Systematic identification and characterization of genes in the regulation and biogenesis of photosynthetic machinery

Graphical abstract

Highlights
- Identification of 70 poorly characterized photosynthesis genes in Chlamydomonas
- Proteomic analysis of mutants allows assignment of function to poorly characterized genes
- Characterization of 5 photosystem I mRNA maturation factors validates this resource
- MTF1 and PMR1 identified as master regulators of photosynthesis

Authors
Moshe Kafri, Weronika Patena, Lance Martin, ..., Frederick R. Cross, Martin Wühr, Martin C. Jonikas

Correspondence
mjonikas@princeton.edu

In brief
Kafri et al. used a genetic screen to identify 70 previously uncharacterized genes required for photosynthesis in the model alga Chlamydomonas reinhardtii. Using mutant proteome profiling, they assign many of the genes to pathways, revealing biogenesis and regulatory factors including the master regulator PMR1, which regulates chloroplast genes via nuclear-expressed factors.
Resource

Systematic identification and characterization of genes in the regulation and biogenesis of photosynthetic machinery


1Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA
2Laboratory of Cell Cycle Genetics, The Rockefeller University, New York, NY 10021, USA
3Lewis-Sigler Institute for Integrative Genomics and Department of Chemistry, Princeton University, Princeton, NJ 08544, USA
4Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA
5These authors contributed equally
6Lead contact
*Correspondence: mjonikas@princeton.edu
https://doi.org/10.1016/j.cell.2023.11.007

SUMMARY

Photosynthesis is central to food production and the Earth’s biogeochemistry, yet the molecular basis for its regulation remains poorly understood. Here, using high-throughput genetics in the model eukaryotic alga Chlamydomonas reinhardtii, we identify with high confidence (false discovery rate [FDR] < 0.11) 70 poorly characterized genes required for photosynthesis. We then enable the functional characterization of these genes by providing a resource of proteomes of mutant strains, each lacking one of these genes. The data allow assignment of 34 genes to the biogenesis or regulation of one or more specific photosynthetic complexes. Further analysis uncovers biogenesis/regulatory roles for at least seven proteins, including five photosystem I mRNA maturation factors, the chloroplast translation factor MTF1, and the master regulator PMR1, which regulates chloroplast genes via nuclear-expressed factors. Our work provides a rich resource identifying regulatory and functional genes and placing them into pathways, thereby opening the door to a system-level understanding of photosynthesis.

INTRODUCTION

In photosynthetic eukaryotes, the photosynthetic apparatus consists of a series of protein complexes in the chloroplast thylakoid membrane that use light energy to produce NADPH, ATP, and other cellular energy carriers. NADPH and ATP, in turn, power many pathways, notably CO2 assimilation into sugar by the Calvin-Benson-Bassham metabolic cycle. As a sophisticated system central to cellular fitness, hundreds of genes encoded in both the nucleus and chloroplast are required to assemble these complexes and regulate their activity under nuclear control. In plants and green algae, this coordination is known to involve a range of different mechanisms, including post-transcriptional regulation of chloroplast-expressed genes by nuclear-encoded proteins, translational regulation of chloroplast-expressed subunits by assembly intermediates of photosynthetic complexes, and protease-mediated degradation of unassembled subunits. Although photosynthesis and its regulation have been extensively studied for 70 years, phylogenetics suggests that hundreds of genes participating in photosynthesis remain to be identified and characterized. Indeed, approximately half of the GreenCut2 genes—a set of 597 genes conserved only in the green photosynthetic eukaryotic lineage and therefore likely to be involved in photosynthesis—have not been functionally characterized.

Genetic screens have been done in land plants and algae to identify missing photosynthesis genes. Land plant screens have identified photosynthesis-deficient mutants based on leaf coloration, seedling lethality, and chlorophyll fluorescence. As a complementary system to plants, the leading unicellular model eukaryotic alga Chlamydomonas reinhardtii has provided advantages of higher throughput and physiology that facilitate the identification and characterization of genes essential to photosynthesis. These characteristics have been leveraged to identify and characterize many core components of the photosynthetic electron transport chain.

In the past decade, several hundred candidates for genes involved in photosynthesis have been uncovered by screens of two large Chlamydomonas mutant collections, Niyogi CAL and CLIP. However, these screens had many false positives

https://doi.org/10.1016/j.cell.2023.11.007
Figure 1. Identification of 115 genes required for photosynthesis

(A) Schematic of biogenesis and regulation of the photosynthesis machinery.

(B) Photosynthetic growth phenotype validation for 1,781 previously identified photosynthesis-deficient Chlamydomonas mutants.25 Photosynthesis-deficient mutants can grow in the dark with acetate but have growth defects in 750 μE/m²/s light without acetate (wild type [WT]). See also Figure S3A.

(C) Normalized colony photosynthetic growth for different mutants (blue) or WT (black). Growth was measured using a metric that incorporates colony size and color (see STAR Methods). Shown is the median of 4 replicates.

(D) Most of the mutant strains have second-site mutations that could cause the photosynthetic phenotype. We used backcrossing to allow segregation between the insertion and second-site mutations. For higher throughput, we developed a pooled backcrossing method (Figure S1).

(E) Histogram of normalized light growth after backcrossing for all strains (black) and for strains disrupted in “genes whose disruption likely did not result in a photosynthesis defect” (GNRPD, STAR Methods, red). Mutants disrupted in 136 genes showed normalized light growth after backcrossing below the threshold of 0.34 (~1.55 on a log2 scale). These genes are linked to the phenotype with FDR < 0.11 (Figure S1).

(F) Validation of the insertion mapping of ~86% of the candidates using PCR and sequencing (see also Figure S2).

(G) Approximately 39% of our hits had a previously known role in photosynthesis (29 in Chlamydomonas and 16 in land plant homologs), compared with 6% in the initial set.

(H) The hits are enriched in chloroplast-predicted proteins (PredAlgo10) and in GreenCut2 green lineage-specific genes.10

(I) Our 115 photosynthetic hits captured most of the previously identified high-confidence hits (41 of 51) and increased the confidence of ~14% of the previously low-confidence hits (32 of 219) (STAR Methods; Figure S1G).
and there are indications that fewer than half of these candidates are actually involved in photosynthesis. Current challenges facing the field include (1) determining which of these candidates are genuinely involved in photosynthesis and (2) determining the functions of validated photosynthesis genes. Here, we address these two challenges by combining genetics and proteomics to identify and functionally characterize genes required for photosynthesis with high confidence on a global scale. We first identified with high confidence (false discovery rate [FDR] < 0.11) a total of 115 genes required for photosynthesis—including 70 genes whose molecular function in photosynthesis had not been previously characterized in any organism—by confirming linkage of each mutation with the observed photosynthetic defect and validating insertion site mappings. We then determined the proteomic profiles of mutants representing these genes to initiate their functional characterization, including assigning 34 of them to specific photosynthetic pathways. As proof of principle for the utility of our resource, we performed additional analyses, which revealed that five of these factors work with known factors to regulate mRNA maturation of key photosystem I (PSI) subunit PsA. We also discovered and characterized two post-transcriptional regulators of photosynthetic apparatus biogenesis, providing insights into how cells leverage the chloroplast translation machinery and the regulation of nuclear gene expression to control photosynthetic complex abundance. Together, our dataset opens the door to rapid characterization of photosynthesis genes and provides systems-level insights into photosynthesis regulation.

RESULTS

A framework for high-confidence identification of genes with roles in photosynthesis

Previous large-scale Chlamydomonas screens suffered from the limitation that most mutant strains carried mutations in multiple genes, preventing high-confidence identification of the specific gene whose disruption causes the observed photosynthetic defect unless multiple independent mutants in the same gene showed the same defect. Here, we overcame this limitation by developing a high-throughput implementation of traditional genetic linkage analysis between a mutation and an observed photosynthetic defect, which allowed us to identify with high confidence the specific gene whose disruption is responsible for the defect, even if that gene was disrupted in only one mutant.

Pooled backcrossing and mapping validation of putative photosynthetic genes

We started this study with a set of 1,781 mapped random-insertion mutants from the CLiP library of Chlamydomonas mutants that we previously identified to have a photosynthetic growth defect. We first validated the mutants’ phenotypes using an automated spot test on agar (Figures 1B and 1C; STAR Methods).

To determine whether a given mapped insertion or another unknown mutation was the cause of the observed photosynthetic defect, we determined whether the insertion was genetically linked to the defect using backcrossing. Backcrossing involves mating a mutant of interest with a wild-type (WT) strain and analyzing the progeny. This process results in random segregation of the different mutations present in the original mutant strain, thereby allowing the impact of each mutation on the phenotype of interest—in our case, defective photosynthetic growth—to be separated. If all progeny carrying a particular insertion exhibited a defect in photosynthetic growth, we concluded that the insertion is genetically linked to the defect, indicating that the disruption of the gene likely caused the defect (Figure 1D).

To overcome the limited throughput (~10 mutants per experiment) of traditional backcrossing, we developed a pooled method that allowed us to backcross nearly 1,000 mutants per experiment (Figure S1A; STAR Methods; Breker et al.26). We backcrossed pools of hundreds of mutants and then grew the pooled progeny under photosynthetic and heterotrophic conditions. We determined the relative abundance of each insertion after growth under each condition by sequencing the unique DNA barcode(s) associated with that insertion (Figure 1E; Table S1; STAR Methods). Depletion of a barcode in the photosynthetic condition pool indicated linkage of the corresponding insertion to the photosynthesis defect.

We sought to estimate the frequency of incorrect identification of causal genes in this approach. Such errors could arise in rare cases where the insertion is not causal but merely in the genomic vicinity of the causal mutation or could be due to measurement noise. We quantified the frequency of such errors with a FDR metric. To calculate the FDR, we used a set of genes whose disruption likely did not result in a photosynthesis defect and measured their prevalence among our hits (Figures 1E and S1B–S1F; STAR Methods). To overcome the limited throughput (~10 mutants per experiment) of traditional backcrossing, we developed a pooled method that allowed us to backcross nearly 1,000 mutants per experiment (Figure S1A; STAR Methods; Breker et al.26). We backcrossed pools of hundreds of mutants and then grew the pooled progeny under photosynthetic and heterotrophic conditions. We determined the relative abundance of each insertion after growth under each condition by sequencing the unique DNA barcode(s) associated with that insertion (Figure 1E; Table S1; STAR Methods). Depletion of a barcode in the photosynthetic condition pool indicated linkage of the corresponding insertion to the photosynthesis defect.

Some of the insertions from the starting collection of 1,781 mutants are known to be mapped to incorrect sites in the genome. Therefore, we validated the mapping of our linked insertions using colony PCR (Figure S2) or whole-genome sequencing (Figures 1F and S2; Table S1; STAR Methods). Altogether, we identified with high fidelity 115 genes required for photosynthesis from our initial set of 1,781 photosynthesis-deficient mutants (Figure S2A).

Approximately 40% of the 115 genes have a known role in photosynthesis in Chlamydomonas (29 genes) or in land plants (16 genes) (Figure 1G), a substantial enrichment compared with ~6% of the genes in the initial 1,781 mutants. The 115 genes are also enriched in metrics associated with photosynthesis: they show a 2.5-fold enrichment in predicted localization to the chloroplast and a 4-fold enrichment in genes conserved specifically in the green lineage (Figure 1H).

A subset of our data provides orthogonal validation of candidate photosynthesis genes. Our 115 genes required for photosynthesis include 41 of the 51 genes identified with high confidence (FDR < 0.3) in previous large-scale photosynthesis screens based on the CLiP mutant collection (Figure 1).
A. WT Photosynthetic phenotype
   Mutant (chl)
   WT allele
   Rescued strain (CHLM)
   Alleviated phenotype

B. RAA6
C. RAA15
D. RAA17
E. CPLD64
F. PIR1
G. CPL6
H. CGL54
I. PSR1
J. CPL12
K. TPK1
L. PSR5
M. PSB27
N. HCF173
O. CHLM-Venus
P. PIR1-Venus
Q. CPL6-Venus
R. CGL54-Venus
S. CPL12-Venus
T. PSB27-Venus

(legend on next page)
This high overlap shows the quality of both datasets. Our 115 genes also include 32 of 219 genes that were previously low-confidence candidates (no FDR was calculated) in the CLIP and Niyogi CAL collections (Figure S1), increasing the confidence that these 32 genes do indeed participate in photosynthesis. Of the remaining 42 genes, 38 had not previously been identified as being required for photosynthesis in any organism.

Altogether, our 115 genes included 70 genes whose molecular function in photosynthesis had not been previously characterized in any organism (Figure S2A). We have noted in Table S1 additional information from other sources that further supports or weakens our confidence in their involvement in photosynthesis. The study of these genes represents a new frontier for photosynthesis research.

**Hit validation and protein localization**

To experimentally validate the involvement of the genes we identified in photosynthesis, we sought to genetically rescue the photosynthetic defect of the mutants that have insertions in genes not previously known to function in photosynthesis. Gene rescue involves testing whether transforming a mutant with a WT copy of the gene alleviates the phenotype (Figure 2A). Despite challenges to gene rescue in *Chlamydomonas* due to difficulties with PCR amplification and expression of heterologous genes,28–30 we rescued mutants in 16 genes out of 36 that we attempted. Considering the low efficiency of construct integration in any organism (Figure S2A), we have noted in Table S1 additional information from other sources that further supports or weakens our confidence in their involvement in photosynthesis. The study of these genes represents a new frontier for photosynthesis research.

**Mutant proteomic profiling informs gene function**

To expand the understanding of the 115 genes identified as required for photosynthesis and to elucidate the specific roles of poorly characterized genes within this set, we sought to use mass-spectrometry proteome profiling (Figure 3A) to assess the impact of the loss of each gene on the proteome. We reasoned that this would be an informative approach to characterize mutants deficient in photosynthesis because the core activities of photosynthesis are mediated by a series of highly expressed protein complexes whose abundance is affected by photosynthetic activity, regulation, and biogenesis. Indeed, many known photosynthesis-deficient mutants show differences in protein complex abundance.32–34 Much of the regulation of the photosynthetic apparatus is thought to occur post-transcriptionally, making protein levels a more informative readout than mRNA.5

When grown in light, our strains exhibit growth defects, which could confound the proteomic readout. To minimize such issues, we grew cells in the dark with acetate as carbon and energy source, taking advantage of the facts that under this condition, growth defects associated with deficient photosynthesis are eliminated, and WT cells assemble a functional photosynthetic apparatus.19

We obtained proteome profiles of mutants each disrupted for one of 100 genes (Figure S2A; Table S3), with at least two experimental replicates for each gene (Figures 3A and 5; STAR Methods). Our profiling dataset captured known co-depletion of proteins that form complexes such as LCIB and LCIC35 (Figure 3B) and known regulatory effects such as the depletion of cytochrome b$.f$ in the tca1 mutant.36 We observed that mutants lacking PHB2 and PDC2 were predicted to be the two subunits of pyruvate dehydrogenase E1; indeed, PHB2 and PDC2 were co-depleted in the *pdc2* mutant (Figure 3D). Another example is CrHCF173, a homolog of the *Arabidopsis* translation initiation factor AtHCF173 that is required for PsbA translation initiation.38 As was shown for AtHCF173, we observed that mutation of CrHCF173 led to the downregulation of PsbA and the entire photosystem II (PSII) complex38,40 (Figure 3E).

The similar behavior of *Chlamydomonas* mutants compared with their land counterparts highlighted the potential of our method to dissect the roles of poorly characterized genes in photosynthesis.

**Figure 2. Gene rescue and protein localization**

(A) Schematic of the genetic rescue procedure for the known chlorophyll biosynthesis gene CHLM. In the dark with acetate, *chlm* grows almost as well as wild type but is yellow18; under high light, the mutant has a severe growth defect. Transformation of the mutant with a Venus-tagged CHLM alleviates both the color and growth phenotypes.

(B–N) The colony growth of wild type, mutants, and the mutants we rescued by transforming with the wild-type genes. (See also STAR Methods and Figure S3B.)

(O) Localization of CHLM-Venus in the wild-type background. A similar localization was observed in the rescued strain.

(P–U) Localizations of Venus-tagged proteins. CPL6, CGL54, and HCF173 are in the mutant background; PIR1, CPL12, and PSB27 are in the wild-type background due to insufficient expression in the rescued mutant strain.
<table>
<thead>
<tr>
<th>Systematic ID</th>
<th>Name</th>
<th>Figures</th>
<th>Localization</th>
<th>Suggested function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cre17.g728850</td>
<td>RAA15</td>
<td>Figures 2C, 4, and 5</td>
<td>predicted mitochondrion</td>
<td>our proteomics and transcriptomics data suggest that this protein participates in splicing of the 2nd intron of psaA mRNA (see main text)</td>
</tr>
<tr>
<td>Cre13.g566400</td>
<td>RAA17</td>
<td>Figures 2D, 4, and 5</td>
<td>predicted mitochondrion</td>
<td>our proteomics and transcriptomics data suggest that RAA17 stabilizes the 3rd exon of psaA (see main text)</td>
</tr>
<tr>
<td>Cre10.g448950</td>
<td>PMR1</td>
<td>Figures 6 and 4</td>
<td>chloroplast, cytosol, nucleus</td>
<td>photosynthesis master regulator 1—our proteomics and metabolomics data suggest that PMR1, a noxim homolog (Figure S8), participates in retrograde regulation and affects mRNA levels of ROGEs (see main text)</td>
</tr>
<tr>
<td>Cre12.g560550</td>
<td>MTF1</td>
<td>Figures 6 and 4</td>
<td>chloroplast</td>
<td>methionyl-tRNA formyltransferase 1—our data indicate that MTF1 is the chloroplast methionyl-tRNA formyltransferase and suggest that it participates in the regulation of the chloroplast-expressed genes (see results and discussion)</td>
</tr>
<tr>
<td>Cre12.g485850</td>
<td>CPLD64</td>
<td>Figures 2E and 4</td>
<td>predicted chloroplast</td>
<td>in our data, the cpld64 mutant showed depletion of the cytochrome b$<em>{6}$f complex; CPLD64 has a predicted transmembrane motif (InterPro: IPR009688); these observations suggest that CPLD64 participates in the biogenesis or stability of the cytochrome b$</em>{6}$f complex in the thylakoid membrane</td>
</tr>
<tr>
<td>Cre01.g014000</td>
<td>PIR1</td>
<td>Figures 2F, 2P, 4, and 5</td>
<td>chloroplast</td>
<td>photosystem I required 1—in our proteomics data, the pir1 mutant showed depletion of PSI; our RNA-seq data suggest that PIR1 does not participate in the psaA mRNA maturation process; it may participate in PsA or PsAB translation</td>
</tr>
<tr>
<td>Cre06.g279500</td>
<td>CPL6</td>
<td>Figures 2J, 2Q, and 4</td>
<td>chloroplast</td>
<td>CPL6 contains a DnaJ heat shock protein domain; we observed that the cpl6 mutant did not exhibit depletion of any photosynthetic complex, suggesting that its chaperone activity is not needed for complex formation in the dark; cpl6 cannot grow under high-light conditions even when supplied with a carbon source (acetate), suggesting that CPL6 may contribute to repairing light damage to the photosynthetic machinery</td>
</tr>
<tr>
<td>Cre02.g073850</td>
<td>CGL54</td>
<td>Figures 2H, 2R, and 4</td>
<td>pyrenoid periphery</td>
<td>CGL54 is in the same protein superfamily as cyanobacterial Psb27 (Figure S9), which is involved in PSII biogenesis; however, a different gene, PSB27, shows higher homology to Psb27 (Figure S9) and the cgl54 mutant did not lead to the depletion of PSII, suggesting that CGL54 has a different function; CGL54 localized to the pyrenoid periphery, similarly to the PSI-interacting protein PSBP4, suggesting that CGL54 may interact with PSI</td>
</tr>
</tbody>
</table>

(Continued on next page)
plant homologs suggests that lessons we learn in Chlamydomonas will also inform our understanding of photosynthesis across the green lineage.

Altogether, ~2,000 proteins were observable in most of the 100 mutant proteomes (Figure S4C; Table S5), providing extensive opportunities for analysis. Here, we focus on the major photosynthetic protein complexes.

### 23 poorly characterized genes impact biogenesis or regulation of individual chloroplast protein complexes

While we observed many cases of mutants that impacted individual components of photosynthetic protein complexes, such as mutants that lack the PSI core subunits PSAE and PSAF (Figure 4A), more than half of our mutants showed proteomic defects in one or more entire complexes (Figures 4B–4I). 41 mutants led to the primary depletion of just one of the eight chloroplast protein complexes we investigated (Figures 4B–4H). These data allowed us to immediately assign roles for 23 poorly characterized genes in the biogenesis or regulation of PSI, cytochrome bo$_{6}$f, PSI, the light-harvesting complexes, or the chloroplast ribosome.

#### PSII

PSII uses light energy to extract electrons from water in the first step of the photosynthetic electron transport chain. In our dataset, mutations in seven genes led to the depletion of the entire PSII complex (Figure 4B). Three of these genes were not previously associated with PSII in any organism. One of the three, PIR1 (Cre-16.g658950), encodes a protein that is predicted to localize to the chloroplast27 and has 6-fold higher transcript levels in light compared to dark,41 so it may participate in the regulation of PSII in response to light.

#### Cytochrome bo$_{6}$f

Cytochrome bo$_{6}$f pumps protons into the thylakoid lumen powered by photosynthetic electron flow. In our dataset, mutation of four genes led to the depletion of the entire cytochrome bo$_{6}$f complex (Figure 4C). Of these four genes, two poorly characterized ones, CPLD64 (Cre12.g485850), which we validated by genetic rescue (Figure 2E; Table 1), and CBR1 (Cre12.g901550), are conserved in land plants (Table S2) and were predicted to localize to the chloroplast.27 Given these observations, we speculate that CPLD64 and CBR1 participate in the biogenesis or stability of the cytochrome bo$_{6}$f complex.

#### PSI

PSI uses light energy to energize electrons, enabling the reduction of NADP to NADPH. In our dataset, mutations in 18 genes led to the depletion of the entire PSI complex (Figure 4D). Twelve of these genes were not previously identified as genes required for photosynthesis, including RAA12, RAA15, RAA17-18, HELS/CPLD46, PIR1, and PIR2, which we describe in detail in later sections. Other interesting poorly characterized genes included RMT2 (Cre12.g524500) and PIR3 (Cre01.g012200). RMT2 was named based on sequence homology to Rubisco large subunit N-methyltransferase (KEGG: K00949); TPK1’s photosynthetic effect is likely due to the participation of TPK1 in the chloroplast pentose phosphate pathway.

#### Light-harvesting complexes

Light-harvesting complexes channel light excitation energy to the photosystems (Figure 4F). In our dataset, mutations in five genes affected the light-harvesting complexes—these genes include LHR1 (Cre02.g142266), whose Arabidopsis homolog CYP97A3 is required for light-harvesting complex II biogenesis,42 and four poorly characterized genes. Two of the poorly characterized genes, LHR4 (Cre01.g016350) and LHR5 (Cre01.g001000), were required for normal levels of light-harvesting complex I; whereas the other two, SRR16 (Cre10.g458350) and LHR2 (Cre14.g616700), affected the LHCBM proteins, the core complex of light-harvesting complex II.

#### Chloroplast ribosome

Mutations in three genes, PSR26 (Cre50.g761497), HEL41 (Cre07.g349300), and PSR8 (Cre02.g110500), led primarily to
the depletion of chloroplast ribosomal proteins (Figure 4H). The helicase HEL41 was previously found to physically associate with the chloroplast ribosomal large subunit6 and in our dataset had a particularly strong effect on the abundance of the large subunit, suggesting that HEL41 directly impacts ribosomal protein levels by contributing to biogenesis or stability of the large ribosomal subunit.

11 poorly characterized genes impact biogenesis or regulation of multiple photosynthetic complexes

Mutations in seven known and eleven poorly characterized genes led to the depletion of multiple complexes (Figure 4I). The known genes illustrate how the depletion of multiple complexes can result from different mechanisms. For example, mutants lacking chlorophyll biosynthesis genes CHLD (Cre05.g242000)25 or CHLM (Cre12.g498550)25 showed a depletion of chlorophyll-binding proteins, including both PSI and PSII complexes (Figure 4I). Other known mutants are in regulatory genes, for example, the kinase CPL3 (Cre03.g185200).25

The poorly characterized genes affecting multiple complexes included the conserved predicted xanthine dehydrogenase/oxidase XDH1 (Cre12.g545101), whose mutation led to decreased levels of PSI and PSII and their light-harvesting complexes similar to mutants in chlorophyll biosynthesis enzymes (e.g., chld and chlm). These observations suggest a role for XDH1 in pigment metabolism, possibly by preventing the activation of chlorophyll degradation by xanthine.45 The poorly characterized genes also included the conserved predicted chloroplast-localized protein MSR8 (Cre09.g400312), whose disruption impacted both PSII and light-harvesting complex II.

Disruption of the poorly characterized genes PMR1 and MTF1 led to the depletion of the entire photosynthetic apparatus; we discuss their characterization below.

Characterization of factors that regulate photosynthetic apparatus biogenesis

We hypothesized that many of the poorly characterized genes encode proteins that regulate the photosynthetic machinery because many (14/24) of the known genes whose disruption led to strong depletion of the photosynthetic complexes in our proteomic experiment encode regulatory proteins (Figures 4B–4I). We focused on two subsets of the poorly characterized genes: ones whose disruption specifically impacted PSI levels and ones whose disruption had broad effects on the photosynthetic apparatus.

Regulators of PSI psaA mRNA maturation

The mRNAs encoding chloroplast-expressed proteins are constitutively expressed, and the abundance of the proteins they encode is primarily regulated post-transcriptionally.46 A central mechanism for this post-transcriptional regulation involves the regulators of organelle gene expression (ROGEs), nuclear-encoded factors that each promote mRNA stability/maturation (M factors) or translation (T factors) of a specific chloroplast-encoded subunit of a photosynthetic complex.46 In the absence of a T or M factor, the abundance of the regulated subunit drops, translation of other subunits decreases, and unassembled subunits are degraded, leading to depletion of the entire complex.46

Figure 3. Proteomic data reproduce known phenotypes and validate predicted phenotypes

(A) In each experiment, ten mutant strains and a wild-type control were grown under dark conditions. After extraction and digestion, we labeled peptides with tandem mass tags (TMTs) and analyzed them using SPS-MS3 mass spectrometry. At least two independent experiments were carried out for each mutant (STAR Methods). The normalized log2 of mutant/WT protein abundance in two replicates is plotted.

(B) LCIB and LCIC protein abundances are shown in the icib mutant.

(C) Cytochrome b6f protein subunit abundances are shown in the tca1 mutant.

(D) Abundance of predicted Chlamydomonas pyruvate dehydrogenase E1 alpha subunit PDC2 and beta subunit PDH2 in the pdh2 mutant.

(E) Abundance of PsbA and other components of the PSII complex in the hcf173 mutant. The bigger dots represent other subunits of the complex of interest. See also Figure S4.
We identified six known M factors among the genes required for accumulating the entire PSI complex in our proteomics (Figure 4D). One of these M factors, MAC1, is required for psaC mRNA stability.43 The other five, RAA1, RAA3, RAA4, RAA6, and RAA8, participate in the maturation of psaA mRNA.48–52

We hypothesized that other genes with similar proteomic patterns might also be M factors. We focused on seven poorly characterized genes (HEL5, RAA17, RAA18, RAA12, RAA15, PIR1, and PIR2), of which we validated three (RAA17, RAA15, and PIR1) by gene rescue (Table 1), whose mutants exhibited strong and specific depletion of the PSI complex (Figures 5A and S6). To determine whether any of these genes are M factors, we profiled the chloroplast transcriptome in mutants representing these genes and known factors (STAR Methods). Mutations in five of the poorly characterized genes, HEL5, RAA17, RAA18, RAA12, and RAA15, resulted in less than 15% of the WT levels of mature psaA mRNA, similar to mutants of known psaA mRNA maturation factors (Figures 5B and 5C), suggesting that these five genes are psaA M factors.

PsA A is one of the two central chloroplast-encoded components of PSI.16 In Chlamydomonas, its maturation involves a sophisticated mRNA splicing mechanism.53 PsA mRNA starts as four separate transcripts that hybridize to form a structure containing two introns, which are spliced out to generate the mature mRNA (Figure 5B). This process is mediated by a ribonucleoprotein complex that includes at least 14 splicing factors.53,54 These splicing factors are classified based on their impact on the splicing of the two introns. By evaluating the relative splicing of each intron in the mutants using paired-end RNA sequencing (RNA-seq), we were able to classify HEL5 as impacting intron 1, RAA15 and RAA18 as impacting intron 2, and RAA12 as impacting both introns (Figure 5D). RAA17 appears to represent a new maturation factor group, which we propose directly affects exon 3 stability (Figures 5B–5E).

HEL5 is required for splicing psaA intron 1
HEL5 (Cre01.g027150) belongs to the DEAD-box helicase superfamily (Interpro: IPR011545). Its Arabidopsis homolog ISE2 appears to be a general splicing factor that participates in the mRNA processing of chloroplast ribosome subunits, ATP synthase subunit AtP, and protease ClpP1,55 While Chlamydomonas HEL5 appears to contribute to the biogenesis or stability of the chloroplast ribosome (Figure 5D), it does not affect the ATP synthase or Clp protease. Instead, we observe that the primary function of HEL5 seems to be the splicing of psaA intron 1 (Figures 5C and 5D), illustrating how the specificity of a splicing factor can change across evolution.

RRA15 and RRA18 are required for splicing psaA intron 2
In mutants lacking RAA15 (Cre17.g728850) or RAA18 (Cre07.g351825), we observed a 96% decrease in mature psaA intron 2 compared with WT, suggesting that these genes encode intron 2 splicing factors (Figures 5A, 5D, and 5E). Transforming the WT allele of RAA15 into the corresponding mutant alleviated the mutant’s growth defects to almost-WT levels (Figure 2C), providing confidence that a mutation in this gene causes the observed photosynthesis phenotype. RAA15 was previously pulled down with known intron 2 splicing factors RAA2 and RAA7,51,52 suggesting that these three factors function together.

RRA12 is required for splicing psaA introns 1 and 2
RRA12 (Cre17.g698750) is a member of the octotricopeptide repeat (OPR) family of regulatory RNA-binding proteins56 required for photosynthesis (Table S1), whose two mutant alleles showed depletion of PSI (Figures 4D and S5A). Its transcriptomic profile was similar to that of RAA1, a known M factor required for psaA intron 1 and 2 splicing18 (Figures 5D and 5E). Much like RAA1, we observed that RAA12 mutation leads to the depletion of mature forms of both introns 1 and 2 (Figure 5D). Furthermore, similarly to RAA1, RAA12 was previously co-precipitated with known M factors: intron 1 splicing factors RAA4 and RAT2,51,57 and intron 2 splicing factor RAA7.56 These results suggest that RAA12 is required for the maturation of both introns.

RRA17 regulates psaA exon 3 stability
Transforming the WT RRA17 (Cre13.g566400) allele into the RRA17 mutant rescues the mutant’s growth to WT-like levels even under high-light conditions (Figure 2D), confirming that RRA17 is required for photosynthesis. The RRA17 mutant exhibits almost-complete depletion of exon 3 (<2% of WT levels), a phenotype not exhibited by any of the other mutants of known factors in our dataset, suggesting that RRA17 is a different kind of maturation factor that specifically protects the third exon. RRA17 is a member of the OPR family of RNA-binding proteins; thus, it is possible that it could directly bind to psaA. The decreased level of exon 3 is likely the cause of the decreased level of the mRNAs with spliced intron 2 observed in the raa17 mutant. RRA17 expression is light dependent: its expression level is 5-fold higher in light compared with dark,11 suggesting that it participates in psaA dark-to-light acclimation.

RAT2 is required for psaA maturation but is not a limiting factor in the dark
RAT2 is a previously known psaA maturation factor that participates in processing the intron 1 RNA component tscA48 (Figures 5B and 5F). As expected, a mutant strain lacking RAT2 showed photosynthetic defects in our screen, but
surprisingly, it did not lead to the depletion of PSI in our protein profiling (Figure 4K). A potential explanation for this discrepancy is that the rat2 mutant has substantially more mature psaA than any other maturation factor mutant in our dataset (Figures 5C–5E). This level of mature psaA mRNA may be sufficient for PSI production in the dark, conditions under which materials were collected for our proteomic analysis. Under light conditions requiring active photosynthesis, the lower levels of psaA mRNA would not meet the higher demand for PSI production, resulting in a photosynthesis defect.

**psaA mRNA maturation**

In addition to identifying and characterizing five M factors, our RNA profiling provides insights into the overall maturation process of psaA. In nearly all mutants that primarily impact one intron (with raa15 being the only exception), we observed that splicing of the other intron is also impacted (Figure 5D), suggesting that each splicing site requires integrity of the other for maximal activity.

HEL5, RAA17, and RAA18 were not identified in the previous immunoprecipitation of the psaA mRNA maturation complex.
### A

<table>
<thead>
<tr>
<th>Relative protein level \log_{(mutant/WT)}</th>
<th>MTMT1 (n=6)</th>
<th>rescued MTMT1 (n=4)</th>
<th>MTMT2 (n=3)</th>
<th>PMR1 (n=4)</th>
<th>rescued PMR1 (n=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSI</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PSII</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ATP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Photosynthetic apparatus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B

<table>
<thead>
<tr>
<th>PS II</th>
<th>MTMT1</th>
<th>rescued MTMT1</th>
<th>MTMT2</th>
<th>PMR1</th>
<th>rescued PMR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PsA</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PsB</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PsC</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ATP synthase</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

### C

**MTMT1-Venus**

Chloroplast - expressed
Nucleus - expressed

### D

**PSI** + PSI ATP

### E

**Relative protein level \log_{(mutant/WT)}**

### F

**mtf1**

**cif2**

**child**

### G

<table>
<thead>
<tr>
<th>PS II</th>
<th>PMR1</th>
<th>rescued PMR1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PsA</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PsB</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PsC</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ATP synthase</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

### H

**PMR1**

### I

**Relative phosphatase activity**

### J

**NADP^+ (F/NADH) peak area ratio**

### K

**WT**

**pmr1**

**mtf1**

### L

**mRNA relative expression \log_{(mutant/WT)}**

### M

**ROGEs**

### N

**Relative protein level \log_{(mutant/WT)} rep 3-4**

### O

**PMR1**

**CIF2**

**MTMT1**

### P

**PMR1-Venus**

**Cytoplasm**

**Chloroplast**

**chlorophyll**

**bright field**

**Pyramid**

(legend on next page)
suggesting that they are only transiently associated or act independently and demonstrating the complementary value of our approach. Together, the above findings broaden our understanding of psaA maturation, a key process in PSI biogenesis and regulation, and illustrate how our data can be used to rapidly functionally characterize factors with roles in photosynthesis.

Functional specialization of chloroplast translation initiation factors

One of the most striking observations from our data was the identification of genes whose mutants exhibited decreased levels of all four major light-reaction complexes. Two of these genes, CIF2 and MTF1, are required for chloroplast translation initiation. CIF2 (Cre07.g341850) likely functions as the chloroplast translation initiation factor 2 (IF2), which attaches the fMet-tRNA to the translation initiation complex, based on its homology to the characterized Arabidopsis IF2, FUG1, and CIF2’s physical interaction with the Chlamydomonas chloroplast ribosome.64

MTF1 is the chloroplast’s MTF and is required for translation of nearly all chloroplast-encoded proteins

MTF1 (Cre12.g560550) is a conserved gene whose mutant shows a severe photosynthetic phenotype. In our proteomic experiments, loss of MTF1 expression had the strongest phenotype: the disruption of this gene resulted in the depletion of the entire photosynthetic apparatus and most of the chloroplast-encoded proteins (Figures 6A, 6B, and S7A). We validated this phenotype by genetic rescue, which alleviated the observed growth defect in the mutant to nearly WT growth under high-light conditions (Figure 6C), and recovered expression of chloroplast-expressed proteins (Figures 6A and 6B).

MTF1 was previously annotated as a putative MTF based on sequence similarity to known enzymes. MTF generate fMet-tRNA, which is the tRNA needed for translation initiation in bacteria.62 In contrast to bacteria, eukaryotes do not use fMet-tRNA for cytosolic translation, but the chloroplast and mitochondria within eukaryotic cells require this tRNA for translation initiation. Indeed, we found that MTF1 has a similar AlphaFold-predicted structure to the known E. coli enzyme MTF, with the active-site key residues and hydrophobic pocket conserved63,64 (Figures S7E and S7F). These similarities validate the annotation of MTF1 as a MTF.

In theory, MTF1 could provide fMet-tRNA for the chloroplast or the mitochondria. We found that Venus-tagged MTF1 localized exclusively to the chloroplast (Figure 6D). The strong effect of mtf1 mutants on chloroplast-expressed proteins and not on mitochondrial-expressed proteins (Figure 6A) also suggest that it is primarily active in the chloroplast. Consistent with the idea that MTF1 primarily affects chloroplast-encoded photosynthetic subunits, we observed that in the mtf1 mutant, chloroplast-expressed subunits tended to be more depleted than their nuclear-expressed counterparts (Figure 6E), suggesting that the depletion of the nuclear-expressed subunits was a secondary effect due to degradation of incompletely assembled complexes. Together, our results strongly suggest that MTF1 is the MTF that mediates chloroplast translation initiation.

Translation initiation factors MTF1 and CIF2 are dispensable for normal levels of several chloroplast-expressed proteins

If all chloroplast-expressed proteins required formylmethionine-tRNA as is thought to be the case for E. coli proteins,65 we would have expected that MTF1 and CIF2 would be required for translation of all chloroplast-expressed proteins. Surprisingly, we found that mtf1 and cif2 mutations did not affect levels of the two chloroplast-expressed proteins required for chlorophyll biosynthesis in the dark, chlB and chlL (Figures 6A, S7A, and S7B). Consistent with this observation, mtf1 and cif2 mutants were green when grown in the dark (Figure 6F), whereas strains without the chlB/L/N complex are yellow in the dark.66 mtf1 and cif2 mutants also did not show downregulation of chloroplast-
expressed RNA polymerase (\(Rpo\) genes, including the essential protein \(rpoA\)) (Figures 6A, S7A, and S7B). These observations suggest that translation of certain subsets of chloroplast-expressed proteins can use non-canonical translation initiation mechanisms that do not require formylmethionine-tRNA.

**PMR1 is a master regulator of photosynthesis**

Our data suggest that the poorly characterized protein PMR1 (Cre10.g448950) coordinates the expression of multiple photosynthetic complexes by acting at the level of nuclear gene expression control. \(pmr1\) mutants showed severe photosynthetic growth deficiency and depletion of light-reaction complexes (most significantly PSI and PSII, and light-harvesting complex I) (Figures 4I, 6A, 6G, and 6N). These defects were all rescued by transforming the mutant strain with the WT allele (Figures 6A, 6G–6H, and S7A). Consistent with a regulatory role of PMR1, expression of the WT PMR1 allele under the strong PSAD promoter in the rescued strain led to the overexpression of most of the photosynthetic complexes (Figures 6A, 6G, and S7H).

PMR1 is a member of the CCR4-NOT family and shows the highest sequence homology (Table S6) and a similar predicted structure (Figure S7I) to nocturnin (NOCT) (Kyoto encyclopedia of genes and genomes [KEGG]; K18764), a protein that has been identified as a circadian-controlled master regulator that affects metabolism and hundreds of transcripts in animals. 68–70 Consistent with NOCT-like characteristics, we observed that PMR1 has periodic expression (Figure S7J), and the disruption of its expression influences the levels of hundreds of mRNAs (Figure S7K).

Recent work showed that human and fly NOCT act as phosphatases that convert NADP(H) to NAD(H),72 which then has secondary effects on the transcriptome. We sought to determine using an \(in\) \(vitro\) assay whether PMR1 also acts as an NADP(H) phosphatase but observed only very minor activity (50-fold lower than NOCT) (Figures 6I and S7L). We further analyzed the predicted binding pocket for the adenine in NADP\(^+\) in PMR1 by structural predication, as compared with that of NOCT, and identified two residues that are different in PMR1: K192 and K377 in PMR1, corresponding to R290 and S369 in NOCT. K192 could disrupt the binding of NADP(H) and K377 may partially block the binding pocket, decreasing enzymatic activity on NADP\(^+\) (Figures 6J and S7M). Finally, an NADP(H) phosphatase mutant would be expected to show an increase in the ratio of NADP(H) to NAD(H),72 but the \(pmr1\) mutant did not show an increase in this ratio (Figure 6K). Together, these results suggest that PMR1’s primary activity is not as an NADP(H) phosphatase; instead, PMR1 may directly regulate mRNA levels, similar to the rest of the characterized members of the CCR4-NOT family.73,74

Our RNA-seq analysis suggests that PMR1 regulates the levels of photosynthetic complexes through broad control of the ROGEs, nuclear-encoded factors that each regulate the levels of photosynthetic complexes (Figure 6L). Instead, the \(pmr1\) mutant exhibited strong depletion of \(~\)20 ROGEs that together regulate all major photosynthetic complexes, most notably ROGEs required for biogenesis of PSI and PSII (Figure 6M; \(p < 0.0016\), Mann-Whitney U test comparing the ROGE mRNA distribution to the distribution of all measured mRNAs). Since the depletion of even one ROGE can lead to the depletion of an entire photosynthetic complex, we propose that this downregulation of ROGEs explains the observed broad and specific (Figures 6A, 6N, and S7N) downregulation of all photosynthetic complexes in the \(pmr1\) mutant (Figure 6O).

If PMR1 directly regulates the mRNA of nuclear-expressed genes, we would expect it to localize to the cytosol and/or nucleus. Consistent with this, fluorescently tagged PMR1 localized to the cytosol and nucleoplasm (Figures 6P and 6Q). Intriguingly, a substantial fraction of the protein also localizes to the chloroplast. This additional site of localization suggests the possibility that PMR1 participates in retrograde regulation—signaling from the chloroplast to the nucleus and cytosol to regulate nuclear-expressed genes.75

**DISCUSSION**

In this study, we identified with high confidence (FDR < 0.11) 115 genes required for photosynthesis, including 70 whose functions in photosynthesis had not been previously characterized in any organism. We then showed that mutant proteomes provide key insights into the functions of these genes in photosynthesis, in many cases allowing the assignment of genes to specific pathways.

We identified five ROGEs that are essential for the biogenesis of PSI. Including these genes, 76% (16/21) of genes with known functions in our dataset that lead to the depletion of an entire photosystem complex are ROGEs (Figure 4), demonstrating their significant impact on photosynthesis.

Growing evidence indicates that ROGEs play a regulatory role rather than being merely required for complex biogenesis:72 different ROGEs affect different chloroplast-encoded genes,5 are differentially transcriptionally regulated,76 and participate in feedback loops,6,76 a classical transcription network motif.77 Moreover, several ROGEs can coregulate the same protein56,76 (Table S4), and the expression of photosystem proteins with a stronger effect on growth, including the largest subunit of each complex, tends to be impacted by more ROGEs (Table S4). Our results further support a regulatory role for ROGEs by showing that different ROGEs can be limiting factors in different conditions; RAT2 is a limiting factor for \(psaA\) expression in the light but not in the dark (Figures 4K and 5C–5E), and by discovering that multiple ROGEs are controlled by a master regulator (Figure 6O). Together, ROGE-mediated regulation raises the intriguing possibility that during the endosymbiosis process, as transcriptional regulation in the chloroplast was lost,5 ROGEs evolved to replace transcription factors in a regulatory network for chloroplast-expressed proteins.

In order to respond effectively to changing conditions, the cell must simultaneously regulate multiple photosynthetic complexes. Such coordinated regulation cannot be achieved by the ROGEs alone, since each regulates only one or two chloroplast-encoded proteins.5 Our results suggest the existence of two mechanisms that operate on a larger scale to coordinate the expression of multiple complexes.
First, the cell appears to leverage the chloroplast translation machinery to coregulate multiple complexes. Specifically, while translation factors MTF1 and CIF2 may look like classical housekeeping genes, our data suggest that they are leveraged for regulatory functions. Whereas classical housekeeping translation initiation factors mediate all translation,\textsuperscript{76} MTF1 and CIF2 each affect specific subsets of chloroplast-expressed proteins, a property associated with regulatory factors.\textsuperscript{75} CIF2 is mostly required for expression of photosynthetic machinery, whereas MTF1 loss also affects ribosomal large subunits (Figures 6A and S7A–S7C). Consistent with a regulatory role, MTF1 overexpression leads to overexpression of proteins downregulated in the mtf1 mutant (Figure 6A). The differences in the proteomic impacts of mtf1 and cif2, combined with the differential regulation of the MTF1 and CIF2 transcripts (Figure S7G), suggest that MTF1 and CIF2 coordinate chloroplast gene expression in response to light and nitrogen availability.

Second, our data suggest that the master regulator PMR1 regulates the mRNA levels of multiple nuclear-encoded ROGEs, thus coordinating the expression of the overall photosynthetic apparatus. We hypothesize that the higher-level regulatory mechanisms mediated by PMR1, MTF1, and CIF2 are essential for the cell’s rapid and coordinated response to changes in growth conditions.

More than 65% of the 115 genes we identified as required for photosynthesis have homologs in land plants (Figure S1H). In most cases, the functions of these conserved genes appear to be similar in \textit{Chlamydomonas} and land plants, supporting the value of \textit{Chlamydomonas} as a model system and expanding the significance of our findings. Genes with no clear homologs in land plants could reflect homolog search failure due to sequence divergence\textsuperscript{80,81} and/or different evolutionary innovations in the algal lineage such as the algal-specific CO\textsubscript{2}-concentrating mechanism (CCM), the study of which has the potential to enhance crop yields.\textsuperscript{82} We anticipate that future studies of the genes identified here and explored in our proteomics dataset will enable further discoveries in photosynthesis.

Limitations of the study

Considering our FDR cutoff of 0.11, up to 11% of our hits may be false positives. We have validated by genetic rescue 12 of the 70 genes not previously known to be required for photosynthesis; future work on other genes will require independent validation. In addition, although protein localization by Venus-tagging is generally reliable,\textsuperscript{30,31} increased confidence in the conclusions on cellular localization will require validation by an independent method such as immunofluorescence.\textsuperscript{31} While we have initiated here the characterization of several of the identified genes, additional work is needed to fully characterize the molecular mechanisms by which they and other factors impact photosynthesis.

STAR+METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
  - EXPERIMENTAL MODEL AND SUBJECT DETAILS
    - Strains and culture conditions
  - METHOD DETAILS
    - Automated spot tests
    - Pooled backcrossing
    - Validating insertion sites by PCR
    - Validating insertion sites by DNA sequencing
    - Selection of 115 high-confidence hits
    - Comparison to hits from previous large-scale studies
    - Mutant gene rescue protocol
    - Confocal microscopy
    - Proteomic analysis
    - Sample processing and mass spectrometry
    - Western blotting
    - Chloroplast transcriptome profiling (Chloroplast RNAseq)
    - Nuclear RNAseq
    - Measurement of NADP\textsuperscript{+} and NAD\textsuperscript{+} in wild type and pmr1 mutant (in-vivo)
    - Protein purification
    - Measurement of NADP(H) dephosphorylation activity of PMR1 in vitro
    - Indirect Immunofluorescence Assay
  - QUANTIFICATION AND STATISTICAL ANALYSIS
    - Barcode quantification, normalization, and growth score calculation
    - Mass spectrometry data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.11.007.

ACKNOWLEDGMENTS

We thank Michelle Warren-Williams for media preparation and propagating strains; the Princeton University genomics core facility and its manager Wei Wang for help with sequencing; Princeton University Confocal Microscopy manager Gary Laevsky for instrumentation support; members of the Jonikas laboratory and Felix Willmund for helpful discussions; Alexei V. Korenykh, Jin Du, Michael Estrella, and Kristina Solorio-Kirpichyan for providing NOCT protein; Olivier Vallon, Yana Kazachkova, Silvia Ramundo, Shan He, Alice Lunardon, Jessica H. Hennacy, Moritz T. Meyer, and Eric Franklin for feedback on the manuscript; and Marie Bao, as part of Life Science Editors, for help with editing the manuscript. Figure 3A was created using BioRender.com. The project was funded by the Princeton Catalysis Initiative, U.S. National Institutes of Health grant R35GM128813, U.S. National Foundation grant MCB-1914989, EMBO fellowship ALTF 1006-2017, HFSP fellowship LT000031/2018-L, HHMI/Simons Foundation grant 55108535, and the Lewis-Sigler Scholars Fund. M.C.J. is a Howard Hughes Medical Institute investigator.

AUTHOR CONTRIBUTIONS

We support inclusive, diverse, and equitable conduct of research.

REFERENCES


### STAR METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit anti- PsbA</td>
<td>Agrisera</td>
<td>Cat# AS05 084A; RRID: AB_2172617</td>
</tr>
<tr>
<td>Rabbit anti- PsbC</td>
<td>Agrisera</td>
<td>Cat# AS11 1787</td>
</tr>
<tr>
<td>Rabbit anti- PsA</td>
<td>Agrisera</td>
<td>Cat# AS06 172; RRID: AB_2237771</td>
</tr>
<tr>
<td>Rabbit anti- ATPC</td>
<td>Agrisera</td>
<td>Cat# AS08 312; RRID: AB_2290280</td>
</tr>
<tr>
<td>Rabbit anti- AtpB</td>
<td>Agrisera</td>
<td>Cat# AS05 085; RRID: AB_2258955</td>
</tr>
<tr>
<td>Rabbit anti- α tubulin</td>
<td>Agrisera</td>
<td>Cat# AS10 680; RRID: AB_10748620</td>
</tr>
<tr>
<td>Goat anti-Rabbit IgG(H+L)</td>
<td>Invitrogen</td>
<td>Cat# A11034</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-PMR1</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UltraPure Low-Melting Point Agarose</td>
<td>Invitrogen</td>
<td>Cat# 16500100</td>
</tr>
<tr>
<td>TRIzol Reagent</td>
<td>Invitrogen</td>
<td>Cat# 15596026</td>
</tr>
<tr>
<td>cOmplete, EDTA-free Protease inhibitor</td>
<td>Roche</td>
<td>Cat# 5056489001</td>
</tr>
<tr>
<td>Anti-FLAG M2 Magnetic Beads</td>
<td>Sigma-Aldrich</td>
<td>Cat# M8823</td>
</tr>
<tr>
<td>3×FLAG peptide</td>
<td>Sigma-Aldrich</td>
<td>Cat# F4799</td>
</tr>
<tr>
<td>4×Laemml sample buffer</td>
<td>Bio-Rad</td>
<td>Cat# 1610747</td>
</tr>
<tr>
<td>Guanidine hydrochloride</td>
<td>Sigma-Aldrich</td>
<td>Cat# 369080-1KG</td>
</tr>
<tr>
<td>Cetyltrimethylammonium bromide</td>
<td>Sigma-Aldrich</td>
<td>Cat# 57-09-0</td>
</tr>
<tr>
<td>HEPES</td>
<td>Sigma-Aldrich</td>
<td>Cat# H3375-25G</td>
</tr>
<tr>
<td>NEM</td>
<td>Sigma-Aldrich</td>
<td>Cat# 128-53-0</td>
</tr>
<tr>
<td>DTT</td>
<td>Sigma-Aldrich</td>
<td>Cat# 3483-12-3</td>
</tr>
<tr>
<td>EPPS</td>
<td>Sigma-Aldrich</td>
<td>Cat# 000010</td>
</tr>
<tr>
<td>Methanol</td>
<td>Fisher scientific</td>
<td>Cat# A955-4</td>
</tr>
<tr>
<td>Acetonitrile</td>
<td>Fisher scientific</td>
<td>Cat# A456-4</td>
</tr>
<tr>
<td>Chloroform</td>
<td>VWR</td>
<td>Cat# BDH83626.400</td>
</tr>
<tr>
<td>H2O</td>
<td>VWR</td>
<td>Cat# 87003-652</td>
</tr>
<tr>
<td>Formaldehyde solution 4 %, pH6.9</td>
<td>Sigma-Aldrich</td>
<td>N/A</td>
</tr>
<tr>
<td>Formic acid</td>
<td>Sigma-Aldrich</td>
<td>Cat# F0507</td>
</tr>
<tr>
<td>NH4HCO3</td>
<td>Sigma-Aldrich</td>
<td>Cat# 09830</td>
</tr>
<tr>
<td>MAX efficiency transformation Reagent for Algae</td>
<td>Invitrogen</td>
<td>Cat# A24229</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phusion High-Fidelity DNA polymerase</td>
<td>New England BioLabs</td>
<td>Cat# M0530L</td>
</tr>
<tr>
<td>MinElute Gel Extraction Kit</td>
<td>QiAGEN</td>
<td>Cat# 28606</td>
</tr>
<tr>
<td>Gibson Assembly Master Mix</td>
<td>NEB</td>
<td>Cat# E2611L</td>
</tr>
<tr>
<td>MT10plex Isobaric Label Reagent Set plus TMT11-131C Label Reagent</td>
<td>Thermo</td>
<td>Cat# A34808</td>
</tr>
<tr>
<td>QIAprep Spin Miniprep Kit</td>
<td>QiAGEN</td>
<td>Cat# 27106</td>
</tr>
<tr>
<td><strong>Experimental Models: Organisms/Strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C.reinhardtii: wild-type CC-4453</td>
<td>Chlamydomonas Resource Center</td>
<td>CC-4533 cw15</td>
</tr>
<tr>
<td>C.reinhardtii: wild-type CC-1690</td>
<td>Chlamydomonas Resource Center</td>
<td>CC-1690</td>
</tr>
<tr>
<td>CLIP library mutants in Table S1</td>
<td>Li et al.25; Chlamydomonas Resource Center</td>
<td><a href="https://www.chlamycollection.org/">https://www.chlamycollection.org/</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
### Resource Availability

#### Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Martin C. Jonikas (mjonikas@princeton.edu).

#### Materials availability
All unique/stable reagents generated in this study are available from the lead contact upon request.

#### Data and code availability
- Data have been deposited and are publicly available:
  - The raw RNA and DNA sequencing data are available in NCBI with accession ID SRP441891: [https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP441891&o=acc_s%3Aa](https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP441891&o=acc_s%3Aa)

## Reagents or Resources

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlamydomonas rescued strains listed in &quot;mutant gene rescue protocol&quot; section</td>
<td>This paper, Chlamydomonas Resource</td>
<td><a href="https://www.chlamycollection.org/">https://www.chlamycollection.org/</a></td>
</tr>
<tr>
<td>Oligonucleotides and Recombinant DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pLM005</td>
<td>Mackinder et al.</td>
<td>GenBank: KX077945.1</td>
</tr>
<tr>
<td>pRAM118</td>
<td>Itakura et al.</td>
<td>GenBank: MK357711</td>
</tr>
<tr>
<td>Plasmid constructs generated and listed in &quot;mutant gene rescue protocol&quot; section</td>
<td>This paper, Chlamydomonas Resource Center</td>
<td><a href="https://www.chlamycollection.org/">https://www.chlamycollection.org/</a></td>
</tr>
<tr>
<td>Software and Algorithms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MATLAB</td>
<td>MathWorks</td>
<td>N/A</td>
</tr>
<tr>
<td>Python</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>PyMOL</td>
<td>pymol.org</td>
<td><a href="https://pymol.org/2/">https://pymol.org/2/</a></td>
</tr>
<tr>
<td>Cutadapt 1.18.</td>
<td>cutadapt</td>
<td><a href="https://cutadapt.readthedocs.io/en/v1.18/">https://cutadapt.readthedocs.io/en/v1.18/</a></td>
</tr>
<tr>
<td>Fiji</td>
<td>Schindelin et al.</td>
<td><a href="https://imagej.net/software/fiji/downloads">https://imagej.net/software/fiji/downloads</a></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electroporation Cuvette, 2 mm gap</td>
<td>Bulldog Bio.</td>
<td>Cat# 12358-346</td>
</tr>
<tr>
<td>Ibidi USA μ-Slide 8 well, Glass bottom</td>
<td>Ibidi</td>
<td>Cat# NC0704855</td>
</tr>
<tr>
<td>Poly-L-lysine coated glass slides</td>
<td>Sigma-Aldrich</td>
<td>Cat# P0425</td>
</tr>
<tr>
<td>Kontes Duall #22 homogenizer</td>
<td>Kimble</td>
<td>Cat# KT885450-0022</td>
</tr>
<tr>
<td>Cryomill</td>
<td>Retsch</td>
<td>Part NO. 20.749.0001</td>
</tr>
<tr>
<td>10% Mini-PROTEAN® TGX™ Precast Protein Gels</td>
<td>BIO-RAD</td>
<td>Cat# 4561036</td>
</tr>
<tr>
<td>Lumigrow Lumibar lights</td>
<td>Lumigrow Lumibar</td>
<td>Cat# 8100-5502</td>
</tr>
<tr>
<td>Vacuum filter flask, with a fritted glass support base.</td>
<td>Wilmad Labglass</td>
<td>Cat# BP-1752-001</td>
</tr>
<tr>
<td>Nylon membrane filters (0.5 μm pore size)</td>
<td>GVS Magna™</td>
<td>Cat# 1213776</td>
</tr>
<tr>
<td>Oasis HLB 96-well μElution Plate, 2 mg Sorbent per Well, 30 μm, 1/pk</td>
<td>Waters</td>
<td>Cat# 186001828BA</td>
</tr>
<tr>
<td>Electroporator</td>
<td>NEPA GENE</td>
<td>NEPA21 type II</td>
</tr>
<tr>
<td>SPS Confocal Microscope</td>
<td>Leica</td>
<td>TCS SP5</td>
</tr>
<tr>
<td>Singer Rotor HAD</td>
<td>Singer Instruments</td>
<td>Cat# ROT-001</td>
</tr>
<tr>
<td>PhenoBooth imager</td>
<td>Singer Instruments</td>
<td>N/A</td>
</tr>
<tr>
<td>Nikon Confocal laser scanning Microscope</td>
<td>Nikon</td>
<td>A1Rsi</td>
</tr>
<tr>
<td>Typhoon FLA9500 fluorescence scanner</td>
<td>GE Healthcare</td>
<td>N/A</td>
</tr>
<tr>
<td>iBright imaging system</td>
<td>invitrogen</td>
<td>iBright 1500</td>
</tr>
</tbody>
</table>
The raw proteomic data are available in ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD036908: http://www.ebi.ac.uk/pride/archive/projects/PXD036908.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains and culture conditions
We performed all experiments on Tris Acetate Phosphate (TAP) TAP or Tris Phosphate (TP) media with revised trace elements. TP media had the same recipe as TAP, but the acetic acid was omitted and HCl was added instead to adjust the pH to 7.5. We propagated strains robotically on TAP agar as previously described. All mutants used in this study were from the C. PhotosyntheLiP library. We used the library’s parental strain, CC-4533, as wild type. We backcrossed mutants to a CC-1690 mt+ transformant carrying a hygromycin resistance cassette (WT-hyg), which has high mating efficiency with the CLiP strains.

We performed spot tests and backcrossing with a subset of 1,781 out of the 3,109 mutants deficient in photosynthetic growth identified previously. This subset had been propagated in the laboratory as colony arrays in 96-colony format since the library’s original construction; whereas propagation of the remaining strains had stopped by the time this study began.

We focused our efforts on characterizing insertions with mapping confidence levels of 1-3. The 1,781 mutants carried insertions into 1,616 genes mapped with confidence levels 1-3.

METHOD DETAILS

Automated spot tests
We used a RoToR robot (Singer) to replicate colony arrays in 384-colony format from the TAP agar plates on which the 1,781 mutants were propagated onto three agar plates: one TAP, and two TP. We grew the TAP plate in the dark for about a week before imaging; and we acclimated the two TP plates overnight at 750 μE/m²/s for 2-3 days before imaging (using Lumigrow Lumibar lights, catalog number 8100-5502; equal levels of red, blue, and white light). We photographed the plates using a PhenoBooth imager (Singer). We performed the experiment in four replicates: two independent experiments with a technical replicate in each experiment.

To calculate the “normalized colony photosynthetic growth” we analyzed the pictures using MATLAB. We selected parameters and the algorithm to match as closely as possible our observations by eye. We used a MATLAB script to identify and remove the background and to calculate colony size, which we determined based on the number of green pixels. We further added a 0.5-1 adjustment based on how dark the green pixels are, because when colonies are more dense, they become a darker green. We limited the effect of the color to 0.5-1 to put more emphasis on the colony’s actual size, which we felt more closely reflects the colony’s growth. This normalization is done automatically using MATLAB based on the color levels, and all the values are relative. We normalized the colony size in each plate by the median size of the 10 largest colonies. We then normalized the size of each colony on the high light plates by the size of the corresponding colony on the corresponding TAP dark plate. We performed the second normalization to rule out mutants with a slow growth phenotype that is not specific to photosynthesis.

Pooled backcrossing
We performed initial backcrossing experiments with two subsets of mutants labeled MK (26 plates) and AB (10 plates), which together contained the 1,781 mutants, with some mutants being present in both subsets. After obtaining initial results with these subsets, we re-arrayed the most promising mutants in 96-colony format onto four plates labeled NP. The NP plates included 1) mutants containing insertions linked to photosynthetic defects in the initial backcrosses, 2) insertions in genes that were identified as high-confidence hits in our previous study, and 3) mutants that were yellow or brown. Additionally, to check the method’s replicability, we generated a control plate which contained mutations in genes that were not hits and carried insertions whose disruption likely did not result in a photosynthesis defect. The genes disrupted in mutants on the control plate included 1) known flagellar genes and 2) genes that were represented by more than 35 barcodes, no more than 2 of which were hits in our original pooled photosynthesis screen (in other words, many mutants were available for these genes and the vast majority of these mutants did not exhibit a photosynthesis defect). Using the NP and control plates, we performed a final backcrossing experiment that included two biological repeats of the NP plates and one biological repeat of the control plate.

The backcrossing approach was adapted from the pooled mating (Multiplexed Bulked-Segregant Pool) protocol described previously. Our protocol is illustrated in Figure S1. Each experimental replicate consisted of the following steps:

1) Mating: We scraped and pooled mt- mutant strains from 96-colony format arrays into flasks containing low-nitrogen gamete-induction medium. We pooled 60-150 colonies into each 250 ml flask containing 50 ml of gamete-induction medium. We
resuspended a similar quantity of WT-hyg into separate flasks containing the same media. We used a cell counter to verify that the strains and the WT-hyg cells were at a similar concentration. We shook flasks at 90 RPM for 5-7h in low light (∼40 μE) for mating induction. Then for each flask of mutant strains, we mixed 700ul of mutant strains (mt-) and 700ul of WT-hyg in a 1.5 ml Eppendorf tube, incubated them at low light (∼40 μE) without shaking for one hour, then gently spread them on two TAP agar plates. We incubated the plates overnight in very low light (∼30 μE). In the morning, we wrapped the plates in aluminum foil and kept them in the dark for 7 days.

2) Meiosis: We removed most of the unmated cells by scraping the agar surface using a sharp razor, and moved the plates to low light (∼30 μE) for meiosis induction and initial proliferation for ∼5 days. We used a light microscope to check the sporulation efficiency.24 We pooled the strains into liquid media (TP) for competitive growth.

3) Light and cassette selections (competitive growth): We added hygromycin to our media to ensure that only backcrossed strains were measured. The mutant library does not have hygromycin resistance, so the original CLiP mutants cannot grow on this media. The WT-hyg strain has hygromycin resistance but does not have barcodes, so it will not affect the barcode counting. We inoculated pooled strains at ∼2 × 10⁴ cells ml⁻¹ into TAP + hygromycin (15 μg/ml) 1L bottles for dark growth (3 replicates) and TP + hygromycin (15 μg/ml) 1L bottles for high light growth (3 replicates; except of the 1st experiment where we also included hygromycin (15 μg/ml) + paromomycin (5 μg/ml) conditions). We bubbled air into the bottles and stirred them using magnetic stirrers at 200 rpm. We exposed the TP cultures to 100 μE for overnight light acclimation, then to 750 μE for the remainder of the growth (using Lumigrow Lumbar lights, catalog number 8100-5502; equal levels of red, blue, and white light). When the cells reached a concentration of approximately 2 × 10⁶ cells ml⁻¹, we harvested 10⁸ cells for DNA extraction by centrifugation and flash-freezing the pellet in liquid nitrogen.

Next, we extracted the DNA and prepared the barcode libraries as described,24 and sent the libraries for Illumina sequencing at the Princeton Genomics Core Facility.

After demultiplexing, the barcodes where quantified, normalized, and used to calculate the growth score as described in “barcode quantification, normalization, and growth score calculation” in the “quantification and statistical analysis” section below.

Validating insertion sites by PCR
We adapted the check PCR protocol from the CLIP website (https://www.chlamylibrary.org/about), where we used the G1 and G2 primers to validate the existence of the expected insertion (Figure S2). We used the primers suggested for each strain on the CLIP website. We considered the mapping validated if we got a larger PCR product for the mutant than for the wild type, or if we obtained a PCR product for the wild type and not for the mutant in at least 2 experiments (Figure S2).

Validating insertion sites by DNA sequencing
The strains were grown in the dark condition, and the DNA was extracted using the same method as above. The DNAs were sent to Princeton Genomics Core Facility for library preparation and whole genome sequencing.

The paired-end 150nt reads were aligned to a reference file that combined the v5.5 Chlamydomonas genome (from Phytozome), the chloroplast and mitochondrial genomes (from NCBI: chloroplast_BK000554.2.gb and mitochondrion_U03843.1.gb) and our CIB1

<table>
<thead>
<tr>
<th>Name</th>
<th>Plates in backcrossing</th>
<th>Competition experiments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp1 (AB replicate 1)</td>
<td>AB set (10 plates)</td>
<td>2 TAP Hygromycin dark and 2 TP Hygromycin light; 1 TAP hygromycin + paromomycin dark and 2 TP hygromycin + paromomycin light</td>
</tr>
<tr>
<td>Exp2A (MK 1-12 replicate 1)</td>
<td>1st half of MK set (12 plates)</td>
<td>3 TAP hygromycin dark and 3 TP hygromycin light</td>
</tr>
<tr>
<td>Exp2B (MK 13-26 replicate 1)</td>
<td>2nd half of MK set (14 plates)</td>
<td>3 TAP hygromycin dark and 3 TP hygromycin light</td>
</tr>
<tr>
<td>Exp3A (MK 1-12 replicate 2)</td>
<td>1st half of MK set (12 plates)</td>
<td>3 TAP hygromycin dark and 3 TP hygromycin light</td>
</tr>
<tr>
<td>Exp3B (MK 13-26 replicate 2)</td>
<td>2nd half of MK set (14 plates)</td>
<td>3 TAP Hygromycin dark and 3 TP Hygromycin light</td>
</tr>
<tr>
<td>Exp4 (AB replicate 2)</td>
<td>AB set (10 plates) + 3 plates from MK set.</td>
<td>3 TAP hygromycin dark and 3 TP hygromycin light</td>
</tr>
<tr>
<td>NP plates</td>
<td>2 biological replicates of NP set (4 plates) + 1 biological repeats of control set (1 plate).</td>
<td>For each biological replicate: 3 TAP hygromycin dark and 3 TP hygromycin light</td>
</tr>
</tbody>
</table>
cassette,\textsuperscript{25} using the command “bowtie2 –sensitive-local -k 10 -l 100 -X 650 -S”. The resulting SAM files were filtered to extract only read pairs indicating insertion junctions (where the primary alignment was discordant with one side aligning to the CIB1 cassette and the other side aligning to the genome). The resulting genomic positions corresponding to likely cassette insertion positions were clustered (using scipy.cluster.hierarchy.fclusterdata(d=3000, criterion=’distance’, method=’average’)). For each mutant, all clusters containing 4 or more reads were plotted to show the detailed read locations and orientations, as well as the putative insertion positions according to the original library data.\textsuperscript{25}

Additionally, for each such plot, the concordant read pairs spanning each genomic position were counted and plotted. The resulting plots were evaluated manually to determine the most likely insertion position(s), based on the numbers of matching reads, whether the reads originated from both sides of the insertion position (much less likely for junk fragments), and whether there were concordant read pairs spanning the position (real cassette insertions should not have concordant read pairs spanning them, since the cassette is much longer than the sequenced fragment size).

Selection of 115 high-confidence hits

In our experiment, 148 mutants in 136 genes showed normalized light growth after backcrossing that fell below the 0.34 threshold (FDR = 0.1).

First, we validated that the insertions were mapped to the correct genes. We validated the mapping for 117/136 of those genes (86\%) by PCR and DNA sequencing (Figures 1F and S2; Table S1). The 19 unvalidated genes were removed from the list.

Next, we removed some of the hits to improve the quality of the data set as described below:

1) Six genes (Cre06.g262900, Cre03.g158950, Cre12.g521450, Cre13.g578600, Cre17.g728700, Cre02.g106950) were represented by only one mutation that was in a strain that also included a mutation in an established photosynthetic gene or in a gene with multiple hits in our data set. In these cases, we assumed that the phenotype originated from the well-established gene and removed the 2nd gene from the hit list.

2) Five strains had two hits in each (LMJ.RY0402.172741: Cre13.g584250 + Cre12.g554400, LMJ.RY0402.187220: Cre11.g481115 + Cre07.g326010, LMJ.RY0402.210483: Cre10.g458700 + Cre03.g211185, LMJ.RY0402.166642: Cre03.g155001 + Cre16.g660390 & Cre16.g660430, LMJ.RY0402.176469: Cre06.g296500 & Cre06.g296550 + Cre16.g687294 & Cre16.g687406). While both genes may be required for the photosynthetic growth, it is more probable that one is the real hit and the other is piggybacking on its phenotype. Hence, we counted them as one and concentrated on the one more likely to be connected to photosynthesis (Cre13.g584250, Cre11.g481115, Cre10.g458700, Cre03.g155001, Cre16.g687294). In Table S1, we state the reason for the choice and mention that the effect can be from the other gene.

3) We removed Cre09.g407650 from the gene hits list because we observed in the proteomic data that Cre09.g407650 is not downregulated in the corresponding mutant (Figure S5C). The insertion in that mutant was in the 3’ UTR, consistent with a mild effect on protein levels.

We then added 10 genes as described below:

In our statistical analysis, we looked at genes with insertion mapping confidence levels of 1-3 and excluded confidence level 4 insertions because only 58\% of these mutants are correctly mapped.\textsuperscript{25} However, there were 3 cases where we did validate the insertion of confidence level 4 hits (LMJ.RY0402.124891: Cre16.g665750, LMJ.RY0402.207089: Cre01.g040050, LMJ.RY0402.097626: Cre12.g501550), so we added those three genes to the hit list.

Last, we added 7 genes based on manual analysis of the data (LMJ.RY0402.176891: Cre01.g022681, LMJ.RY0402.119871: Cre06.g273700, LMJ.RY0402.091258: Cre09.g415500, LMJ.RY0402.174216: Cre09.g415700, LMJ.RY0402.049481: Cre02.g901750, LMJ.RY0402.049829: Cre11.g467573, LMJ.RY0402.208107: Cre16.g668700). In most of these cases, the gene was not a hit in the original analysis because it was not a hit in one replicate, but the replicate is not reliable due to an obvious reason such as very low reads. After removing a problematic experiment, the gene is a hit. In Table S1, we mention in each of these cases why the gene was included in the hit list.

After these edits, our list contained 115 high-confidence genes.

Comparison to hits from previous large-scale studies

We compared our 155 high-confidence genes to two sets of hits: 1) previously-identified high-confidence hits, and 2) previously low-confidence hits; which we obtained from three previous large-scale studies.\textsuperscript{21,24,25}

Previously-identified high-confidence hits consisted of high-confidence hits from Li et al.\textsuperscript{25} and genes in the photosynthesis clusters in Fauser et al.\textsuperscript{24} Fauser et al. clustered mutants together based on their phenotype in over 100 different conditions. The work identified two clusters of genes relevant to photosynthesis. The first cluster is the light-sensitive group, where all the hits are relevant to our study; the second cluster is the photoautotrophic light-insensitive. In this second cluster, the clustering is based on phenotypes across many conditions; however, the only condition similar to our experiments is Photoautotrophic 1–3, so we took only the genes whose mutants exhibited pronounced phenotype in this condition: Cre14.g616600, Cre01.g016514, Cre03.g194200, Cre03.g187800, Cre10.g423500, Cre06.g259100, Cre11.g467712. We merged the hits from Li and Fauser. This procedure yielded 51 high-confidence hits, of which 41 were also high-confidence hits in our study.
Previously low-confidence hits consisted of a subset of the 260 low-confidence hits from Li et al.28 and the 253 low-confidence hits from Fauser et al.24 that were represented in the collection of mutants we analyzed. Neither data set had FDR calculations. While both datasets include genes truly required for photosynthesis, methodological limitations of the studies mean that these datasets also include a substantial number of false positives, making validation by our orthogonal method valuable. In low-confidence hits from Li et al., many of the genes are represented by only one mutant, and others are represented by several mutants but only a small fraction of these mutants shows a photosynthetic phenotype. So, there is a high chance that the photosynthetic phenotype comes from a second-site mutation. In the Wakao study, the authors showed that in most cases their insertion is linked to the photosynthetic phenotype; however, their insertions typically were associated with large deletions that affected several genes. Wakao et al. chose to assign the phenotype to one of the disrupted genes in each of the mutants, primarily based on the literature. Although this connection is often correct, it does not have an experimental/statistical basis.

To create the low-confidence data sets, we first merged the Li and Wakao datasets with 260 and 253 hits respectively. We then took the subset of this merged list of genes that overlaps with the ~1,161 genes that were included in our initial data set. If a gene was also in the previously-identified high-confidence hits, it was removed from this list. This procedure yielded 219 previously low-confidence hits, of which 32 were high-confidence hits in our study.

Mutant gene rescue protocol

The plasmids for complementation were generated as described previously.31 Of the 16 plasmids were based on the pLpM005 backbone, and the remaining 12 were based on the pRAM118 plasmid where the paromomycin resistance cassette was replaced with a hygromycin resistance cassette.86 All plasmids expressed the gene of interest from a PSAD promoter and appended a Venus-3xFLAG tag to the protein sequence.

In the gene rescue protocol, we transformed mutant cells with the linearized plasmid expressing the gene disrupted in the mutant. The linearization and transformation process was carried out as previously, until the selection, which was carried out as follows. For plasmids with hygromycin resistance cassette, we used hygromycin-based selection. The cells were plated on 1.5% agar TAP plates with hygromycin (20 μg/ml) and paromomycin (μg/ml) and placed under very dim light for five days, then transferred to light (~100 μE) for 1–2 weeks until colonies of a sufficient size for picking appeared. For plasmids with paromomycin resistance cassette, we could not use drug selection because CLiP strains already have paromomycin resistance, so we used light selection instead. This selection could be used only for mutants that grow poorly under light conditions. For each of these strains, we included a control where we transformed the mutant with a different plasmid of similar size to determine if transformation with any plasmid could reverse the phenotype, e.g., by creating a second-site suppressor mutation. We only considered a rescue successful when the transformation of the correct gene led to growth under light conditions and the control transformation did not. We plated the cells on 1.5% agar TP plates with paromomycin (20 μg/ml). We gradually increased the light intensity to allow for light acclimation. We left the plate on the shelf overnight for five days under 30 μE, three days under ~100 μE, and finally 3-4 days under ~600-700 μE light.

Next, we validated the rescues by robotic spot tests. After the rescue, we picked ~40 transformants from each rescued mutant to check their photosynthetic phenotype. We used RoToR robot (Singer) to replicate each plate with transformants, wild type and mutants to TP and TAP plates, in order to check their growth under TP highlight (550-1100μE) compared to their growth under TAP dark conditions. Then we took 2-4 promising colonies (3 replicates for each) into the plate with wild type and the original mutants (RP 1-4 plates). We used those plates to validate our rescued phenotype. We have at least two independent experiments for each RP plate.

Gene rescue is notoriously challenging in Chlamydomonas due to difficulties with PCR amplification and expression of heterologous genes.28–30 so we performed this part as a “screen”. We used plasmids with the 36 genes we managed to clone (Cre01.g014000, Cre01.g015500, Cre01.g016350, Cre01.g02681, Cre01.g040050, Cre02.g073850, Cre02.g106950, Cre02.g142266, Cre06.g158950, Cre03.g188700, Cre05.g243800, Cre05.g248600, Cre06.g258566, Cre06.g262900, Cre06.g279500, Cre07.g350700, Cre09.g396920, Cre10.g420561, Cre10.g433400, Cre10.g448950, Cre10.g466500, Cre11.g467682, Cre12.g485850, Cre12.g498550, Cre12.g521450, Cre12.g566400, Cre13.g578650, Cre13.g584250, Cre13.g608000, Cre16.g658950, Cre16.g675246, Cre15.g782850, Cre15.g860550, Cre09.g96250, Cre16.g672949), to try to rescue its mutant strain once, and continued with the strains that we managed to rescue. Our success rate of ~44% is close to the maximum expected even if all were real hits, considering that only 30%–50% of transformed constructs express in medium-throughput efforts.31 Many of the failed rescues are likely due to challenges with transformation into Chlamydomonas,28–31 detrimental effects of the GFP tag or the constitutive promoter with some of the genes, and the inherent limitations of our approach, including that rescue of each mutant was only attempted once.

The rescued mutants generated in this study are listed below:

1) Rescued CHLM was generated by transforming plasmid A134 (hygromycin resistance) into the CLiP mutant LMJ.RY0402. 228123, which carries a disruption in the Cre12.g498550 gene.
2) Rescued PSR1 was generated by transforming plasmid pRAM+49;50 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.077016, which carries a disruption in the Cre10.g433400 gene.
3) Rescued CPL6 was generated by transforming plasmid A249 (hygromycin resistance) into the CLiP mutant LMJ.RY0402. 046095, which carries a disruption in the Cre06.g279500 gene.
4) Rescued CPL12 was generated by transforming plasmid A253-2 (hygromycin resistance) into the CLiP mutant LMJ.RY0402. 180319, which carries a disruption in the Cre10.g466500 gene.
5) Rescued CGL54 was generated by transforming hygromycin resistance plasmid (N/A) into the CLiP mutant LMJ.RY0402.057931, which carries a disruption in the Cre02.g073850 gene.
6) Rescued TPK1 was generated by transforming plasmid pRAM+77;78 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.207089, which carries a disruption in the Cre01.g040050 gene.
7) Rescued PSR5 was generated by transforming plasmid pRAM+69;70 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.176891, which carries a disruption in the Cre01.g022681 gene.
8) Rescued CPLD64 was generated by transforming plasmid A258 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.234057, which carries a disruption in the Cre12.g485850 gene.
9) Rescued TBA2 was generated by transforming plasmid pRAM+103;104 hygromycin resistance) into the CLiP mutant LMJ.RY0402.164167, which carries a disruption in the Cre13.g578650 gene.
10) Rescued PIR1 was generated by transforming plasmid A202 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.044496, which carries a disruption in the Cre01.g014000 gene.
11) Rescued PMR1 was generated by transforming plasmid B451 (hygromycin resistance) into the CLiP mutant LMJ.RY0402.248779, which carries a disruption in the Cre10.g448950 gene.
12) Rescued MTF1 was generated by transforming plasmid M1A (hygromycin resistance) into the CLiP mutant LMJ.RY0402.193706, which carries a disruption in the Cre12.g560550 gene.
13) Rescued PBS27 was generated by transforming plasmid Y7 (paromomycin resistance) into the CLiP mutant LMJ.RY0402.255772, which carries a disruption in the Cre05.g243800 gene.
14) Rescued RAA6 was generated by transforming plasmid T675 (paromomycin resistance) into the CLiP mutant LMJ.RY0402.208103, which carries a disruption in the Cre07.g350700 gene.
15) Rescued RAA5 was generated by transforming plasmid T666 (paromomycin resistance) into the CLiP mutant LMJ.RY0402.254076, which carries a disruption in the Cre17.g728850 gene.
16) Rescued RAA17 was generated by transforming plasmid J6/T791 (paromomycin resistance) into the CLiP mutant LMJ.RY0402.133008, which carries a disruption in the Cre13.g556400 gene.

Confocal microscopy
We performed confocal imaging as described previously. Colonies were transferred to a 96-well microtiter plate with 100 μL TP liquid medium in each well and then pre-cultured in air under 150 μmol photons m⁻² s⁻¹ on an orbital shaker. After ~16 hr of growth, 10 μL cells were transferred onto a μ-Slide 8-well glass-bottom plate (Ibidi) and 200 μL of 1% TP low-melting-point agarose at ~35 °C was overlaid to restrict cell movement. Cell samples were imaged using a Leica SP5 confocal microscope with the following settings: Venus, 514 nm excitation with 530/10 nm emission; and chlorophyll, 514 nm excitation with 685/40 nm emission. All confocal microscopy images were analyzed using Fiji. For each strain, a confocal section through a cell showing the predominant localization pattern was captured and analyzed.

Proteomic analysis
Based on our screen results we chose mutants in 100 genes for proteomic profiling (Figure S2A; Table S3). The list includes 3 poorly-characterized genes that were not in the final hits but are hits in other data sets: PSR23 and PIIR2 are high confidence genes in Li et al., and PSR24 is a hit in 2 hit lists: low confidence in Li et al. and in Wakao et al. We grew starter cultures in TAP dark for about a week, then moved them to ~700 ml of TAP (initial concentration ~10⁶ per ml) in 1L bottles and continued growth in the dark. We bubbled air into the bottles and stirred them (using a magnetic stirrer) set to 200 RPM until they reached ~2x10⁶ cells ml⁻¹. We pelleted ~5x10⁷ cells in 50 ml falcons, transferred the pellets to 1.5 ml tubes, pelleted them again, froze them on dry ice, and stored them at -80 °C.

For each proteomic 11-plex, we prepared 10 samples + a wild-type control. The wild-type control was used in most 11-plexes except for PSR24, which carries a disruption in the Cre13.g556400 gene. The proteins we focused on, in most cases, are quantified from multiple peptides (e.g., the number of quantified peptides for PsaA is 8-9, PsaF is 4-8, PetA is 8-14, PetC is 3-5, PsaB is 12-20, PsaC is 9-15, AtpA is 14-18, and AtpB is 17-21) and thus, these proteins are more likely to be quantified accurately.

1) We measured the overall protein abundance in each sample before we mixed them into the 11-plex to reduce the chance that one sample will dominate the 11-plex.
2) We focused on groups of proteins (like PSI proteins or Rubisco complex); the chance that an artifact will impact an entire complex is extremely low.
3) Each peptide is analyzed independently in the mass spectrometer, so proteins quantified from several peptides are much more reliable. The proteins we focused on, in most cases, are quantified from multiple peptides (e.g., the number of quantified peptides for PsaA is 8-9, PsaF is 4-8, PetA is 8-14, PetC is 3-5, PsaB is 12-20, PsaC is 9-15, AtpA is 14-18, and AtpB is 17-21) and thus, these proteins are more likely to be quantified accurately.
4) To reduce the chance that the specific set of mutants in an 11-plex will affect the results, the mutants were selected at random, and the replicate for each mutant was in an 11-plex containing a different set of other mutants.
Sample processing and mass spectrometry
TMT-labeled (11-plex) peptides were prepared mostly as previously described.91 Frozen cell pellets were resuspended in 6 M guanidine hydrochloride (GdCl), 2% cetyltrimethylammonium bromide (CTAB), 50 mM HEPES, 1mM EDTA, and 5mM dithiothreitol (DTT) (pH 7.4). The resuspension lyses the algae to visual homogeneity. Mutant algae cultures grow to different densities and generate pellets of different mass. Diversity in pellet mass was normalized by diluting cells to that of the least dense culture by visual inspection. The final volume ranged from 200-1200 µL. 200 µL of each resuspension was removed to a new Eppendorf prechilled on ice. The lysed algae were sonicated at 20% power for 25 s. Proteins were denatured further at 60 °C for 20 min. After cooling, cysteines were alkylated by the addition of 1 m MeOH. The proteins were resuspended in 50 µL of 6 M GdCl and 10 mM EPPS (3-[4-(2-hydroxyethyl)-1-piperazinyl]propane sulfonic acid) (pH 8.5). The resuspended pellets were frozen.

Pellets were thawed and their protein concentrations quantified using the BCA assay from Pierce with the BSA standard curve diluted in 10 mM EPPS pH 8.5 6M GdCl. 30 µg of each pellet was diluted to 15µL with 10mM EPPS pH 8.5 in 6M GdCl. The 15 µL of 2 µg/µL denatured protein solution was diluted with 75 µL 20 ng/µL LyseC in 10mM EPPS pH8.5, and vortexed to digest overnight at room temperature. A second round of digestion followed with the addition of 270 µL of 20 ng/µL each LyseC and Trypsin in 10 mM EPPS pH 8.5, vortexing and overnight incubation at 37°C. The solvent was removed under reduced pressure in a SpeedVac and resuspended in 30 µL of 200 mM EPPS (pH 8.0) to a concentration of 1 g/L. Ten microliters were removed from each resuspension and charged with 2µL of different TMT-isobaric mass tag N-hydroxysuccinimide (NHS) ester (20 g/liter). The acylation proceeded overnight at RT and was quenched at RT with 0.5 µL of 5% hydroxylamine for 20 min, followed by 1 µL of 5% phosphoric acid.

Peptides were enriched from the acidified TMT labeling reactions by solid-phase extraction using a Waters Oasis HLB Elution 96-well plate (3 mg/well). One well per multiplexed quantitative proteomics experiment was wetted with 400 µL MeOH and then hydrated with 200 µL water. The 11 labeling reactions are pooled and diluted into 400 µL and allowed to adsorb HLB resin under gravity flow. The adsorbed peptides were washed with 100 µL water, followed by centrifugation for 1 min at 180 rpm. The peptides were eluted with sequential additions of 100 µL of 35% acetonitrile (1% formic acid [FA]) and 100 µL of 70% acetonitrile (0.1% FA). Eluent solvent was removed under reduced pressure in a SpeedVac. The peptides were resuspended in 20 µL of 1% FA and subjected to quantitative multiplexed proteomics by nano-ultraperformance liquid chromatography-tandem mass spectrometry (nanoUPLC-MS/MS).

Peptides were separated on a 75 µm inner diameter microcapillary column. The tip for the column was pulled inhouse and the column was packed with approximately 0.5 µm (5µm, 100 Å, Michrom Bioresources) followed by 40 cm of Waters BEH resin (1.7 µm, 120 Å). Separation was achieved by applying a 3-22% Acetonitrile gradient in 0.125%, formic acid with 2% DMSO over 165 min at ~300 nL/min. Electrospray ionization was enabled by applying a voltage of 2.0 kV through an IDEX high-pressure fitting at the inlet of the microcapillary column.

TMT3 data collection was performed as previously described91 on a Fusion Lumos Tribrid Mass Spectrometer (Thermo). The instrument was operated in data-dependent mode (10 ions/scan) with an MS1 survey scan performed at a resolution setting of 120k (m/z 200) with a scan range of m/z 350 to 1,350, an RF (radio frequency) lens of 60%, automatic gain control (AGC) target of 1e6, and a maximum injection time of 100 ms. Ions with charge states 2-6 were filtered by intensity with a threshold of 5e3. A dynamic exclusion window of +/-0.01ppm for 90s was used. MS2 quadrupole isolated ions (0.5 isolation window) were activated with CID at 35% collision energy and Q 0.25 and analyzed in the ion trap with an AGC target of 1.5e4 and 75ms maximum injection time. 10 data dependent MS3 synchronous precursor selections (2 isolation window) were selected from range 400-2000 m/z. The MS3 activation is HCD with 55% collision energy. The ions are analyzed in the orbitrap at 50,000 resolution with an AGC of 1.5e5 and an maximum injection time of 100 ms.

The proteomic (mass-spectrometry) data analysis is described in “mass spectrometry data analysis” in the “quantification and statistical analysis” section.

Western blotting
Cultures were grown as for the proteomics experiment. 100 µL of cells (1-2 x 10⁶ cells mL⁻¹) were lysed directly in 100 µL of 2x Laemli Sample Buffer (BioRad) + 5mM DTT, boiled at 95 °C for 10 min, and sonicated 3 s at 45% amplitude. Cell lysates were separated on a
10% SDS-polyacrylamide gel (BioRad), and transferred to a PVDF membrane using a semi-dry transfer system (BioRad). Membranes were blocked in 5% milk in PBS-T for 1 hour. The indicated primary antibody (PsBA - AS05 084A, Agrisera; PsBC - AS11 1787, Agrisera; PsSA - AS06 172, Agrisera; ATPC - AS08 312, Agrisera; AtpB - AS05 085, Agrisera) was added and incubated with shaking overnight at 4°C, followed by three washes in 1X PBS-0.1% Tween. Secondary antibody was added for 1 hour at room temperature, followed by three additional washes in PBS-T. Blots were imaged using ECL reagent on an iBright imaging system. To control for total protein levels, we again washed the blots 3x 5 min in PBS-T and re-blotted overnight for α-tubulin (AS10 680, Agrisera).

**Chloroplast transcriptome profiling (Chloroplast RNAseq)**

The RNA seq experiments were split into two experiments; each experiment had its own wild type. In each experiment, we had 2-3 replicates for each mutant strain and 2-4 replicates for the wild type.

The strains were grown in the same conditions as for the proteomic analysis. When the cultures reached ~2x10^6 cells/ml, we pelleted 13 ml of culture in 15 ml round Falcon tubes. We then used TRIzol extraction (following the manufacturer’s protocol) to obtain the total RNA. The RNA was sent to the Princeton Genomics Core Facility for RNAseq and Next Generation Sequencing. The chloroplast mRNA does not have polyA, so the facility used the Qiagen FastSelect – rRNA Plant Kit for RNA depletion. The facility then used the PrepX™ RNA-Seq for Illumina Library kit to generate the library for RNAseq.

mRNA analysis: First, non-coding RNA sequence was filtered out: each dataset was aligned (using the bowtie2 –fast command) against the dataset of non-coding RNAs, gotten only unaligned reads were included in the rest of the analysis. Next the reads were aligned against a reference file containing the updated chloroplast and mitochondrial genomes, a set of Chlamydomonas rRNA sequences (downloaded from [https://www.arb-silva.de/](https://www.arb-silva.de/)), and Chlamydomonas nuclear coding sequences (v5.5 from Phytozome, file Creinhardtii_281_v5.5.cds_primaryTranscriptOnly.fa), using the bowtie2 –fast option. For each sample, the number of reads in each chloroplast gene was calculated in python, with each side of each read considered separately, and with gene positions based on the chloroplast gff3 file from Gallaher et al.93

The reads were used to estimate the mRNA levels of the different chloroplast-expressed photosynthetic genes. The reads were normalized by the total chloroplast gene reads.

Our RNA seq reads were paired-end, allowing us to estimate splicing efficiency by analyzing where each side maps on the genome: when paired reads mapped to adjacent exons, the intron between them was considered spliced out. If the read in one end was in exon 1 and the read in the second end was in exon 2, this read was considered to be from a fully-mature mRNA. The overall coverage was much higher in our second experiment, so we normalized the 1st experiment using the wild-type ratio between the experiments, allowing us to present them together.

**Nuclear RNAseq**

The mRNA of pmr1 (2 independent experiments) and wild type (2 independent experiments) was also used for polyA-based RNAseq. The library preparation and Next Generation Sequencing were done at the Princeton Genomics Core Facility.

The paired-end reads were aligned against the primary transcriptome (v5.5, from Phytozome) using the bowtie2 –fast command, and the number of reads aligning to each transcript were counted in python for each sample.

We normalized the number of reads to 50M, then we averaged (using the geometric mean) the 2 experimental repeats of pmr1 and the 2 experimental repeats of wild type, and then calculated the relative reads by log2(pmr1/wild type).

**Measurement of NADP⁺ and NAD⁺ in wild type and pmr1 mutant (in-vivo)**

We used liquid-chromatography mass spectrometry to measure the cellular levels of NADP⁺ and NAD⁺ in wild type and pmr1 mutant. The protocol was adapted from Yuan et al.94 In short, we grew starter cultures at TAP dark for about a week, then inoculated experimental cultures in ~700ml of TAP in 1L bottles at an initial concentration ~10⁶ per ml. We grew the experimental cultures in the dark stirred using a magnetic stirrer at 200 RPM and bubbled with air until they reached ~2x10⁹ cells ml⁻¹. We harvested ~10⁷ cells using vacuum filter, and immediately dunked the filter’s membrane into 1.5 ml of 40:40:20 (v/v/v) methanol:acetonitrile:H₂O solution with 0.5% formic acid to extract the metabolites. All reagents were precooled to -20 °C and the protocol was performed on ice. After neutralizing by NH₄HCO₃ (132 L), and pelleting, we took 100 µl supernatant for LC-MS.

The LC-MS method was modified from Yang et al.95 Water-soluble metabolite measurements were obtained by running samples on the Orbitrap Exploris 480 mass spectrometer (Thermo Scientific) coupled with hydrophilic interaction chromatography (HILIC). An XBridge BEH Amide column (150mm X 2.1 mm, 2.5 uM particle size, Waters, Milford, MA) was used. The gradient was solvent A (95.5%:5.5% H₂O:acetonitrile with 20 mM ammonium acetate, 20 mM ammonium hydroxide, pH 9.4) and solvent B (100% acetonitrile) 0 min,90% B; 2min,90% B; 3min,75% B; 7min,75% B; 8min,70% B; 9min, 70%; 10 min, 50% B; 12 min, 50% B; 13 min, 25% B; 14 min, 25% B; 16 min, 0.5% B, 20.5 min, 0.5% B; 21 min, 90% B; 25 min, 90% B. The flow rate was 150 µL/min with an injection volume of 5 µL and a column temperature of 25 °C. The MS scans were in polarity switching mode to acquire data from both positive and negative ions across a mass range of 70–1000 m/z, with a resolution of 120,000. Data were analyzed using the EI-MAVEN software (v 0.12.0; Elucidata).

We included a total of 3 replicates from each strain from 2 independent experiments.
**Protein purification**

The *pmr1* rescued cells expressing PMR1-Venus-3×FLAG and the control cells expressing Venus-3×FLAG were pre-cultured in 50 mL TAP medium with 5 μg mL⁻¹ until the cell density reached ∼2-4 ×10⁶ cells mL⁻¹. Then, the culture was diluted into 1,000 mL TAP liquid medium to a concentration of ∼2 ×10⁸ cells mL⁻¹. Cells were grown with air bubbling and constant stirring at 210 RPM under 150 μmol photons m⁻² s⁻¹ light until the cell density reached ∼2-4 ×10⁸ cells mL⁻¹. Cells were collected by centrifugation at 3,000 g for 4 min in an Avanti J-26X centrifuge with an 8,100 rotor (Beckman) at 4 °C. The pellets were washed in 35 mL ice-cold washing buffer (25 mM HEPES, 25 mM KOAc, 1 mM Mg(OAc)₂, 0.5 mM CaCl₂, 100 mM Sorbitol, 1mM NaF, 0.3 mM Na₃VO₄, and cOmplete EDTA-free protease inhibitor (1 tablet/500 mL)) and then resuspended in a 1:1 (v/w) ratio of ice-cold 2×IP buffer (50 mM HEPES, 50 mM KOAc, 2 mM Mg(OAc)₂, 1 mM CaCl₂, 200 mM Sorbitol, 1mM NaF, 0.3 mM Na₃VO₄, and cOmplete EDTA-free protease inhibitor (1 tablet/50 mL)). 3 mL cell slurry was immediately added to liquid nitrogen to form small popcorn pellets which were stored at -80 °C until needed. Cells were lysed by cryogenic grinding using a Cryo-mill (Retsch) at frequency of 25 oscillations per second for 20 min. The ground powder was defrosted on ice for 45 min and dounced 25 times on ice with a Kontes Duall #22 homogenizer (Kimble). Proteins were solubilized by incrementally adding an equal volume of ice-cold 1×IP buffer plus 2 % digitonin (RPI) followed by an incubation of 45 min with nutation at 4 °C. The cell debris were removed by spinning at 12,700 x g for 30 min at 4 °C. The supernatant was then mixed with 50 μL anti 3×FLAG magnetic beads (Sigma) which had been previously washed sequentially with 1×IP buffer 3 times and 1×IP buffer plus 0.1 % digitonin 2 times. The mixture was incubated with nutation at 4 °C for 1.5 hr, followed by the removal of supernatant. The beads were washed 4 times with 1×IP buffer plus 0.1 % digitonin followed by a 30 min competitive elution with 45 μL of storage buffer (20 mM HEPES, pH 7.4, 350 mM KCl, 1 mM EDTA, 10% (vol/vol) glycerol, and 5 mM DTT) and 2 μg/μL 3×FLAG peptide (Sigma-Aldrich). Protein purity was assessed by SDS-PAGE followed by Coomassie blue staining.

**Measurement of NADP(H) dephosphorylation activity of PMR1 in vitro**

The NADP(H) dephosphorylation reaction was carried out at 22 °C using 1 mM NADP(H) (Roche) and 0.5 μM Nocturnin, PMR1-Venus-3xFLAG, or Venus-3xFLAG. Reactions contained 20 mM Tris·HCl (pH 8.0), 70 mM NaCl, and 2 mM MgCl₂. At the indicated time points, the reaction was quenched using 4 volumes of cold methanol, and then further diluted 100-fold with methanol before LC-MS analysis. The LC-MS conditions were the same as in the in-vivo experiment, except that we used an Exploris 240 mass spectrometer, and the mass range of 600–800 m/z was scanned. In our conditions the dominant form of NADP(H) was NADP⁺ (98%), so we followed this form in the experiment.

**Indirect Immunofluorescence Assay**

Indirect immunofluorescence was performed as described previously. First, cells were harvested by centrifugation and rinsed with PBS buffer twice. Next, 100 μL of cells were spotted onto Poly-L-lysine-coated glass slides (Sigma-Aldrich). Cell fixation was done by incubating with acetone at -30 °C for 1.5 hr, followed by the removal of supernatant. The beads were washed 4 times with 1×IP buffer plus 0.1 % digitonin followed by a 30 min competitive elution with 45 μL of storage buffer (20 mM HEPES, pH 7.4, 350 mM KCl, 1 mM EDTA, 10% (vol/vol) glycerol, and 5 mM DTT) and 2 μg/μL 3×FLAG peptide (Sigma-Aldrich). Protein purity was assessed by SDS-PAGE followed by Coomassie blue staining.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

The data quantification and overall statistical analysis were done using MATLAB. The sequencing data were analyzed by Cutadapt, Bowtie 2, and python. The confocal images were analyzed by Fiji. The structural data were analyzed and displayed using PyMOL.

The number of experimental repeats (n) is provided in the legends of the corresponding Figures. The error bars represent standard deviation (SD) and are described in the legends. The definition of center (mean or median) is described in the legends.
3) To reduce the noise, we removed samples with very low read counts in the TAP condition (<7 in the first experiment and <10 in the rest).

4) We calculated the relative growth as log₂ (averaged TP light reads / averaged TAP dark reads). In the first experiment, we had two different conditions; one was grown in hygromycin and paromomycin, and the other only in hygromycin; we analyzed them separately.

5) We normalized the NP experiment results – the overall distribution of relative growth rates in the NP experiment was shifted because most of the strains in this competition have a photosynthetic defect, so we scaled the results from this experiment by 0.6 to get a similar distribution to the other experiments.

6) For the final "growth score," we used the median of the five experiments with the strongest photosynthetic growth defects (for all but 122 genes, it is the same as using all the data). We used the five experiments with the photosynthetic growth defects because there are slightly different conditions between experiments, which can affect the phenotype. Furthermore, in some repeats, we were unable to see an effect because we did not manage to remove all the diploid cells. Lastly, the possibility that the mutants will have a phenotype "by chance" in more than five different experiments is very low, so even slightly lower effects for genes with many experimental repeats can be tolerated. The growth score and the light/dark ratio of backcrossing experiments for all the strains are shown in Table S1.

We used the "growth score" to set the 0.34 threshold to identify hits and to calculate the FDR (see below, and Figure S1). To reduce noise, we counted as hits only the strains that had reads above the threshold in at least two experiments.

7) FDR calculation (see also Figure S1) – to calculate the False Discovery Rate (FDR) we first estimated how many of the 1,616 mutated genes in our starting set are required for photosynthesis. We sampled 350 genes at random from the 1,616 and searched the literature for genes among them that are required for photosynthesis. Approximately 6.25% of the genes were known to be required for photosynthesis. Considering previous estimates indicating that approximately half of the genes required for photosynthesis remain to be discovered, we estimate that an additional 6.25% of the genes in the initial set are also required for photosynthesis; thus, we estimate that 12.5% of the initial genes are required for photosynthesis, and the remaining 1,414 (87.5% of the initial 1,620 genes) in our starting set are not required for photosynthesis. Next we defined a set of genes that we called "Genes whose disruption likely did Not Result in a Photosynthesis Defect" (GNRPD). We assigned genes from our set of 1,616 to GNRPD if they were represented by more than 20 insertions, where at most two mutants showed a photosynthetic defect in the Li et al. experiment. We selected the threshold of 0.34 as a compromise between low false-discovery rates and a relatively large number of hits. A phenotype threshold of 0.34 resulted in 136 hit genes identified, which included 2/204 (~1%) of the GNRPDs. We assume that the same percentage (~1%) of the 1,414 estimated genes in our starting set that are not required for photosynthesis in the original mutant set, will go into the hits, resulting in a calculated FDR < 0.11 when using a threshold of 0.34. With a threshold of 0.49, the same calculation yields 227 hit genes with an FDR < 0.3.

As a sanity check for the FDR calculation, we also calculated the hit p-value based on linkage distance (Figure S1).

**Mass spectrometry data analysis**

Mass spectrometry raw data were analyzed using GFY software licensed from Harvard to quantified proteins relative abundance.

We normalized each protein’s abundance in each sample by that protein’s abundance in the corresponding wild-type/control sample, then normalized the protein’s abundance in the sample by the sample’s median to account for systematic difference likely coming from technical difference in the amounts of proteins entered into the TMT labeling.

To decrease the noise, we used 11-plex-median-based normalization (Figure S4). We divided the abundance of each protein in a given sample by the median abundance of this protein in its 11-plex. This normalization sets the median of each 11-plex to 1 on a linear scale (0 on a log scale). This normalization process intends to correct two kinds of artifacts: 1) when one protein is over/under-represented in all samples of one specific 11-plex (as in the case of S4A); and 2) to set the overall median relative abundance of this protein across all 11-plexes to 1 on a linear scale (0 on a log scale), to control for systematic effects. Systematic effects, such as the underrepresentation of ribosomal proteins in the data before the 11-plex-median-based normalization visible in Figure S4E, are likely due the reference wild-type control strain that was included in the 11-plexes and used to calculate the proteins relative abundance. This normalization improves the overall quality of the data, as seen in Figure S4.

We are aware that if most of the mutants in a group have similar proteomic effects, the median normalization could lead to the over- or underestimation of the abundance of that protein. This effect is rare because the mutants were selected at random, and a change in the median value would require five of the mutants to have a similar effect on the proteome. Furthermore, we randomized the mutants present in the 11-plexes of the two repeats. If the two repeats disagreed, we assumed there was an experimental problem and performed an additional repeat. Thus, it is unlikely that our normalization strategy would produce significant artifacts in the proteomics data shown in the figures.
More than 65% of our hits are conserved in land plants.

Figure S1. Pooled backcrossing and FDR calculation, related to Figure 1

(A) The mutants carried barcoded cassette insertions and additional mutations. To determine which barcoded cassette insertions are linked to the observed photosynthetic phenotype, we mated the mutants, which were paromomycin-resistant and mt−, with a hygromycin-resistant mt+ strain. The resulting progeny included mixed genotypes where the insertions and the second-site mutations segregated randomly. We grew the progeny under a dark control condition, where all viable strains grew, and photoautotrophically (~750 µE/m²/s), where mutants in genes required for photosynthesis were depleted. By sequencing the pools of barcodes associated with insertions, we could identify barcodes that were depleted under the photoautotrophic condition and thus were genetically linked to genes required for photosynthesis.

(B) Calculation of the “estimated number of genes in our starting set that are not required for photosynthesis.” Our dataset included 1,616 genes with mapping confidence level <4. We sampled 350 genes at random from the 1,616 and screened the literature for genes among them that are required for photosynthesis. 6.25% of the genes were known to be required for photosynthesis. Considering previous estimates indicating that approximately half of the genes required for photosynthesis remain to be discovered,25 we estimate that an additional 6.25% of the genes in the initial set are also required for photosynthesis; thus, we estimate that 87.5% of the genes in our starting set are not required for photosynthesis. Given these numbers, the estimated number of genes in our starting set that are not required for photosynthesis is 1,414 (87.5% of the initial 1,616 genes).

(C) The false discovery rate (FDR) calculation is based on a set of specific genes that we called genes whose disruption likely did not result in a photosynthesis defect (GNRPD). Genes from our set of 1,616 genes were assigned to GNRPD if they were represented by more than 20 insertions in Li et al. experiment and at most two mutants showed a photosynthetic defect. ~1% of the GNRPDs were among the 136 hit genes identified with a phenotype threshold of 0.34. We assume that the same ratio (~1%) of the estimated number of genes in our starting set that are not required for photosynthesis (see B) in the original mutant set will go into the hits, yielding an estimated FDR < 0.11. In the manuscript, we focused on this threshold due to its low FDR and due to the shape of the distributions (Figure 1E): the GNRPD distribution goes down to almost zero below this threshold.

(D) The same calculation as (C) was repeated for lower-confidence hits (phenotype threshold of 0.49) and indicated that these lower-confidence hits have FDR < 0.3. Despite the higher false discovery rate, these hits still include many genes genuinely required for photosynthesis.

(E) We used whole-genome paired-end Illumina sequencing to identify the insertion locus of the hygromycin resistance cassette in the wild-type strain. We identified the insertion in chromosome 2 around position 2.25 Mb. Cassette-genome reads are chimeric paired reads where one read maps to the genome and the other maps to the cassette. Genome-genome reads are paired-end reads where both reads mapped to adjacent regions in the genome; such reads are depleted in the proximity of an insertion. The insertion appears complex, and the data are insufficient to fully map the insertion site with confidence, but a model consistent with the observed data is that there is a genomic inversion from ~2248150 to ~2248450, followed by the cassette insertion, followed by poorly mappable sequence, and then genomic sequence resuming at ~2248800 (with a genomic deletion from ~2248450 to ~2248800).

(F) To estimate the linkage distance, defined as the minimum distance between two markers needed for independent segregation in our backcrossing, we evaluated barcode read counts from backcrossed pools grown in the dark on hygromycin, where all barcodes should be present except for the ones that were too close to the hygromycin cassette for recombination to occur. We measured the linkage distance as the distance from our HygroR marker to the edge of the region of depleted insertions. To reduce noise levels, we used only insertions with a mapping confidence level of 95% and smoothed the data by using a running median. This linkage distance was 0.25 Mb on one side of the mapped hygromycin cassette insertion and 0.6 Mb on the other side of the insertion. Note that the FDR metric we use is more conservative than a p value based only on linkage size. Even if we assume an upper-limit linkage region size of 2 Mb (1 Mb from each side), the chance for a random insertion to be inside the linkage region is <0.017 (2/120 Mb—Chlamydomonas genome size). Even if we assumed a worst-case scenario where each strain has four additional unknown mutations, the chance that all of them will be in the rest of the genome is >0.93 (118/120)⁶, so the chance that at least one of them will be in the linkage region is <0.07, and thus, a p value based only on linkage size would be p < 0.07. We are thus more conservative by stating that our FDR is <0.11.

(G) 26 of our 115 hits (23%) were also hits in Wakao et al.,21 and 68 of the 115 (59%) were also hits in Li et al.25

(H) More than 65% of our hits are conserved in land plants.
A Initial set: 1,781 mutants
Backcrossing
227 linked genes FDR<0.3
136 linked genes FDR<0.11
Mapping validation
117 genes

Manual correction:
- removed 12 suspicious genes
- added 10 genes that are likely correct
  but did not make the automatic cutoff

Our hit set: 115 genes
70 previously-uncharacterized genes
29 Chlamydomonas photosynthesis genes
16 homologs of known photosynthesis genes
93 of our hits + 7
4 known genes
1 hits from other sets
validation
by rescue
12 previously-uncharacterized
4 known

Proteomics on mutants representing 100 genes

B WT
Mutant
WT PCR product ~1kb
Mutant PCR product ~3kb

C WT
Mutant
WT PCR product ~1kb
No PCR product for the Mutant

D WT
Mutant
WT PCR product ~1kb
No PCR product for the Mutant

E WT
Mutant
WT PCR product ~1kb
Mutant PCR product ~1kb

F DNA:
Mutant
WT

<table>
<thead>
<tr>
<th>Length (bp)</th>
<th>Control</th>
<th>Check</th>
<th>Check</th>
</tr>
</thead>
<tbody>
<tr>
<td>3000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

G WT
Mutant
Expected insertion point

(legend on next page)
Figure S2. Supplemental data for the genetic screen, related to Figure 1

(A) Mutant and gene numbers at different stages of this project. For a detailed description of the process, please see the STAR Methods. Mutant and gene IDs are provided in Table S1.

(B–F) Mapping validation by colony PCR. Displayed are four scenarios of cassette insertion and the expected PCR product (B–E), and PCRs examples (F). (B) Clean insertion—the cassette integrates into the genome cleanly; in this situation, the PCR product of the mutant will be approximately 2 kb longer than the WT product. (C) Insertion with significant deletion—in this case, the deletion associated with the insertion removed one of the genomic PCR primer-binding sites; therefore, we will get the PCR product for WT but not from the mutants. (D) Insertion with rearrangement—in this case, the primer sequence is there but in the incorrect orientation, so again we will get a PCR product for WT but not for the mutant. (E) When the insertion is not in our expected gene, we will get the same length of PCR product from the WT and the mutants. (Note that we can get this pattern also if the insertion is associated with a deletion of a similar size.) (F) Example of colony PCR results. The control lane is mutant DNA amplified using control primers to verify the mutant DNA quality. In the upper example, the mutant is ~2 kb longer than the WT, as expected from a clean insertion (B). In the middle example, we have a band for the WT but not for the mutants. Such a result was interpreted as validating an insertion site if it was reproduced at least twice, and is expected for scenarios (C) and (D). The lower example was interpreted as a failure to validate the mapping and is expected for (E). When we failed to get a product with WT, we used different primers or whole-genome sequencing to map the insertion site.

(G) Mapping validation by sequencing. Mutant genomes were sequenced using Illumina paired-end 150 nt reads. We considered an insertion site validated when we found in the expected area chimeric paired-end reads (where one read mapped to the genome and the other to the cassette) and a “hole” in the genome coverage. For more details, see STAR Methods.
Figure S3. Additional images, related to Figures 1, 2, and 6

(A) Images for Figure 1B before background removal.

(B) Unprocessed plate images for Figure 2. In each high-light plate, the three copies of the original mutants are outlined in dashed red and every triplicate of the rescued strains is outlined in dashed black. To reduce the effect of location on the plate, we put one WT next to each mutant trio. The “r” indicates the rescued strain used in the main figure. Similarly, “m” indicates the mutants and “w” the WT used in the main figure. There are differences in the rescue efficiency between

(legend continued on next page)
the different rescued strains, even in the same mutant. Many parameters could contribute to those differences, including insertion site and expression level. The light level for the photosynthetic conditions was approximately 1,100 μE/m²/s.

(C) The unprocessed plate images for Figure 6. The light levels for the photosynthetic conditions were 550 μE/m²/s for pmr1 and 800 μE/m²/s for mtf1.
(D) Confocal images of WT (without any fluorescent tag), and un-tagged Venus.
(E) Additional confocal images for Figures 2 and 6.
(F) Control for PMR1’s indirect immunofluorescence shown in Figure 6Q. As expected, much-lower anti-PMR1 signal is observed in the pmr1 mutant (LMJ.RY0402.206992).
(A and B) Example of the data of two proteins (Cre01.g004900, A and PsaB, B) across all samples Without and with 11-plex-median-based normalization. Each proteomic 11-plex (10 samples and WT) is shown in one color, and the 11-plex’s median is shown as an x in the same color. In addition to the samples included in Figure 4, these data include additional samples including controls and repeats that were subsequently filtered out due to noisy or contaminated samples. Without normalization, we can see systematic differences in protein abundance between the 11-plexes, which are removed by normalizing the protein’s measured abundance using the group median, which we refer to as the 11-plex-median-based normalization. The black lines represent the median of all the samples.

(C) The normalization reduces the noise and systematic errors in the data. Protein levels are shown for proteins measured in at least 65% of the experiments characterizing the 100 mutants. The data are the average of two repeats on the log2 scale. The upper panel is before, and the lower panel is after the 11-plex-median-based normalization. We can see that the normalization removes much of the noise and maintains most of the signal. The left-most ∼90 proteins are the ones shown in Figure 4.

(D) Scatterplots comparing the two replicate measurements of protein levels in the mutants shown in Figures 4B–4I.

(E) The data from Figure 4 are shown without the 11-plex-median-based normalization. A systematic underrepresentation of ribosomal proteins is apparent; we think this is due to a relative overrepresentation of ribosomal proteins in the reference wild-type strain that was included in the 11-plexes and that was used to initially normalize all raw data. As can be seen in Figure 4, the 11-plex-median-based normalization eliminates such systematic under- and overrepresentation.
Figure S5. Proteomic controls, related to Figure 4

(A) Proteomes are compared between two different mutant alleles of the same gene. We only show data where we obtained data for both alleles and at least one of the alleles showed proteomic changes above the noise level. Each axis represents one allele’s log2(mutant/WT) proteomic data. The sample name is shown near each axis.

(B) Genes for which we rescued the mutants and collected proteomics data for both the mutants and the rescued strains. (A and B) Our data suggest that the impact on the photosynthetic complexes is from our mutant gene in all cases except for TRX21. The two trx21 mutants have different phenotypes: one was yellow and had a decreased abundance of chlorophyll-binding proteins (including PSII), and the other was green and only affected PSII and the small subunit of the chloroplast ribosome. These observations suggest that the yellow mutant has an additional mutation leading to the additional proteomic phenotype. Additionally, 5 genes (HCF173, CPLD64, CHLM, RAA6, and RAA17) showed strong proteomic and photosynthetic phenotypes, and their rescue restored the mutant to WT-like growth. This demonstrates that only in rare cases (1/16) does the prominent proteomic phenotype come from a second mutation.

(C) Proteomic validation that the mutated protein is absent from the strains. We show the proteins’ relative abundance without the 11-plex-median-based normalization (see Figure S4E). We observed downregulation of the mutated protein in all cases where we could measure the protein except for Cre09.g407650 (encircled in red), suggesting that Cre09.g407650 is a false positive. The insertion in Cre09.g407650 is in the 3’ UTR and was linked to the phenotype; this insertion is likely not the reason for the photosynthetic phenotype, demonstrating how proteomics can help identify false positives.
Figure S6. Supplemental data for psaA mRNA maturation factors, related to Figure 5
(A–C) Scatterplots of proteomic data in mutants in known psaA maturation factors (RAA1 and RAA4) and mutants in genes with similar proteomic profiles (HEL5, RAA12, PIR2, and RAA18). The data reflect the average normalized log₂ (mutant/WT protein abundance) from two independent experiments.
(D) Scatterplot of two replicates of proteomic data of hel5 mutants.
(E) The mRNA levels (normalized to WT) of psaB, psaj, and psac in the different mutants. The largest effects are 2-fold changes in psaB levels, which is within the noise level and is not expected to affect translation levels. Error bars represent SE. The chloroplast mRNA does not have poly(A), so we used an rRNA depletion kit to remove most of the rRNA before amplifying RNA (see the "chloroplast transcriptome profiling (chloroplast RNA-seq)" section of the STAR Methods).
(F) RAT2 and RAA1 are required for tscA processing. This requirement suggests that tscA processing is carried out in conjunction with the splicing complex organized around RAA1. Error bars represent SE.
Figure S7. Supplemental data for MTF1, CIF2, and PMR1, related to Figure 6

(A) Protein levels of chloroplast-expressed genes and mitochondrial controls in mtf1, cif2, pmr1, rescued MTF1, and rescued PMR1. The data represent the median of 6, 3, 4, 4, and 3 independent experiments, respectively. Those replicates include the 2 experimental replicates shown in Figure 4 for mtf1, cif2, and pmr1, as well as additional experimental replicates.

(B) Detailed comparison of the chloroplast-expressed rpo RNA polymerase, rps and rpl ribosomal proteins, the clpP protease, and chl chlorophyll biosynthesis proteins between mtf1 and cif2 mutants. The median of 6 and 3 repeats ± SE is shown.

(C) Comparison of the chloroplast-expressed ribosomal proteins between mtf1 and cif2 mutants. The data in this control experiment are based only on two experimental replicates of each mutant that were run together in the same 11-plex to allow more-direct compression between them than the data shown in (B), which were obtained from different 11-plexes. Note that the data are very similar in both analyses.

(D) Images of the western blot membranes. We used each membrane to measure the relative abundance of one protein (using immunoblotting), then we used α-tubulin immunoblotting on the same membrane as a loading control (the image of the α-tubulin blot is shown below each membrane). We performed two experimental replicates with each antibody and used the second replicate for Figure 6. Note that the AtpB antibody recognizes not only the chloroplast’s AtpB (55 kDa) but also the mitochondrial ATP synthase beta subunit (70 kDa); see the product information on Agrisera website for antibody AS05 085. For PsbA, a band with a lower molecular weight was observed when its antibody was used on Chlamydomonas; this is thought to be due to D1 degradation (see Agrisera website for antibody AS05 084A).

(E and F) Comparison between the AlphaFold-predicted MTF1 structure I and the crystal structure of E. coli MTF64 (F). The conserved active-site residues (Asn108, His110, and Asp146 in E. coli MTF64; corresponding to Asn160, His162, and Asp198 in MTF1) are shown in red, and fMet is shown in black. For a better comparison of the active sites, we used YRB, a script that displays the hydrophobic pockets (yellow) and negative charges (red) on a protein surface. In both active sites, we can see hydrophobic pockets below the fMet and negatively charged active-site residues above it.

(G) Expression data41,98 for MTF1, CIF2, and Cre09.g392729 (encoding the mitochondrial predicted MTF1 ortholog) are shown for different growth conditions.

(H) There is a strong negative correlation (r = -0.78) between the levels of chloroplast-expressed complexes in the pmr1 mutant and their levels in the PMR1 rescued strain, supporting the idea that PMR1 overexpression in the rescued strain leads to the overexpression of chloroplast-expressed complexes.

(I) Comparison between the AlphaFold-predicted PMR1 structure (green) and the crystal structure of human nocturnin (pink).99

(J) PMR1 diurnal expression. The light period is shown in yellow, and the dark period is shown in gray. The data are from Strenkert et al.71

(K) The pmr1 mutant affects the mRNA expression of many genes. The log2(pmr1−1/WT2) data are shown in red, and as control, the log2(WT1/WT2) data are shown in black.

(L) PMR1 NADP⁺ phosphatase activity in vitro. We started with NADP⁺ and used LC-MS to follow the accumulation over time of NAD⁺ after the addition of one of the following proteins: NOCT (positive control), PMR1-Venus-3xFLAG, or Venus-3xFLAG (negative control—contaminants from the IP may contribute phosphatase activity). The data represent the median of three replicates ± SE. For each protein we calculated a linear fit using MATLAB’s “polyfit” command. The linear fits’ slopes (normalized by 61.8), and their SE (calculated by MATLAB’s “fitlm” command) were used to generate Figure 6L.

(M) Comparison of the substrate-binding pockets between the NOCT structure (pink) and AlphaFold-predicted PMR1 structure (green). NADP⁺ is illustrated in orange. Changing R290 to K192 affects the shape of the substrate-binding region, and lysine K377 may partially block the entrance to the substrate-binding pocket in PMR1.

(N) The pmr1 mutant does not lead to the downregulation of mitochondrial gene mRNAs. Each dot represents the mean of 2 experimental replicate measurements of one gene. The bar represents the median value across all genes in a group.

(O) The predominant form of PMR1 in the rescued pmr1:PMR1-Venus-3xFLAG strain is of the expected molecular weight. We performed an anti-FLAG immunoprecipitation from (1) the rescued pmr1:PMR1-Venus-3xFLAG strain and (2) a strain expressing Venus-3xFLAG. The products were run on a gel and stained with Coomassie (EZBlue gel).