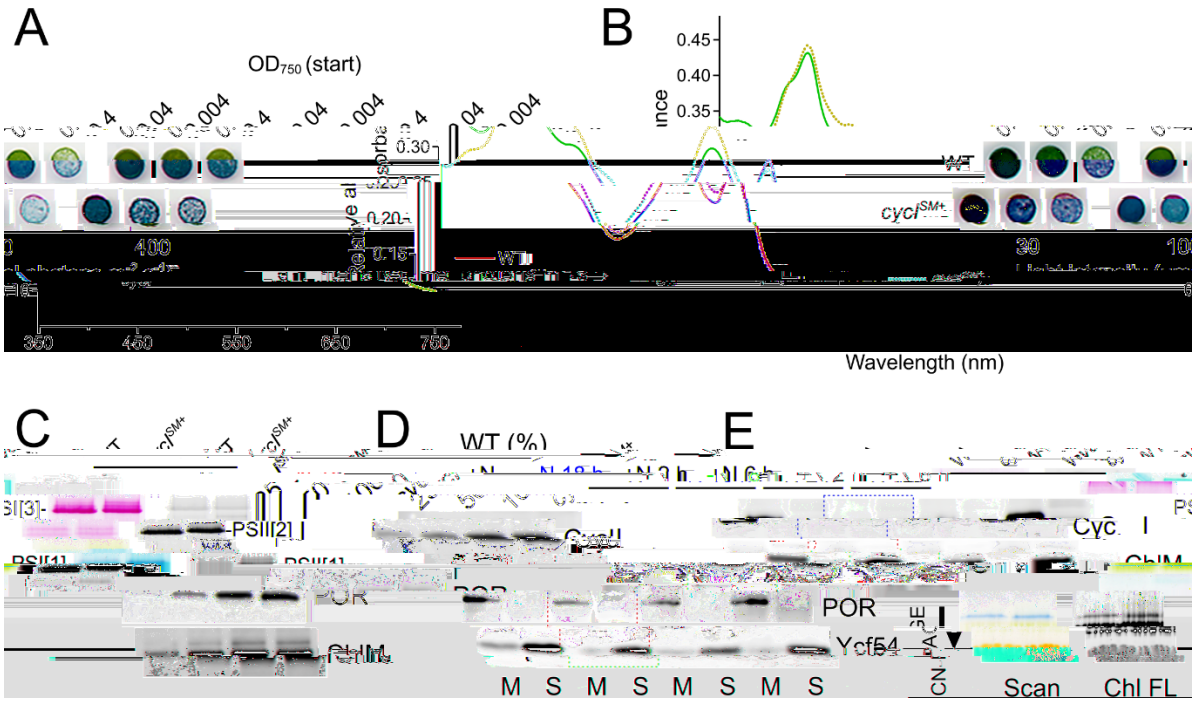

Figure S5. Analysis of the *cycI*$^{SM+}$ strain

(*A*) Drop growth assays of the described strains grown on BG11 agar under different light intensities. Photographs were taken after incubation for 6 d. (*B*) Whole-cell absorption spectra of the indicated strains grown autotrophically under SL. (*C*) CN-PAGE separation of membrane proteins isolated from the indicated strain grown autotrophically under SL. The loading corresponds to the same number of cells from each strain. Pigmented complexes were detected by their colour (Scan) and Chl fluorescence (Chl FL) with excitation by blue light. PSI[1] and PSI[3] indicate monomeric and trimeric PSI, respectively; PSII[1] and PSII[2] indicate monomeric and dimeric PSII, respectively. (*D*) Comparison of the levels of selected Chl biosynthetic enzymes in the WT and *cycI*$^{SM+}$ strains. Membrane fractions were isolated and analyzed by SDS-PAGE with loading on an equal cell number basis, followed by immunodetection with protein specific antibodies. The WT sample was also loaded at 25% and 50% levels for ease of comparison. (*E*) Immunodetection of indicated Chl biosynthetic enzymes in the *cycI*$^{SM+}$ strain upon nitrogen depletion and subsequent restoration with 10 mM $NaNO_3$. Cells were collected before (+N) and after 18 h nitrogen starvation, and after 2 and 6 h nitrogen restoration.

Membrane (M) and soluble (S) protein fractions were isolated from the collected cells and loaded on an equal cell number basis for SDS-PAGE, followed by immunodetection.

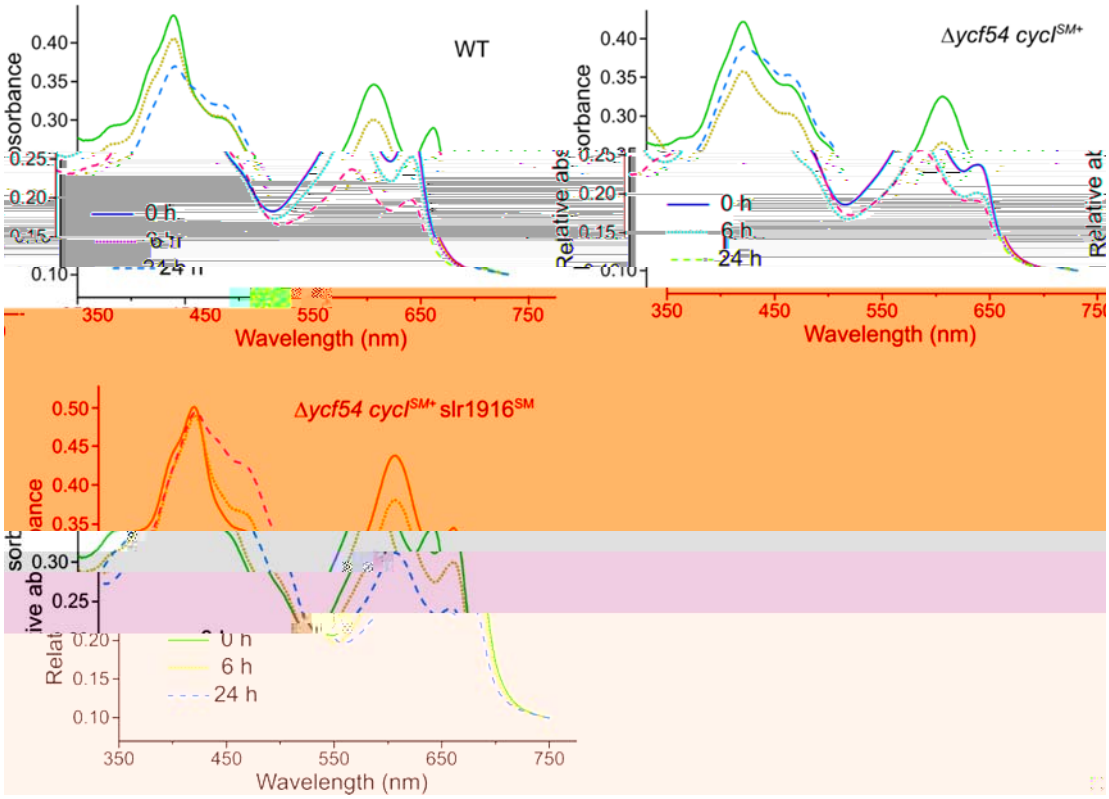

Figure S6. Whole-cell absorption spectra of *Synechocystis* strains before and after treatment with gabaculine

Strains were grown autotrophically under SL and whole-cell absorption spectra were measured before (0 h) and after 6 and 24 h treatment with 5 µM gabaculine.

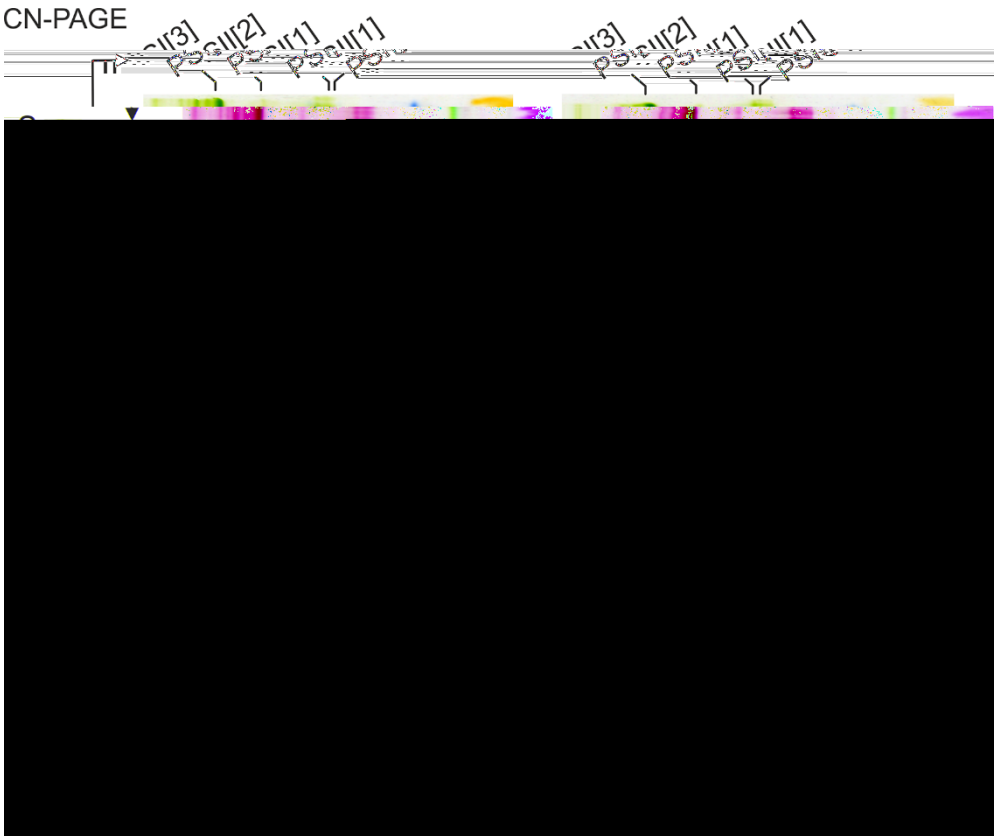

Figure S7. Analysis of membrane proteins isolated from the WT and Δslr1916 strains by 2D CN/SDS-PAGE

Strains were grown autotrophically under SL. Membrane fractions were isolated and solubilized before analysis by CN-PAGE with loading on an equal cell number basis. Pigmented complexes were detected by their colour (Scan). Proteins were subsequently separation in the second dimension by SDS-PAGE and the resulting gel was stained with Coomassie Brilliant Blue. PSI[1] and PSI[3] indicate monomeric and trimeric PSI, respectively; PSII[1] and PSII[2] indicate monomeric and dimeric PSII, respectively. Individual PSI and PSII subunits are marked by hollow arrows. Note that the PsaA/B subunits from the PSI trimer do not usually migrate into the SDS-PAGE gel. PsaC was detectable only in the Δslr1916 strain due to the higher level of PSI in this strain.

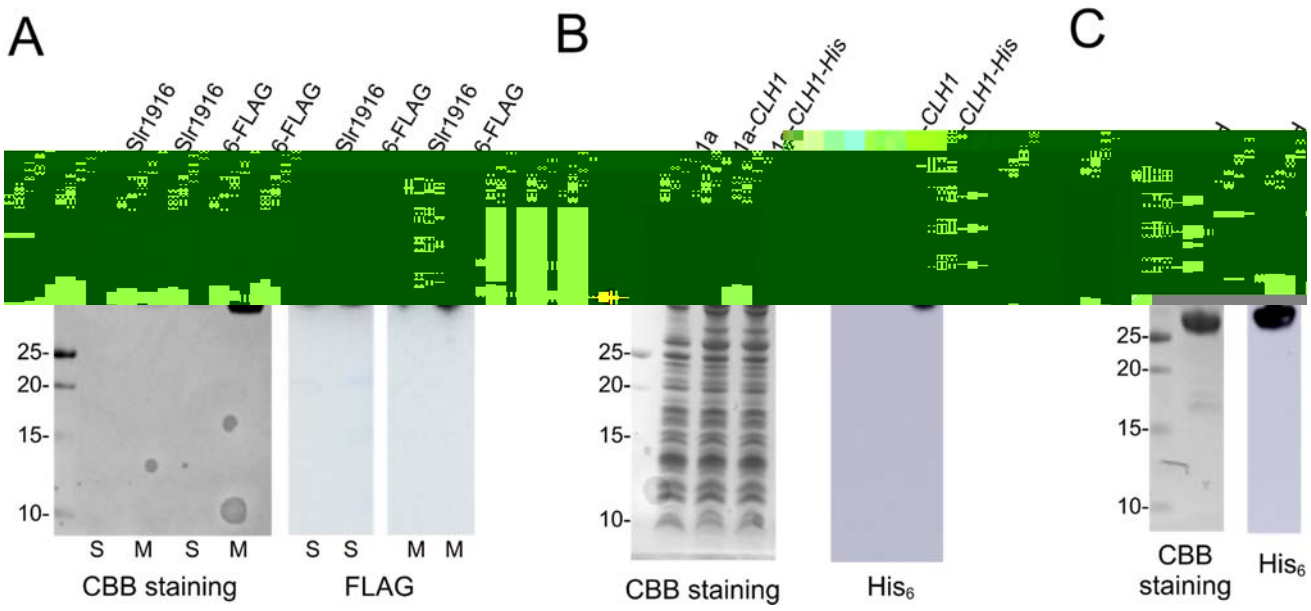

Figure S8. Production of *Synechocystis* Slr1916, *Arabidopsis* CLH1 and *E. coli* MenH

Samples were analyzed by SDS-PAGE, followed by either Coomassie Brilliant Blue (CBB) staining or transfer to a PVDF membrane for immunodetection using antibodies that recognize the FLAG or His_6 tag. (*A*) Immunoprecipitation of FLAG-tagged Slr1916 from *Synechocystis*. The soluble (S) and detergent solubilized membrane fractions (M) were applied to FLAG affinity resin and washed prior to elution with the FLAG peptide. For immunodetection the eluate from the soluble fraction was loaded at 10× the level of that of the membrane fractions. (*B*) Production of recombinant *Arabidopsis* CLH1. Cell lysates from *E. coli* containing empty pET21a vector, pET21a-*CLH1* or pET21a-*CLH1-His* were analyzed by SDS-PAGE. The red arrow indicates the more prominent band in pET21a-*CLH1* and pET21a-*CLH1-His* samples due to production of CLH1, as demonstrated by immunodetection of the His_6 tag. (*C*) Purification of *E. coli* MenH. SDS-PAGE of MenH purified by Ni-affinity and size exclusion chromatography and immunodetection of the His_6 tag.

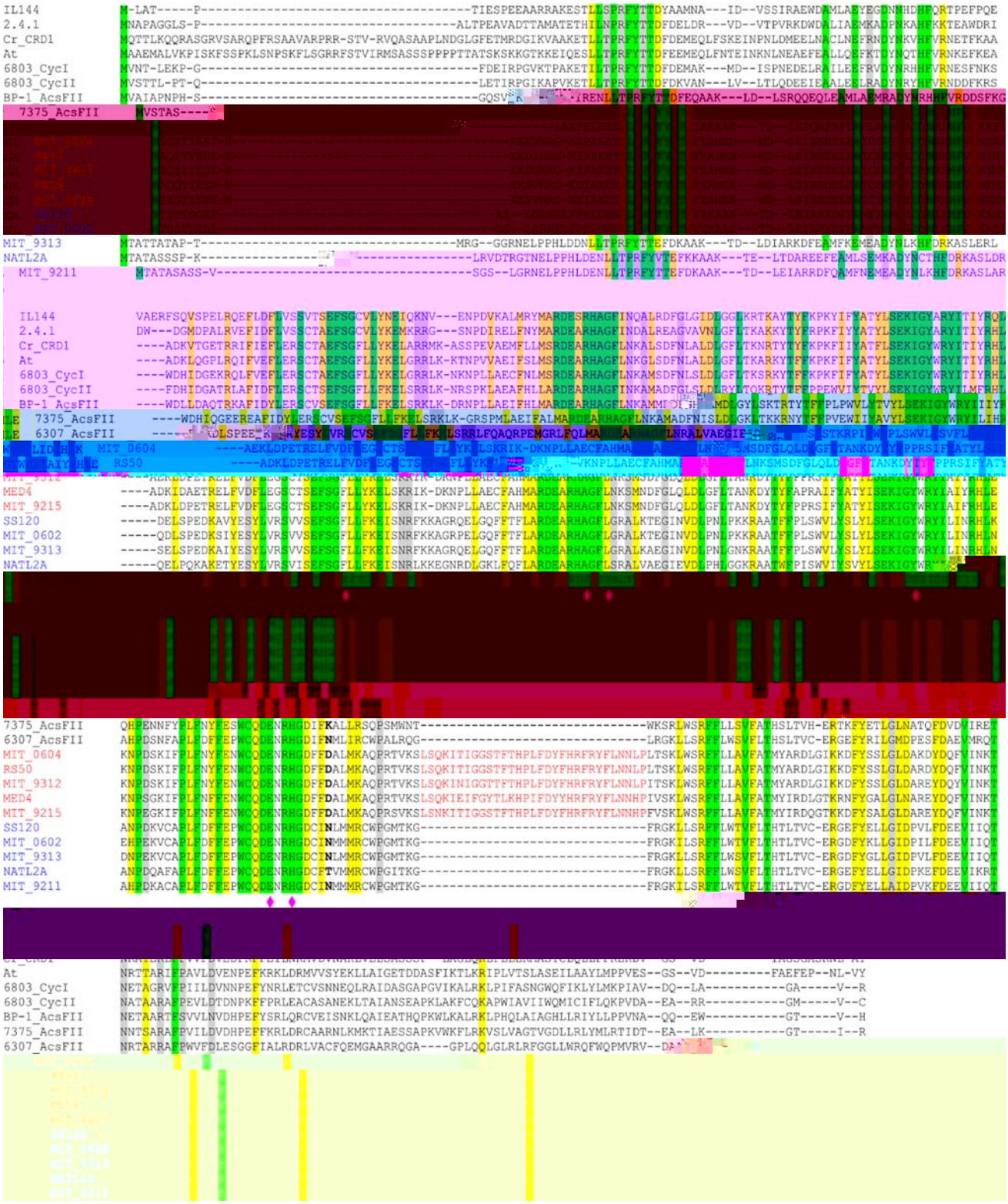

Figure S9. Amino acid sequence alignments of AcsF proteins

Sequences are those from *Rvi. gelatinosus* IL144 (IL144), *Rhodobacter sphaeroides* 2.4.1 (2.4.1), *Chlamydomonas reinhardtii* (Cr_CRD1), *Arabidopsis thaliana* (At), *Synechocystis* sp. PCC 6803 (6803_CycI and 6803_CycII), *Thermosynechococcus elongatus* BP-1 (BP-1_AcsFII), *Leptolyngbya*

sp. PCC 7375 (7375_AcsFII), *Cyanobium gracile* sp. PCC 6307 (6307_AcsFII), 5 HL-adapted *Prochlorococcus* ecotypes (MIT0604, RS50, MIT9312, MED4, MIT9215; colored in red) and 5 LL-adapted *Prochlorococcus* ecotypes (SS120, MIT0602, MIT9313, NATL2A, MIT9211; colored in blue). Conserved, highly similar and similar residues are highlighted in green, yellow and grey, respectively. The putative diiron binding ligands are marked by magenta diamonds. Residues at the equivalent position to the *Synechocystis* CycI D219 residue are indicated in bold. The 31 aa inserts present in AcsF proteins from HL-adapted *Prochlorococcs* ecotypes are marked in red.

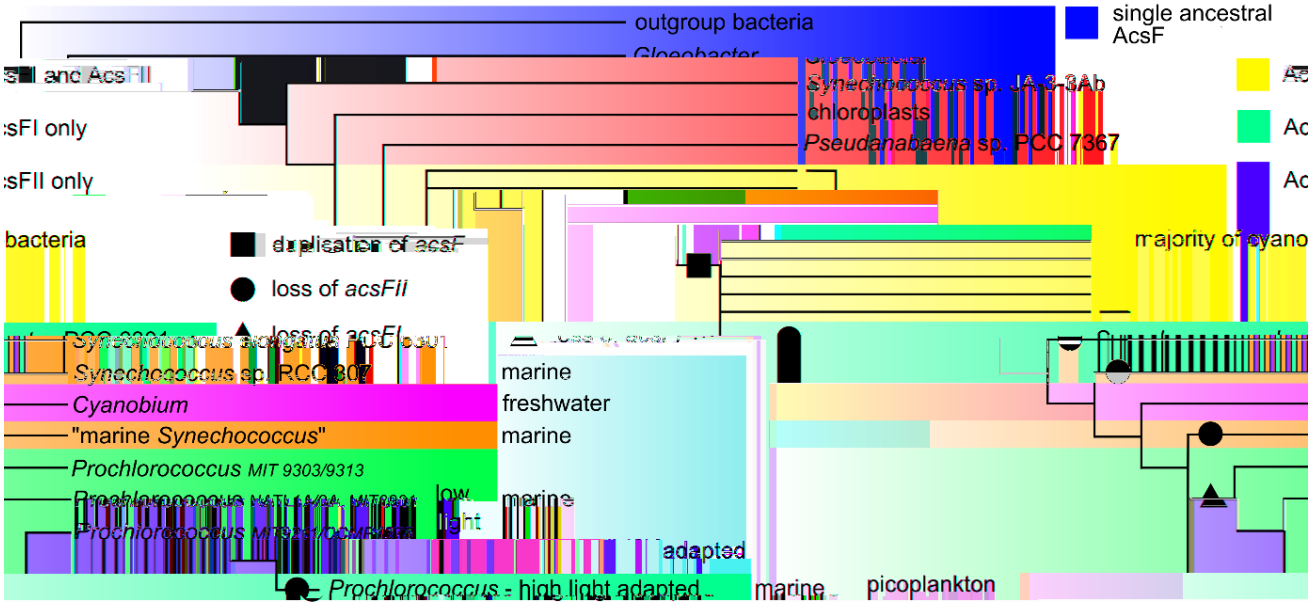

Figure S10. Proposed evolutionary scheme of two cyanobacterial AcsF homologs.

The hypothesis is based on phylogenetic reconstructions of multiple AcsF loci (Fig. 9 in the main text). AcsFI and AcsFII evolved from a single bacterial ancestor by duplication in a deep lineage leading towards most modern cyanobacteria. However, the duplication event occurred only after the emergence of chloroplasts, explaining the absence of AcsFII in plants and algae. Multiple sequential losses of the *acsF* gene copies are predicted in picocyanobacterial clades, presumably due to genome streamlining. Our phylogenetic reconstruction suggests that the highly modified AcsF found in LL-adapted *Prochlorococcus* has evolved from an AcsFII ancestor.

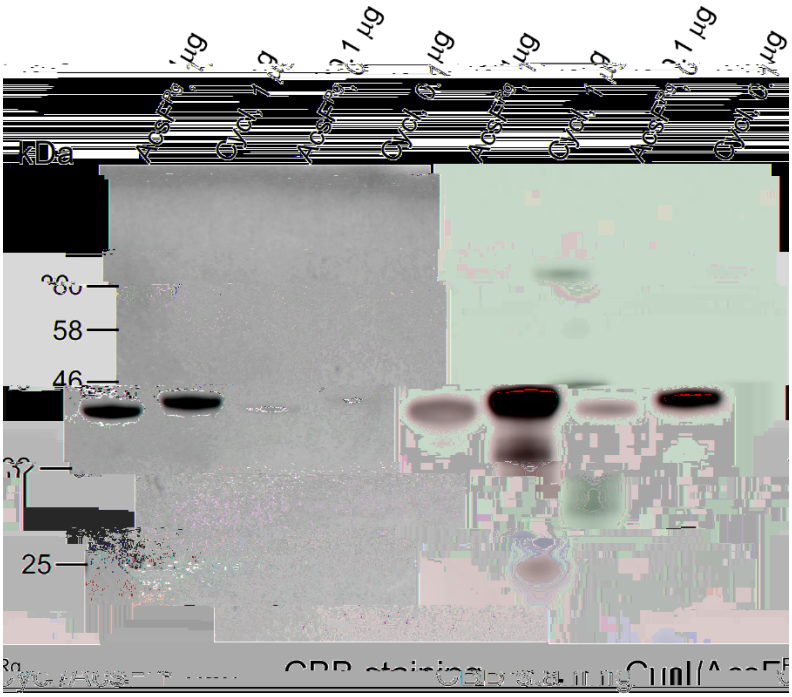

Figure S11. Reactivity of the anti-AcsF antibody against AcsF^{Rg} and CycI

Indicated amounts of purified AcsF^{Rg} and CycI proteins were analyzed by SDS-PAGE. Protein bands were visualized by Coomassie Brilliant Blue (CBB) or transferred to a PVDF membrane for immunodetection (AcsF) using an antibody raised against the *Arabidopsis* AcsF homolog.

Table S1. Strains and plasmids described in this study

Strain/Plasmid	Genotype/characteristics	Source
E. coli		
JM109	Cloning strain for plasmid construction	Promega
S17-1	Conjugation strain for pBBRBB-*Ppuf*$_{843–1200}$ constructs	ref. 1
BL21(DE3)	Expression strain for production of CLH1 and MenH	Novagen
BL21(DE3) pLysS	Expression strain for production of CycI and AcsF^{Rg} polypeptides	Novagen
Rvi. gelatinosus		
WT	IL144	S. Nagashima*
Δ*bchE* Δ*acsF*		

pBBRBB-*Ppuf*$_{843–1200}$	Expression vector carrying the 843–1200 region of the *Rba. sphaeroides puf* promoter, Km^R	ref. 6
pBB[*cycI*]	*cycI* cloned into *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$	This study
pBB[*cycI*^{SM}]	*cycI*^{SM} cloned into *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$	This study
pBB[*cycI-ycf54*]	*cycI-ycf54* cloned into *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$	This study
pBB[*cycI*^{SM}*-ycf54*]	*cycI*^{SM}*-ycf54* cloned into *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$	This study
pBB[*acsF*^{Rg}]	*acsF*^{Rg} cloned into *Bgl*II/*Not*I sites of pBBRBB-*Ppuf*$_{843–1200}$	ref. 7
pET3a	Expression vector carrying T7 promoter, Amp^R	Novagen
pET3a-*acsF*^{Rg}	*acsF*^{Rg} with an added *Spe*I site cloned into *Nde*I/*BamH*I sites of pET3a	ref. 7
pET3a-*cycI*	*cycI* with an added *Spe*I site cloned into *Nde*I/*BamH*I sites of pET3a	ref. 7
pET21a	Expression vector carrying T7*lac* promoter, Amp^R	Novagen
pET21a-*CLH1*	*CLH1* (with stop codon) cloned into *Nde*I/*Xho*I sites of pET21a	This study
pET21a-*CLH1-His*	*CLH1* (without stop codon) cloned into *Nde*I/*Xho*I sites of pET21a	This study
pET28a	Expression vector carrying T7*lac* promoter, Km^R	Novagen
pET28a-*menH*	*E. coli menH* cloned into *Nde*I/*Xho*I sites of pET28a	This study

*Research Institute for Photosynthetic Hydrogen Production, Kanagawa University, Yokohama, Japan.

†School of Life Sciences, Arizona State University, AZ 85281.

1. R. Simon, U. Priefer, A. Puhler, A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in gram negative bacteria. *Nat. Biotechnol.* **1**, 784–791 (1983).
2. G. E. Chen, D. P. Canniffe, C. N. Hunter, Three classes of oxygen-dependent cyclase involved in chlorophyll and bacteriochlorophyll biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6280–6285 (2017).
3. S. Hollingshead *et al*., Synthesis of chlorophyll-binding proteins in a fully segregated Δ*ycf54* strain of the cyanobacterium *Synechocystis* PCC 6803. *Front. Plant Sci.* **7**, 292 (2016).
4. S. Hollingshead *et al*., Conserved chloroplast open-reading frame *ycf54* is required for activity of the magnesium protoporphyrin monomethylester oxidative cyclase in *Synechocystis* PCC 6803. *J. Biol. Chem.* **287**, 27823–27833 (2012).
5. J. W. Chidgey *et al*., A cyanobacterial chlorophyll synthase-HliD complex associates with the Ycf39 protein and the YidC/Alb3 insertase. *Plant Cell* **26**, 1267–1279 (2014).
6. I. B. Tikh, M. Held, C. Schmidt-Dannert, BioBrick^{TM} compatible vector system for protein expression in *Rhodobacter sphaeroides*. *Appl. Microbiol. Biotechnol.* **98**, 3111–3119 (2014).
7. G. E. Chen *et al*., Complete enzyme set for chlorophyll biosynthesis in *Escherichia coli*. *Sci. Adv.* **4**, eaaq1407 (2018).

Table S2. Mutations found in SM1 and SM2 compared to the Δ*ycf54* strain†

#	Type	Start	End	NT change	AA change	Locus	Gene	Gene Product
SM1								
1	SNP	3613	3613	T→C	D219G	sll1214	*cycI*	O_2-dependent MgPME cyclase
2	Del	45844	45844	T→*	frameshift	slr1494	-	ATP-binding cassette transporter
3	Ins	619995	619996	*→G	frameshift	slr1916	-	probable esterase
4	SNP	759568	759568	T→C	V247A	slr2018	-	hypothetical protein
5	Ins	1065632	1065633	*→A	-	IG ssr2406-sll1360	-	-
6	SNP	2873102	2873102	T→C	V41A	slr0076	-	hypothetical protein
7	Ins	2994522	2994523	*→A	frameshift	slr0114	-	putative member of protein phosphatase 2C
8	SNP	3255694	3255694	C→A	A749D	slr0554	-	hypothetical protein
9	Ins	3364585	3364586	*→C	frameshift	sll1496	-	mannose-1-phosphate guanyltransferase
SM2								
1	SNP	3613	3613	T→C	D219G	sll1214	*cycI*	O_2-dependent MgPME cyclase
2	Ins	619923	619924	*→G	frameshift	slr1916	-	probable esterase
3	Del	1802607	1802607	C→*	frameshift	sll1876	*hemN*	O_2-independent Copro'gen oxidase
4	SNP	1921416	1921416	T→C	I159T	slr1160	-	periplasmic protein with unknown function
5	SNP	2320583	2320583	T→C	Y957C	sll0163	-	Trp-Asp repeat protein
6	Del	2595137	2595137	C→*	frameshift	sll0055	-	processing protease
7	SNP	3190324	3190324	T→C	V247A	slr0531	*ggtD*	glucosylglycerol transport system permease protein
8	SNP	3425395	3425395	A→G	I3V	ssr1256	-	hypothetical protein

†Chromosomal variants with nucleotide position referring to the GT-S sequence (NC_017277) and locus details based on CyanoBase (http://genome.microbedb.jp/cyanobase/). Genes mutated in both suppressor mutants are marked in red. IG, intergenic region.

Table S3. Nucleotide sequences of synthesized genes used in this study

Gene	Sequence (5'-3')
acsF^{MED4}-*ycf54*^{MED4}*	ATGGCCCAGCAGACGATCGAGAGCAACAACAAGAAGTCGGTCAACCGCGGCAAGGACATCGC GAAGGACACGATCCTGACCCCGAACTTCTACACGACCGACTTCGAGGCCATGGAGAAGATGG ACCTGAGCATCAACGAGGACGAGCTGGAGGCGATCTGCGAGGAGTTCCGCAAGGACTACAAC CGCCACCACTTCGTCCGCAACAAGGAGTTCGAGGGCGCGGCCGACAAGATCGACGCGGAGAC GCGCGAGCTGTTCGTGGACTTCCTGGAGGGCTCGTGCACCAGCGAGTTCTCGGGCTTCCTGC TGTACAAGGAGCTGAGCAAGCGCATCAAGGACAAGAACCCGCTGCTGGCGGAGTGCTTCGCC CACATGGCCCGCGACGAGGCCCGCCACGCCGGCTTCCTGAACAAGTCGATGAACGACTTCGG CCTGCAGCTGGACCTGGGCTTCCTGACGGCGAACAAGGACTACACCTACTTCGCCCCGCGTG CGATCTTCTACGCCACCTACATCAGCGAGAAGATCGGCTACTGGCGCTACATCGCGATCTAC CGCCACCTGGAGAAGAACCCGTCGGGCAAAATCTTCCCGCTGTTCAACTTCTTCGAGAACTG GTGCCAGGACGAGAACCGCCACGGCGACTTCTTCGACGCCCTGATGAAGGCCCAGCCGCGCA CCGTGAAGTCGCTGAGCCAGAAGATCGAAATCTTCGGCTACACCCTGAAGCACCCGATCTTC GACTACTACCACCGCTTCCGCTACTTCCTGAACAACCACCCGATCGTCAGCAAGCTGTGGTC GCGCTTCTTCCTGCTGGCCGTGTTCGCGACGATGTACATCCGCGACCTGGGCACCAAGCGCA ACTTCTACGGCGCCCTGGGCCTGAACGCCCGCGAGTACGACCAGTTCGTCATCAACAAGACG AACGAGACCAGCGCCAAGGTCTTCCCGGTCGTGCTGAACGTGTACGACAAGTCGTTCTACAA GCGCCTGGACCGCATCGTGGAGAACGGCACGCGCCTGTCGGAGATCGACAAGAAGGAGAACC CGAACGTCATCAAGGTGCTGAGCAAGCTGCCGATCTTCATCTCGAACGGCTACCAGCTGATC CGCCTGTACCTGCTGAAGCCGCTGGAGAGCGACGACTTCCAGCCGTCGATCCGCTAATATAG GAGCTTGGATTATGACGACCTACTTCTTCGTCGCCGCGTCGGAGAAGTTCCTGACGGTGGAG GAGCCGCTGGAGGAGATCCTGAAGGAGCGCATCCGCAACTACAAGGAGAACAAGAAGGAGAT CGACTTCTGGCTGCTGAAGAACCCGTCGTTCCTGAAGTCGAGCGCCTTCCTGGACCTGAGCA AGAAGATCCCGAACACCCCGGCGGCCGTCATCAGCACGGACAAGAAGTTCATCACCTTCCTG AAGCTGCGCCTGGAGTTCGTGGCCGTGGGCGAGTTCGAGTGCCCGAACAGCGAGATCAACGA CCCGTTCAAGGTGGAGTAA
acsF^{9313}†	ATGACCGCCACGACGGCCACGGCCCCGACCATGCGCGGCGGCGGCCGTAACGAGCTGCCGCC GCACCTGGACGACAACCTGCTGACCCCGCGCTTCTACACGACCGAGTTCGACAAGGCGGCCA AGACGGACCTGGACATCGCCCGCAAGGACTTCGAGGCGATGTTCAAGGAGATGGAGGCCGAC TACAACCTGAAGCACTTCGACCGCAAGGCGAGCCTGGAGCGCCTGAGCGAGCTGAGCCCGGA GGACAAGGCCATCTACGAGTCGTACCTGGTCCGCTCGGTCGTGAGCGAGTTCTCGGGCTTCC TGCTGTTCAAGGAGATCAGCAACCGCTTCAAGAAGGCCGGCCGCCAGGAGCTGGGCCAGTTC TTCACCTTCCTGGCCCGCGACGAGGCCCGCCACGCCGGCTTCCTGGGCCGCGCCCTGAAGGC GGAGGGCATCAACGTCGACCTGCCGAACCTGGGCAACAAGCGCGCGGCCACGTTCTTCCCGC TGAGCTGGGTGCTGTACAGCCTGTACCTGTCGGAGAAGATCGGCTACTGGCGCTACATCCTG ATCAACCGCCACCTGAACGACAACCCGGAGAAGGTGTGCGCCCCGCTGTTCGACTTCTTCGA GCCGTGGTGCCAGGACGAGAACCGCCACGGCGACTGCATCAACCTGATGATGCGCTGCTGGC CGGGCATGACCAAGGGCTTCCGCGGCAAGCTGCTGAGCCGCTTCTTCCTGTGGTCGGTCTTC CTGACCCACACCCTGACCGTGTGCGAGCGCGGCGACTTCTACGGCCTGCTGGGCATCGACCC GGTCCTGTTCGACGAGGAAGTCATCATCCAGACCAACAACACGTCGCGCAACGCCTTCCCGT GGGTCTACAACTTCGACGACGGCAAGTTCCTGGAGATGCGCGTGCAGATCCTGAAGGCGTTC CGCAACTGGCGCGAGAGCTCGGGCCTGGCCAAGCCGGTCGCGCTGAGCAAGTTCGTGTCGCT GATCCTGCGCCAGTTCGCCCTGCCGATGCAGAAGACGAACGCGGTCCGCTACGGCTAA
CLH1‡	ATGGCGGCGATAGAGGACAGTCCAACGTTTTCCTCTGTGGTAACTCCGGCGGCTTTTGAGAT AGGCAGCCTCCCGACAACCGAGATACCGGTGGATCCGGTGGAAAATGATTCAACAGCACCGC CAAAACCGGTGAGAATCACCTGTCCAACAGTCGCCGGAACTTATCCCGTCGTTTTATTCTTC CATGGCTTTTATCTTCGCAACTACTTCTACTCTGACGTTCTTAACCACATCGCTTCGCATGG TTACATTCTTGTAGCCCCACAGTTGTGCAAATTATTGCCGCCGGGAGGGCAAGTGGAAGTGG ACGATGCTGGAAGTGTGATAAACTGGGCATCGGAAAACCTCAAAGCTCACCTACCAACTTCG GTAAATGCTAATGGAAAATACACCTCACTCGTGGGCCACAGCCGCGGTGGGAAAACGGCGTT TGCGGTTGCGCTAGGCCATGCCGCAACATTAGACCCATCCATCACGTTTTCAGCTCTAATAG GAATTGATCCAGTCGCAGGAACTAACAAATACATTAGAACCGATCCGCATATCTTAACGTAT AAACCGGAATCTTTCGAGCTGGACATACCGGTTGCAGTGGTGGGAACCGGACTCGGACCGAA GTGGAACAACGTGATGCCACCATGCGCACCAACGGACTTAAACCATGAGGAGTTTTACAAAG AGTGTAAGGCGACGAAAGCCCATTTCGTGGCTGCGGATTACGGACACATGGATATGTTGGAC GATGATTTGCCCGGTTTTGTTGGGTTTATGGCCGGTTGTATGTGTAAGAATGGGCAAAGAAA AAAGTCTGAGATGAGGAGCTTTGTAGGTGGAATTGTGGTTGCGTTTCTCAAGTATAGTTTGT GGGGTGAAAAAGCGGAGATTCGATTGATTGTGAAGGATCCTTCCGTTTCTCCGGCCAAGCTT GATCCTTCACCTGAGTTGGAAGAAGCTTCTGGTATCTTCGTCTAG

*Codon optimized for *Rvi. gelatinosus* and the two genes were synthesized as a single fragment and separated by a 16-bp sequence (underlined), which was designed to provide a ribosome-binding site for *ycf54*^{MED4}.

†Codon optimized for *Rvi. gelatinosus.*

‡ CDS of *Arabidopsis* chlorophyllase-1 (CLH1, AT1G19670) with nucleotide 729 changed from a T to a C to remove an internal *Nde*I site but still encode a His.

Table S4. Primers used in this study

Primer	Sequence (5'-3')
1214seqF	TGTAAAACGACGGCCAGTATGGTTAATACCCTCGAAAAGCC
1214seqR	CAGGAAACAGCTATGACCTTAGCGCACAGCTCCAGCC
1916seqF	TGTAAAACGACGGCCAGTATGCCCACCCTGGATCTTTTGG
1916seqR	CAGGAAACAGCTATGACCTCAGTGATCCGTAGCCAGGATT
1214F_BglII	GAGTCTAGATCTATGGTTAATACCCTCGAAAAGCCC
1214F_NdeI	GGAATTCCATATGGTTAATACCCTCGAAAAGCCCG
1214R_BglII	GAGTCTAGATCTTTAGCGCACAGCTCCAGCCAA
1214R_NotI	GAGTCTGCGGCCGCTTAGCGCACAGCTCCAGCCAAC
1214D219GF	GAGATTTCTTTGGTGCGATTATGCG
1214D219GR	CGCATAATCGCACCAAAGAAATCTC
1214UpF	GCCGATCCGGTTAACCTAGGCA
1214DownR	TGGAGTTGTTGGGAGAGTTCGGTC
1214insideF	GGCCAAGGAAACCATCCTCA
1214insideR	TGGCAAAGACTGAGAGCAGG
1780R_NotI	GAGTCTGCGGCCGCCTAATCCAGGGATGCAAGGGGGT
1780F	GTGGAAAGTTGGGCATTGACG
1780R	CTAATCCAGGGATGCAAGGGG
1874F_NdeI	GGAATTCCATATGGTATCCACTACCCTACCG
1874R_BglII	GAGTCTAGATCTTTAACACACCATCCCCCGAC
psbAIIUpF	AAACGCCCTCTGTTTACCCA
psbAIIDownR	TCAACCCGGTACAGAGCTTC
1916UpF	GGGTGGTGACTATGGAAAATTTG
1916DownR	CACCAAAGCCTAACAGATCAATG
1916SM-CmR1F	CGTGAATGCTGGGGGGGGGTAATACCGGGAAGCCCTGGGC
1916SM-CmR1R	GCCCAGGGCTTCCCGGTATTACCCCCCCCCAGCATTCACG
1916SM-CmR2F	GTGGCAGGGCGGGGCGTAAAAAGTTTCGCTCTGCTGGGG
1916SM-CmR2R	CCCCAGCAGAGCGAAACTTTTTACGCCCCGCCCTGCCAC
1916KO-CmR1F	CGTAGCAATTGCGAGAACTATGGAGAAAAAAATCACTGGATAT
1916KO-CmR1R	ATATCCAGTGATTTTTTTCTCCATAGTTCTCGCAATTGCTACG
1916KO-CmR2F	AGTGGCAGGGCGGGGCGTAAATGGGGGCAATTGTTGGCCGT
1916KO-CmR2R	ACGGCCAACAATTGCCCCCATTTACGCCCCGCCCTGCCACT
psbAIIflankF	CGGTATCGATAAGCTTGATATC
psbAIIflankR	GAATTCGGCTTGATTACGATATC
psbAII-0844-1F	CATAAGGAATTATAACCATATGGCCCAGCAGACGATC
psbAII-0844-1R	GATCGTCTGCTGGGCCATATGGTTATAATTCCTTATG
psbAII-0844-2F	CCAGCCGTCGATCCGCTAAAGATCTTCCTTCAACTCAG
psbAII-0844-2R	CTGAGTTGAAGGAAGATCTTTAGCGGATCGACGGCTGG
psbAII-2196-1F	CATAAGGAATTATAACCATATGACCGCCACGACGGC
psbAII-2196-1R	GCCGTCGTGGCGGTCATATGGTTATAATTCCTTATG
psbAII-2196-2F	CGCGGTCCGCTACGGCTAAAGATCTTCCTTCAACTCAG
psbAII-2196-2R	CTGAGTTGAAGGAAGATCTTTAGCCGTAGCGGACCGCG
1106F_NdeI	GGAACATATGACGACCTACTTCTTCGTC
1106R_BglII	GGAAAGATCTTTACTCCACCTTGAACGGGTC
1916F_NotI	AGAATTCGCGGCCGCACCCACCCTGGATCTTTTGG
slr1916R_BglII	AGTTCAGATCTTCAGTGATCCGTAGCCAGGATTTG

slr1916F_NdeI	ATATGGGCATATGCCCACCCTGGATCTTTTG
slr1916R_NheI	TAAGGGCTAGCGTGATCCGTAGCCAGGATTTGTTG
CLH1F_NdeI	ATATGGGCATATGGCGGCGATAGAGGAC
CLH1Rstop_XhoI	TAAGGCTCGAGGACGAAGATACCAGAAGCTTCTTCC
CLH1R_XhoI	TAAGGCTCGAGCTAGACGAAGATACCAGAAGCTTCTTC
menHF_NdeI	ATATGGGCATATGATCCTGCACGCGCAG
menHR_XhoI	TAAGGCTCGAGTCAGAAACGCAAGATCTGCGC