Title

Discovery of photosynthesis genes through whole-genome sequencing of acetate-requiring mutants of *Chlamydomonas reinhardtii*

Setsuko Wakao[1,2]*, Patrick M. Shih[3,4], Katharine Guan[2,5], Wendy Schackwitz[6], Joshua Ye[2,5], Robert M. Shih[1], Mansi Chovatia[6], Aditi Sharma[6], Joel Martin[6], Chia-Lin Wei[6§], Krishna K. Niyogi[1,2,5]*

Affiliations:

[1]Division of Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[2]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA

[3]Department of Plant Biology, University of California, Davis, CA 95616, USA.

[4]Joint BioEnergy Institute, Emeryville, CA 94608, USA

[5]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA

[6]Joint Genome Institute, Lawrence Berkeley National Laboratory, CA 94720, USA

[§]Present address: Jackson Lab, Farmington CT, 06032

*For correspondence: swakao@berkeley.edu and niyogi@berkeley.edu

Abstract

Large-scale mutant libraries have been indispensable for genetic studies, and the development of next-generation genome sequencing technologies has greatly advanced efforts to analyze mutants. In this work, we sequenced the genomes of 660 *Chlamydomonas reinhardtii* acetate-requiring mutants, part of a larger photosynthesis mutant collection previously generated by insertional mutagenesis with a linearized plasmid. We identified 554 insertion events from 509 mutants by mapping the plasmid insertion sites through paired-end sequences, in which one end aligned to the plasmid and the other to a chromosomal location. Nearly all (96%) of the events were associated with deletions, duplications, or more complex rearrangements of genomic DNA at the sites of plasmid insertion, and 1405 genes in total were affected. Functional annotations of these genes were enriched in those related to photosynthesis, signaling, and tetrapyrrole synthesis as would be expected from a library enriched for photosynthesis mutants. Systematic manual analysis of the disrupted genes for each mutant generated a list of 273 higher-confidence candidate photosynthesis genes, and we experimentally validated two genes that are essential for photoautotrophic growth, *CrLPA3* and *CrPSBP4*. The inventory of candidate genes includes 55 genes from a phylogenomically defined set of conserved genes in green algae and plants. Altogether, 68 candidate genes encode proteins with previously characterized functions in photosynthesis in *Chlamydomonas*, land plants, and/or cyanobacteria, 15 genes encode proteins previously shown to have functions unrelated to photosynthesis, and 190 genes encode proteins without any functional annotation, signifying that our results connect a function related to photosynthesis to these previously unknown proteins. This mutant library, with genome

sequences that reveal the molecular extent of the chromosomal lesions and resulting higher-confidence candidate genes, represents a rich resource for gene discovery and protein functional analysis in photosynthesis.

Introduction

Since the dawn of modern genetics, mutagenesis has been the primary vehicle to perturb the underlying genetic code of organisms, enabling scientists to investigate the genetic determinants underpinning biological systems. In the case of photosynthesis, much has been learned through mutagenesis of the unicellular green alga, *Chlamydomonas reinhardtii*, which has proven to be an indispensable reference organism for investigating the molecular components, regulation, and overall processes of photosynthesis (1,2). *Chlamydomonas* has a haploid genome and an ability to use acetate as a sole carbon source, which facilitates the isolation and analysis of knock-out mutants that are defective in photosynthesis (3). Moreover, the advantage of working with a unicellular alga rather than a whole plant has facilitated the speed with which molecular and genetic studies can be carried out (4). Thus, the development of resources and tools to increase the breadth and depth of genetic studies in *Chlamydomonas* has advanced our ability to understand the molecular basis of photosynthesis.

Numerous large-scale mutagenesis and screening experiments have been carried out in *Chlamydomonas*, with some of the earliest efforts described over half a century ago (3,5,6). Classical mutagenesis studies have utilized chemical and physical mutagens, which induce untargeted genomic lesions and rearrangements across the genome. Identifying the causative mutations requires genetic mapping through crosses, an approach that is robust but time consuming. Insertional mutagenesis approaches, in which a selectable marker is transformed and randomly integrated into the genome, have facilitated molecular analysis, and many PCR-based techniques have been successfully employed in *Chlamydomomas* to rapidly identify flanking sequence tags (FSTs) from the site of marker insertion (7–14). However, the efficiency of FST recovery can be low (7) because of the complexity of events accompanying plasmid insertion such as concatemerization, chromosomal deletion or rearrangement, loss of the primer annealing sites, as well as difficulties with PCR from the *Chlamydomonas* nuclear genome, which is GC-rich and contains a high degree of repetitive sequences (15). High-throughput FST recovery has been achieved in *Chlamydomonas* (8,10) and has offered a large collection of insertional mutants for the scientific community while enabling large-scale mutant analysis of photoautotrophic growth (9).

The advent of next-generation sequencing methods has dramatically improved our ability to identify mutations by whole-genome sequencing (WGS). In *Chlamydomonas*, this approach was initially combined with linkage mapping to identify point mutations in flagellar mutants (11,12), and it was used subsequently for point mutations affecting the cell cycle (13,14) and light signaling (16,17). In the case of insertional mutants, WGS has been used extensively to identify insertion sites in bacteria and some microbial eukaryotes with smaller genomes (18–20) but only for a relatively small number of mutants in *Chlamydomonas* (21). In maize, due to its large genome, high-throughput next-generation sequencing of *Mu* transposon insertion sites has been applied only after enrichment for the transposon sequence (22), whereas the large volume of insertion site information of T-DNA insertion lines in *Arabidopsis* was obtained from traditional PCR-based FST isolation (23–25).

We have previously generated a large insertional mutant population of *Chlamydomonas* by transformation with a linearized plasmid conferring paromomycin or zeocin resistance, and we identified mutants with photosynthetic defects (*i.e.*, acetate-requiring and/or light-sensitive and reactive oxygen species-sensitive mutants) (7,26). However, we were only able to obtain FSTs for 17% of the mutants using PCR-based approaches. Here we employed low-coverage WGS of a subset of 660 mutants to identify the plasmid insertion sites and accompanying structural variants, and we found 1405 genes that are affected by the plasmid insertion in 509 mutants. We generated a list of 273 genes from 348 mutants that we refer to as higher-confidence causative genes, enabling the discovery of 205 potential photosynthesis genes; 190 genes of previously unknown function and 15 genes previously shown to have functions unrelated to photosynthesis. We experimentally validated two genes, *CrLPA3* and *CrPSBP4,* that are required for photoautotrophic growth in *Chlamydomonas*. In addition, our data provide insight into the spectrum of mutations that are induced by insertional mutagenesis in *Chlamydomonas*.

Results

Identification of insertion sites by mapping of discordant read pairs

We re-screened our *Chlamydomonas* photosynthetic mutant collection (7,26) for growth on minimal and acetate-containing media under three light conditions (dark, D; low light of 60-80 μmol photons m^{-2} s^{-1}, LL; and high light of 350-400 μmol photons m^{-2} s^{-1}, HL) and for maximum photochemical efficiency of photosystem (PS) II (F_v/F_m) (S1 Table). An example of the phenotyping is shown in Figure 1. A total of 660 mutants, most of them with a growth phenotype and with resistance to either zeocin or paromomycin, indicative of the presence of the linearized plasmid sequence used for insertional mutagenesis, were chosen for WGS and herein will be referred to as the Acetate-Requiring Collection (ARC).

Genomic DNA was extracted from the 660 ARC mutants and submitted for low-coverage, paired-end WGS with a target depth of sequence coverage for each mutant between 5 and 10. The average sequencing depth across samples was 7.44. Paired-end reads that showed one end mapping to the plasmid used for mutagenesis and the other to a chromosome location were used to identify the plasmid insertion site(s) in each mutant. Plasmid insertion sites were not identified for 72 mutants, because few plasmid sequence reads were detected or the other end mapped to a low complexity region of the *Chlamydomonas* genome. 79 mutants had insertions that were not unique within the population (33 were duplicated, three were triplicated and one was quadruplicated) and were removed from further analysis. The remaining 509 mutant sequences were further analyzed for structural variants (insertions, deletions, and rearrangements) that occurred during insertional mutagenesis.

Figure 2 illustrates the types of structural variants detected by analysis of the paired-end sequence data. Most sequence read pairs were concordant, i.e., they showed the expected orientation and distance with respect to each other when mapped to the *Chlamydomonas* genome (Figure 2, dark gray arrows). In contrast, discordant pairs showed the incorrect orientation or distances that were closer or further from each other

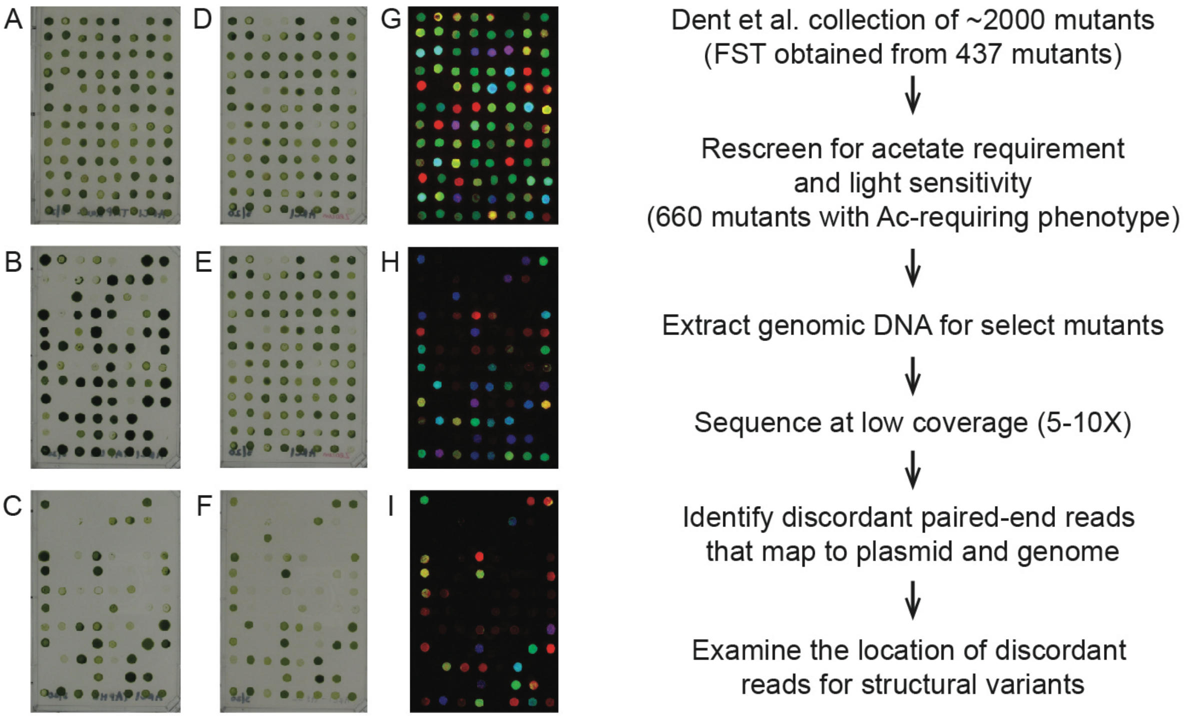


Fig 1. Growth and chlorophyll fluorescence screen pipeline.
Mutants were scored for growth on (A) D+ac, (B) LL+ac, (C) HL+ac, (D) LL+ac+zeocin, (E) LL-min, (F) HL-min. Fv/Fm values were measured on cells grown on (G) D+ac, (H) LL-min, (I) HL-min. FST, flanking sequence tag. A representative plate spotted from a 96-well plate is shown. D, dark; LL, low light; HL, high light; +ac, added acetate; min, minimal media.

than expected based on the genome fragmentation that was performed during sequencing library preparation (genomic DNA was sheared to approximately 600 bp) or on different chromosomes. In Figure 2, the discordant reads are shown as colored arrows, with each color representing a chromosome (or plasmid) to which the corresponding paired-end read was mapped. Each of these genomic sites where sequence read pairs were discordant is listed in S1 Table as a “Discordant site”.

At most of the plasmid insertion sites, two sets of discordant read pairs were found, with their chromosomal reads oriented toward each other and their paired-end reads mapping to the plasmid sequence (Figure 2 blue box). We refer to these 425 events as two-sided insertions, where both sides of the plasmid insertion were unambiguously mapped (S1 Table, column “Number of sides paired with plasmid at site”, 2). Another large group of discordant sites displayed only one set of discordant read pairs located on one side of the plasmid insertion (referred to as one-sided insertions Figure 2; S1 Table, column “Number of sides paired with plasmid at site”, 1). The read-pairs on the other side of the plasmid insertion could not be mapped in 21 of these insertion sites because (i) it was at a repetitive region (14 mutants) and (ii) it had no discordant reads (7 mutants). These 21 one-sided insertions together with the 425 two-sided insertions making a total of 446 insertions and were considered to be simple insertions (S1 Table). In the rest of the one-sided insertions, the other side of the plasmid insertion paired with another chromosomal region indicating an occurrence of a more complex chromosomal rearrangement. Insertions that paired with another chromosomal location was considered a complex insertion. The frequencies of two-sided, one-sided, and complex insertions are shown in S1 Figure.

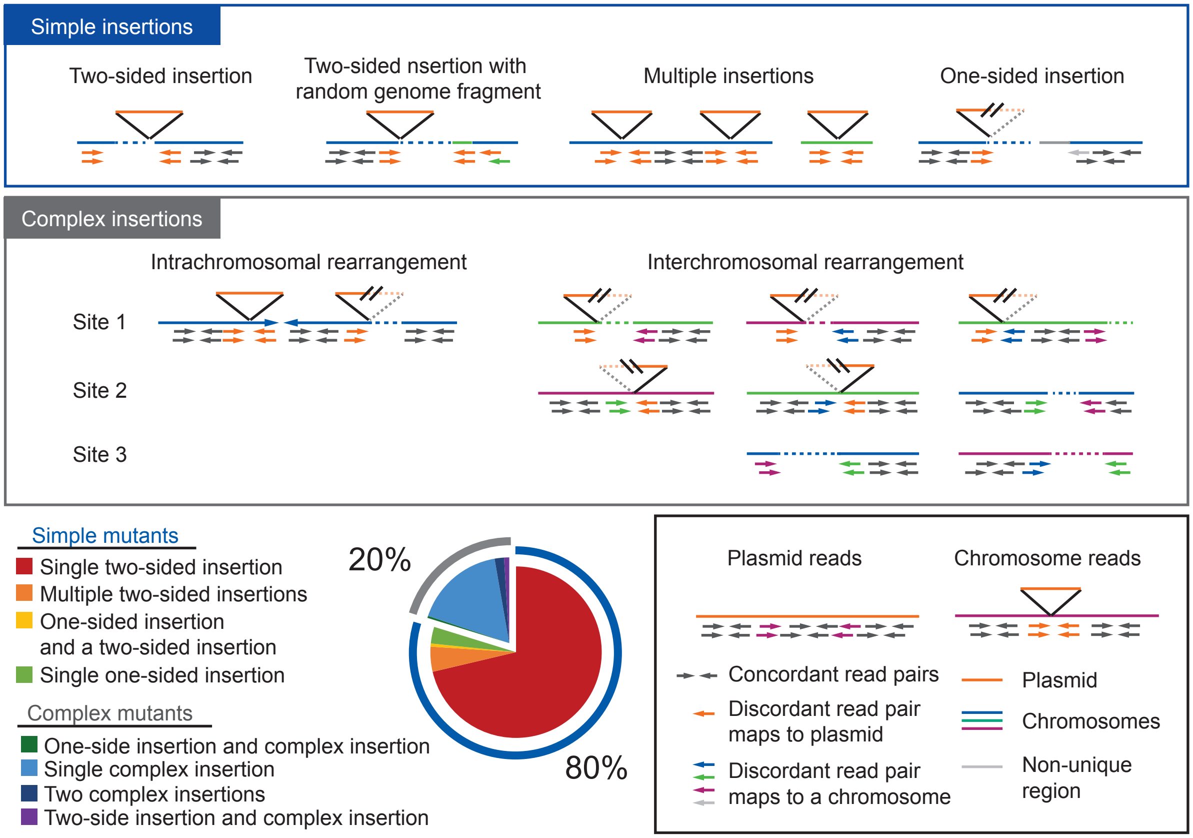
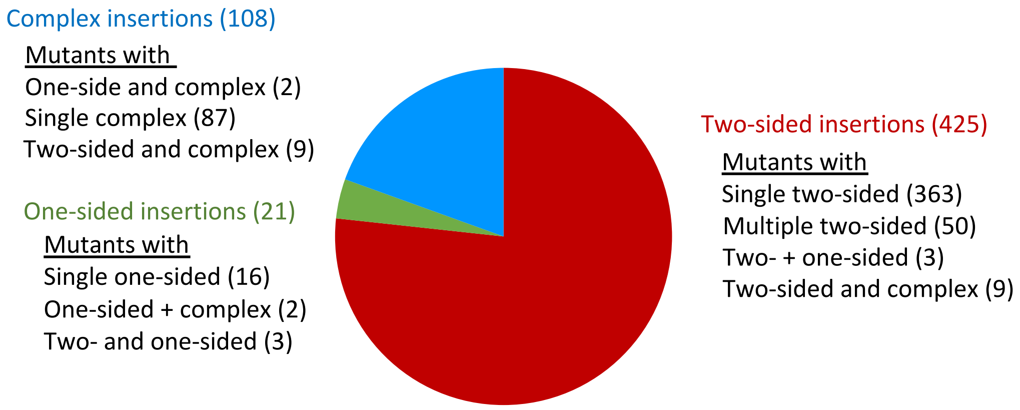


Fig 2. Examples of structural variations and the frequency mutants with simple or complex insertions in ARC. Boxes contain illustration examples of mapped reads as seen in IGV. Black box, mapped reads (concordant and discordant) against plasmid and chromosome. Blue box, examples of "Simple insertions"; Gray box, examples of "Complex insertions". Gray box shows examples of different complex insertions that are intra- or interchromosomal rearrangements. Second from left in gray box shows a possible translocation between two chromosomes. Pie chart shows frequency of "Simple mutants" containing only Simple insertions and "Complex mutants" containing complex insertions.

A total of 406 mutants (80%) contained only simple insertions accounting for 435 out of the total 446 simple insertions (11 mutants contained both simple and complex insertions) (Figure 2, Simple mutants). Among these 406 mutants, 24 mutants had multiple (two or three) two-sided insertions accounting for 50 insertions, and three mutants had one two-sided insertion and one one-sided insertion (Figure 2, S1 Figure). In 17 mutants, the multiple simple insertions occurred on the same chromosome, and six of these had tandem two-sided insertions that disrupted the same or neighboring genes. In 10 two-sided insertions (~1.8%), there appeared to be a short random fragment of another chromosome inserted together with the plasmid (Figure 2, Two-sided insertion with random genome fragment). The original locus of these random fragments did not show a lack of mapped sequence reads but rather showed double the abundance of reads mapping to the small region, indicating that it was an extra copy of the same sequence at the insertion site, similar to what was observed in a previous study but at a lower frequency in ARC (8).

The other group of 103 mutants (20%) contained at least one complex insertion (Figure 2 “Complex mutants”; also see S1 Table, “Pairing with other discordant site(s) of the same mutant”). Nine of these mutants had a coexisting two-sided insertion, two mutants had an additional one-sided insertion, and five mutants contained two independent complex insertions. Some of these rearrangements occurred on a single chromosome, and others involved two or more chromosomes (Figure 2, gray box). Among interchromosomal rearrangements, 13 of them involved two one-sided insertions that were paired to each other (Figure 2 gray box). These together may represent chromosomal translocation events resulting in two chimera chromosomes. In all of these



S1 Fig. Proportion of different types of insertions observed in ARC.
The frequency of the different types of insertions. Some insertions coexist with another insertion in a mutant. The number of mutants grouped by the types of insertions it contains is listed along with the number of insertions accounted for in that group.

possible translocation events, the plasmid sequence was present in one junction and not in the other. The proportion of complex insertion events was similar among the three plasmids used for transformation (pSP124S, pMS188, and pBC1). Validation of these complex structural variants would require *de novo* assembly of sequencing reads. Most mutants only contained only two-sided or only complex insertions; 387 mutants (76%) had only two-sided insertion(s) (Figure 2, red and orange slices), 92 had only complex insertion(s) (18%) (Figure 2, light blue slice), and only a small proportion of mutants contained a mix of two-sided, one-sided, or complex insertions.

In summary, low-coverage WGS data for 509 ARC mutants identified 406 mutants that contained only simple insertions accounting for 435 out of 446 total simple insertions, whereas 103 mutants contained complex insertions that were associated with chromosomal rearrangements such as inversions and translocations.

Analysis of deletions and duplications associated with insertional mutagenesis

Insertional mutagenesis in *Chlamydomonas* has been previously associated with deletions and duplications at the site of plasmid insertion, especially when using glass bead for transformation (e.g. *cpld38*, *cpld49*, *npq4, rbd1*) (27–29). Focusing on the 425 two-sided insertions, we found deletions associated with 374 insertions (88%). A wide range of deletion sizes was observed, with a bimodal distribution peaking at 101-1000 bp and 10 - 100 kb when plotted at $\log_{10}$-scale, the largest deletion being 133 kb (Figure 3A). Duplications occurred less frequently (7%), in a total of 29 insertion events (Figure 3B), and all were less than 1000 bp. Perfect insertions lacking any duplications or deletions were found in only 22 events (5%). Despite the high frequency and relatively large size

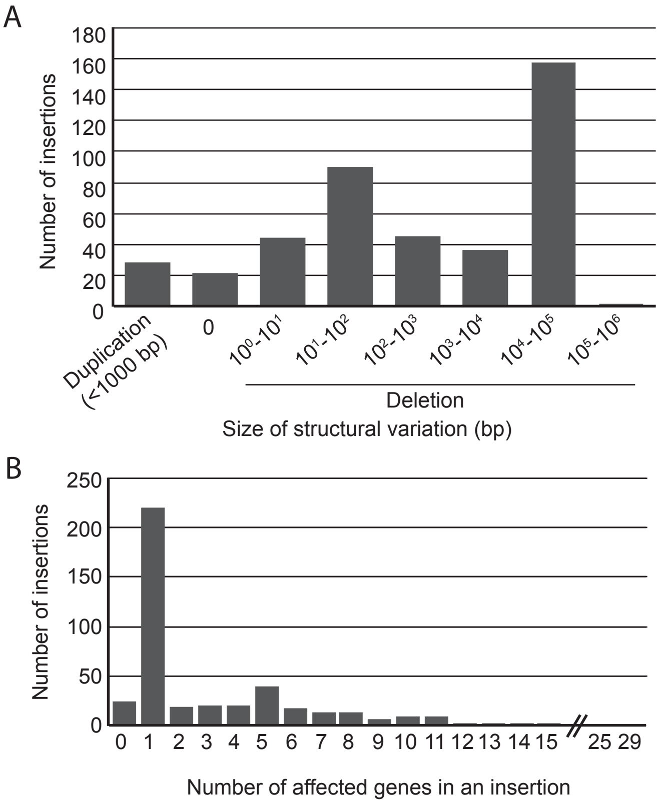


Fig 3. Structural variation accompanying insertions.
(A) Duplication and deletion sizes and (B) number of mutants grouped by the number of genes affected by two-sided insertions.
Only two-sided insertions were included in this analysis.

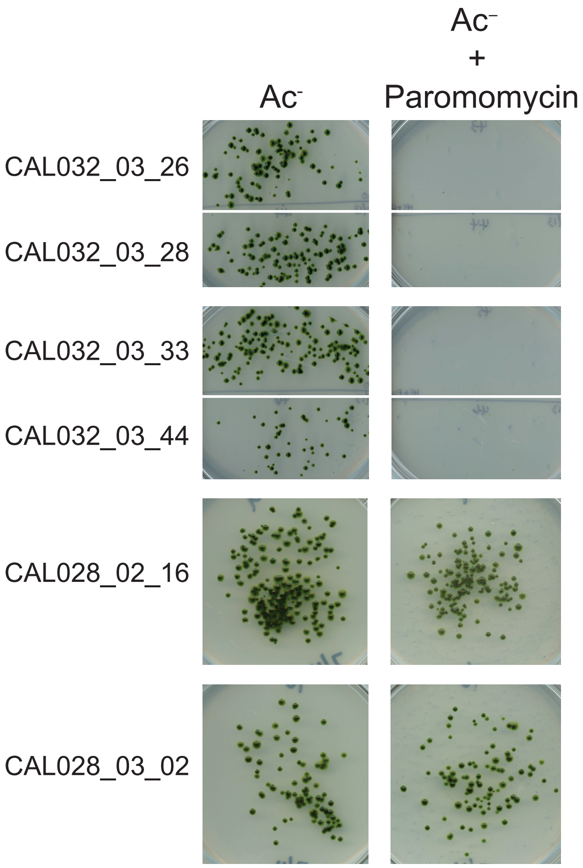
of many deletions, more than half (220 insertions) of the entire set of 425 two-sided insertions affected only a single gene (Figure 3C).

Genetic linkage between acetate-requiring phenotype and antibiotic resistance

To determine if the phenotype of ARC mutants was likely caused by the plasmid insertion, we back-crossed 89 mutants to the wild type (WT) and analyzed the genetic linkage of the acetate-requiring phenotype and antibiotic (paromomycin) resistance in the respective progenies. The acetate-requiring phenotype was closely linked to the antibiotic resistance in 88% (77 out of 88 that produced viable zygospores) of mutants that were tested (S1 Table, column “Genetic Linkage”). In each cross, approximately 100 zygospores were collected and tested for recombination between the acetate-requiring phenotype and paromomycin resistance by selecting for progeny that were able to grow on minimal medium with paromomycin (S2 Figure). The lack of recombination and therefore growth indicates that the genetic distance between the mutation causing the acetate-requiring phenotype and paromomycin resistance is less than 0.5 cM, estimated to be 50 kb on average in the *Chlamydomonas* genome (15).

Identification of secondary mutations using WGS data

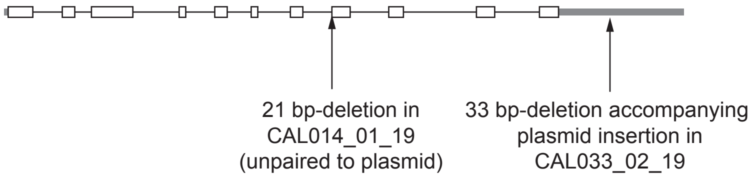
In addition to the deletions associated with plasmid insertions in the ARC mutants, we searched for and found 68 other deletions using Pindel (30) (S2 Table). The size of the deletions ranged from 20 bp to 36 kb, with a majority of them (55 deletions) being less than 100 bp (S2 Table). The deletions were visually confirmed on alignments as direct gaps in reads and/or the lack of reads within the region, depending on the size. This was



S2 Fig. Genetic linkage test of par^{R} and Ac^{-} phenotypes.
Mutants (*ac*$^{-}$ par^{R}) were crossed with WT (*AC*$^{+}$ par^{S}) cells and the zygospores were tested for growth on minimal media with and without paromomycin. Absence of growth on min+paromomycin indicates the genetic linkage of the two phenotypes.

not expected to be an exhaustive search for such deletions. For example, low-coverage regions could be difficult to distinguish from a deletion. Nevertheless, some of the deletions affected clear candidate genes that could be responsible for the mutant phenotype. For example, the CAL014_01_19 mutant was found to contain a 21-bp deletion in Cre01.g013801, a GreenCut2 gene (conserved within genomes of land plants and green algae but absent from non-photosynthetic organisms (15,31)) annotated as a tocopherol cyclase (*VTE1*). The deletion occurred at the junction of intron 7 and exon 8, which could affect splicing and translation of a functional protein (S3 Figure). Because tocopherols are important for photoprotection in *Chlamydomonas* (32) disruption in the *VTE1* gene could explain this mutant's high light-sensitive phenotype (S1 Table). In support of this hypothesis, a second mutant in the ARC, CAL033_02_19, had a 33-bp deletion in this locus. Interestingly, this mutant has a less severe phenotype (S1 Table), consistent with the plasmid insertion and deletion positioned in the 3'-UTR of the gene, which may have led to a partial loss of function (S3 Figure).

Among the 11 mutants whose acetate-requiring phenotype did not cosegregate with its paromomycin resistance, one (CAL036_02_12) had a strong acetate-requiring phenotype (S1 Table) and contained a 36-kb deletion located 2 Mb away from the plasmid insertion on chromosome 7. This resulted in a deletion of seven genes (Cre07.g346050, Cre07.g346100, Cre07.g346150, Cre07.g346200, Cre07.g346250, Cre07.g346300, and Cre07.g346317). One of these (Cre07.g346050) is *COPPER RESPONSE DEFECT 1* (*CRD1*), and *crd1* mutants have a conditional phenotype, lacking accumulation of PSI only under copper deficiency (33). Another mutant (CAL029_03_36) has a one-sided insertion in *CRD1* and was only modestly affected in



S3 Fig. Two mutant alleles in tocopherol cyclase (Cre01.g013801) in ARC.
Schematic representation of the disruption sites in CAL014_01_19 a strictly acetate-requiring mutant and CAL032_02_19, a mutant with comparatively moderate phenotype.

growth in HL (S1 Table), suggesting that the loss of CRD1 is not the cause of the severe phenotype of CAL036_02_12. Another one of the deleted genes is annotated as phytol kinase (Cre07.g346300). Chlorophyll degradation and phytol remobilization through phytol kinase (*VTE5*) and phytol phosphate kinase (*VTE6*) are important for α-tocopherol biosynthesis and their disruption results in high light sensitivity in tomato (34) and *Arabidopsis* (35). The light sensitivity observed in CAL036_02_12 is similar to that of tomato plants silenced for *VTE5* (34) and strongly suggests that Cre07.g346300 is the causative gene for the mutant phenotype. The remaining 10 mutants whose acetate-requiring phenotype is unlinked to the plasmid insertion would be candidates for higher-coverage WGS to search for causative mutations.

Genes with multiple mutant alleles in the ARC

In total, 1405 genes were directly affected by the 554 plasmid insertions in 509 mutants. There are many more affected genes compared to the number of mutants from which they originate due to disruption of multiple genes by large deletions. S3 Table lists all of the disrupted genes and their available annotations.

To begin identifying causative mutations, we searched for genes that were affected in multiple ARC mutants. Figure 4A shows the number of alleles of the 1405 genes that occur in the ARC. Interruption/deletion of 1053 genes only occurred once, while 212 genes have two alleles and 94 genes have three alleles. Some genes appeared on the list of affected genes more than three times (Figure 4A). However, because disruption of multiple genes occurred in approximately half of the ARC mutants, many of these genes represented by multiple alleles are likely not causative for the mutant

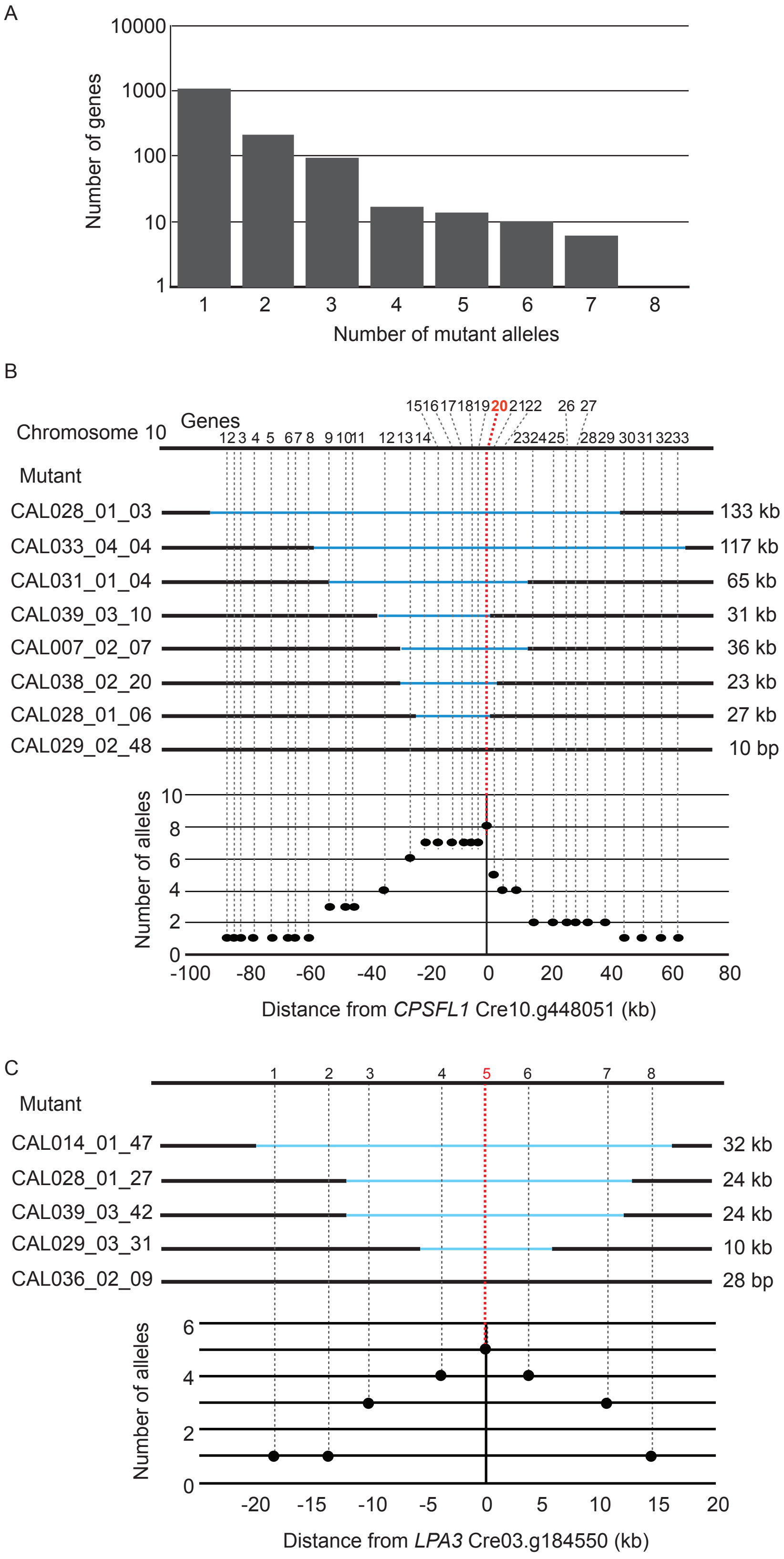


Fig 4. Genes represented by multiple mutant alleles are more likely to be causative genes. (A) Number of genes within all 1407 genes affected in ARC grouped by the number of mutant alleles that represent the gene. Schematic of mutant alleles disrupted in (B) cpsfl1 mutants and (C) lpa3 mutants, and the allele frequencies of surrounding genes. Note not all genes with multiple mutant alleles are causative, but rather occur in ARC out of their physical proximity to the true causative genes.

phenotype. Some of the genes appear more frequently on the list simply because of their proximity to the causative gene. Figure 4B shows an example of such an occurrence for *CPSFL1* (Cre10.g448051). Seven ARC mutants had deletions ranging from 22 to 130 kb in a region on chromosome 10 (CAL028_01_03, CAL033_04_04, CAL031_01_04, CAL039_03_10, CAL007_02_07, CAL038_02_20, and CAL028_01_06) (S1 Table). 33 genes were affected by the deletions in these mutants, including seven genes affected in all seven mutants, which makes it difficult to narrow down to a single causative gene. One additional mutant (CAL29_02_48) had a complex insertion event involving four different chromosomes, but strikingly it shared a single affected gene (*CPSFL1*, containing a 10-bp deletion) with the other seven mutants. All eight mutants exhibited a strict acetate-requirement and severe light-sensitivity phenotype (S1 Table), and in-depth characterization of the CAL028_01_06 mutant showed that *CPSFL1* is involved in carotenoid accumulation and is essential for photoautotrophic growth in *Chlamydomonas* and *Arabidopsis* (36,37).

The *CrLPA3* gene (Cre03.g184550, hereon *LPA3*) is another example of a gene that was affected in multiple mutants (Figure 4C). The CAL014_01_47, CAL028_01_27, CAL039_03_42, CAL029_03_31, and CAL036_02_09 mutants had overlapping deletions ranging from 28 bp to 32 kb in the same region on chromosome 3, and all five mutants exhibited a strict acetate-requiring phenotype in HL (S1 Table). By comparing the disruption frequencies, we identified *LPA3* as the only gene that was affected in all five mutants.

LPA3 and *PSBP4* are essential for photoautotrophic growth

We proceeded to validate the WGS data and identify two genes as necessary for photoautotrophic growth in *Chlamydomonas*. In one case (*LPA3*), multiple alleles were present in the ARC, whereas only a single allele of the other gene *CrPSBP4* (hereon *PSBP4*) was present. Three *lpa3* mutants (CAL028_01_27, CAL039_03_42, and CAL040_01_25) were selected for further analysis (and renamed as *lpa3-1*, *lpa3-2*, and *lpa3-3*, respectively) The WGS data indicated that the *lpa3-1* and *lpa3-2* mutants had very similar deletions of 24 kb that affected the same five genes (S1 Table). The deletion was confirmed by amplifying genomic regions across the predicted deletion by PCR in both mutants (Figure 5A), although it was not possible to amplify the plasmid sequence at the site of the deletion. The *lpa3-3* mutant was predicted from WGS to have a 4-bp deletion and plasmid insertion in the 5'-UTR of Cre03.g184550, which was confirmed by sequencing a PCR fragment of the region from the mutant (Figure 5A), but it was not included in S1 Table, because it was one of the 79 mutants with a non-unique insertion site (see above in section "Identification of insertion sites by mapping of discordant read pairs"). All three mutants had an acetate-requiring phenotype (Figure 4B). The gene Cre03.g184550 encodes a GreenCut2 protein (CPLD28) (31), and is annotated as an ortholog of *Arabidopsis* LOW PSII ACCUMULATION 3 (LPA3). *Arabidopsis* LPA3 has been reported to be involved in the assembly of photosystem II (38), although the publication on the function of this protein was later retracted (39). Complementation with a genomic DNA clone of Cre03.g184550 (*LPA3*) including 1.2 kb upstream of the transcription start site rescued all three mutants, demonstrating that the disruption of this gene was responsible for the acetate-requiring phenotype of these mutants. Mutants lacking LPA3 exhibited very low F_v/F_m values even in the dark (Figure 5C). This

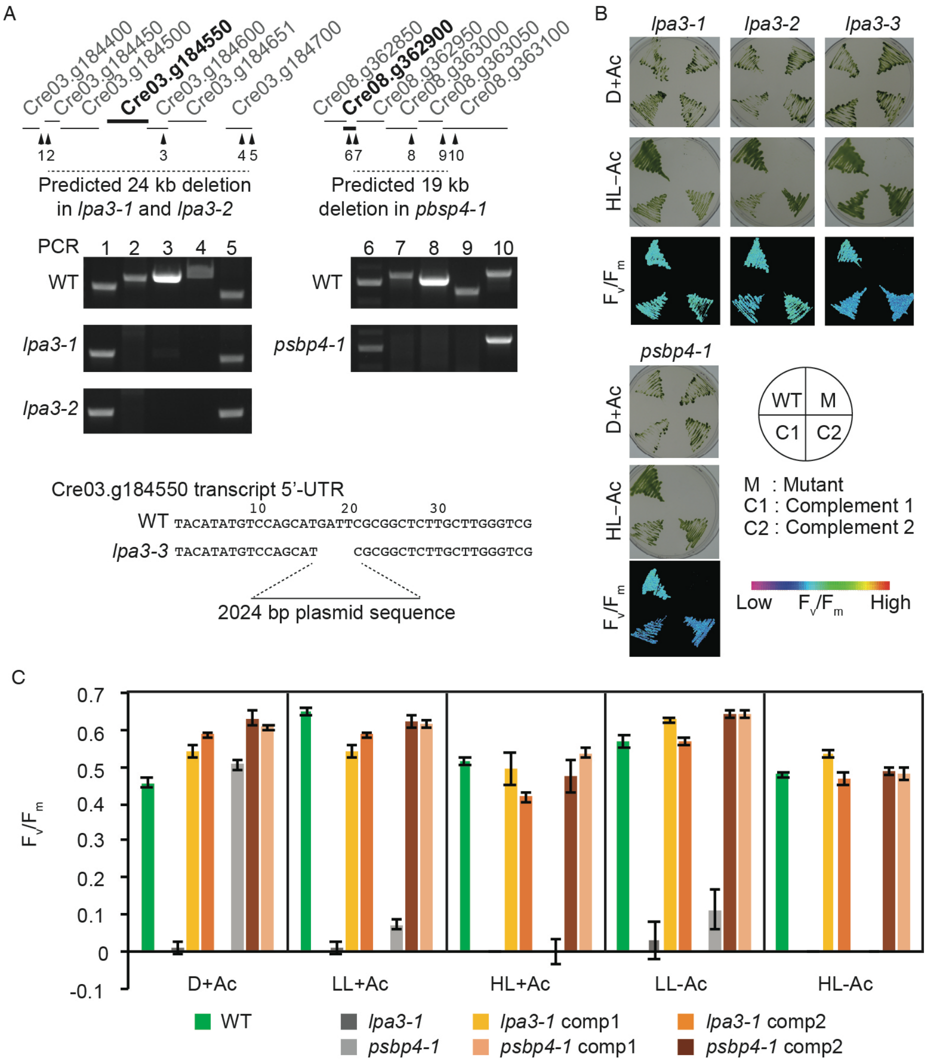


Fig 5. Identification of CrLPA3 and CrPSBP4 required for photoautotrophic growth. (A) Schematic of loci and deletions indicated from whole-genome sequence data in mutants *lpa3-1* (CAL028_01_27), *lpa3-2* (CAL039_03_42), and *lpa3-3* (CAL040_01_25) that share a disruption in Cre03.g184550, gene encoding a predicted ortholog of Arabidopsis LOW PHOTOSYSTEM II ACCUMULATION 3 (LPA3) and mutant *psbp4-1* (CAL032_04_48) that had a deletion encompassing Cre08.g362900, a gene encoding a protein predicted as PSBP4. Numbered arrowheads indicate the PCR probes used in testing for deletions shown in the agarose gel photos. WT and lpa3-3 sequences indicate the plasmid insertion site and associated 4 bp-deletion. (B) Growth and chlorophyll fluorescence phenotype of WT, mutants and their complemented lines. Cells were grown with acetate in the dark, without acetate under 400 μmol photons s^{-1} m^{-2} and imaged for growth and Fv/Fm measurements (HL-Ac). Fv/Fm value are represented by false colors as shown in the reference bar. (C) Fv/Fm values of each genotype under different growth conditions. comp, complemented line.

suggests that *Chlamydomonas* LPA3 is required for the assembly of PSII even in the absence of light, resulting in a much more severe phenotype than *lpa3* single mutants in *Arabidopsis*, which were able to grow in LL on soil (38). The low F_v/F_m phenotype of the mutants was rescued in the complemented lines in all light conditions (Figure 5B, C).

The mutant CAL032_04_48 (renamed as *psbp4-1*) required acetate for growth and exhibited light sensitivity even in the presence of acetate, and its F_v/F_m was reduced compared to that of the WT when grown in the light (S1 Table, Figure 5C). Its WGS indicated two tandem simple insertions disrupting five genes. Among them, Cre08.g362900, annotated as encoding a thylakoid luminal PsbP-like protein (PSBP4), presented itself as a clear candidate to be the gene responsible for the phenotypes. The PSBP4 ortholog of Arabidopsis has been shown to involved in the assembly of PSI (40,41). The deletion in *psbp4-1* was confirmed by PCR (Figure 5A), and the mutant phenotype was rescued by transforming with genomic DNA including Cre08.g362900 and upstream region, demonstrating that disruption of *PSBP4* was the cause of the acetate-requiring and light-sensitive phenotypes of this mutant (Figure 5B, C).

Curation of higher-confidence photosynthesis candidate genes

To identify candidate genes that are likely to be responsible for the ARC mutant phenotypes, we focused on the 406 mutants with only simple insertions (Figure 2). We reasoned that a mutant with a simple insertion event is more likely to have a causative gene within its disrupted gene list than a mutant with a complex insertion event that is accompanied by large-scale chromosomal rearrangements, which could cause unpredictable changes in expression of neighboring genes due to alterations in promoters, enhancers, and chromatin environment. For each of the 406 mutants with simple insertions, we applied a series of criteria to generate a list of genes that are the strongest confidence candidates for being genes that are responsible for the ARC mutant phenotype. If a mutant contained a single, simple insertion that disrupts a single gene then that gene was immediately considered to be a higher-confidence candidate. If a mutant contained a simple insertion with multiple genes disrupted by an associated deletion, then we manually analyzed the genes and selected the best candidate, considering whether it was a GreenCut2 gene and/or whether it encoded a protein with annotation or domains indicating a possible function in photosynthesis (e.g. redox, chlorophyll *a/b*-binding, Fe-S cluster). 78 GreenCut2 genes that were disrupted in 509 ARC mutants (Table 1) and were considered strong candidates unless there was an even stronger candidate based on functional annotation. As was shown for *cpsfl1* (Figure 4B) and *lpa3* (Figure 4C), mutants with overlapping disrupted genes were also compared to find the strongest candidate (gene with highest disruption frequency). Neighboring genes that were co-disrupted with the strongest candidates were deemed non-candidates in all the mutants. As a final criterion, we searched candidate genes derived from analysis of other existing photosynthesis mutant libraries and identified overlaps with *Chlamydomonas* genes whose disruption affected photoautotrophic growth (9), orthologous genes from the maize Photosynthetic Mutant Library (PML, http://pml.uoregon.edu/pml_table.php) (42), and orthologous genes identified from Dynamic Environmental Photosynthetic Imaging (DEPI) of *Arabidopsis* mutants (43).

We were able to identify a higher-confidence candidate gene for 348 out of 436 mutants with simple insertions. Because there were multiple alleles of 59 genes, this resulted in 273 higher-confidence candidate genes, which are shown in Table 2 (and S4 Table with additional details and references). This list includes genes known to be important for photosynthesis, photoprotection, and peripheral functions (S4 Table, Column "Inferred function from Cr and other photosynthetic organisms"). 106 gene products were predicted to be targeted to the chloroplast by protein targeting software Predalgo (https://giavap-genomes.ibpc.fr/cgi-bin/predalgodb.perl?page=main) (44), and among those, 61 were also predicted to be targeted to plastids by ChloroP (http://www.cbs.dtu.dk/services/ChloroP/) (45) (Table 2, S4 Table). 55 GreenCut2 genes are within this higher-confidence list, leaving 23 GreenCut2 genes that were not chosen because there was a stronger candidate gene (see column "Comments" in Table 2), an indication that not all GreenCut2 genes may be critical for photosynthesis. Among the 273 candidates, the photosynthetic functions of 68 genes have been previously described in *Chlamydomonas*, land plants, or cyanobacteria. This leaves 205 genes whose functions remain to be studied in context of photosynthesis, 47 of which have no annotation (S4 Table).

Discussion

We successfully used high-throughput, low-coverage WGS for the identification of plasmid insertion sites in our *Chlamydomonas* photosynthesis mutant collection (ARC). This approach has a much higher efficiency than PCR-based FST isolation. From the larger collection of 2800 mutants (7) from which ARC was derived, we recovered FSTs from only 17% of the mutants, whereas our WGS identified insertions in 509 out of 581 non-redundant ARC mutants (88% success among the population). We attribute this improvement to the fact that insertion site identification by WGS is not dependent on the intactness or sequence continuity of the inserted plasmid sequence, and therefore WGS overcomes complications such as plasmid concatemerization and loss of plasmid ends to which PCR primers need to anneal. Most importantly, it completely bypasses the need for PCR from the GC- and repeat-rich genome of *Chlamydomonas*. Even with relatively low average WGS coverage (~7x), we also identified 68 deletions that were not associated with plasmid insertions, some of which may be causative mutations for photosynthesis-related phenotypes that are unlinked to the plasmid insertion in specific mutants.

A previous study using WGS to identify DNA insertion events in *Chlamydomonas* (21) provides the most direct comparison with our results. Lin et al. (2018) analyzed paromomycin-resistant insertional mutants derived from electroporation instead of the glass bead transformation method that we used to generate either paromomycin- or zeocin-resistant mutants (9). They sequenced 20 transformants in 10 pools of two strains and verified 38 insertions, obtaining an average of 1.9 insertions per strain. In contrast, we found a total of 554 insertions in 509 mutants, resulting in a lower average of ~1.1 insertions per mutant. Lin et al. (2018) found that more than half (11 of 20) of their strains had more than one insertion event, and a larger collection of 1935 mutants derived from electroporation exhibited multiple insertions in 26% of strains (10). We found multiple insertions in 8% (43 out of 509) of the ARC mutants, suggesting that glass bead transformation of *Chlamydomonas* results in a higher frequency of single-copy insertions. Lin et al. (2018) identified one-sided insertions in ~40% of their mutants, whereas we observed only ~4% (21 out of 554 insertion events), despite the lower average WGS coverage in our study (~7x vs. ~15x). The frequency of complex rearrangements in our study (19%) was comparable to that observed by Lin et al. (25%), however, as previously noted by us and others (7,10,21,46), glass bead transformation seems to be frequently associated with larger deletions of genomic DNA at the sites of DNA insertion than electroporation, a finding that was clearly evident in our WGS data (Figure 3A).

In part because of the occurrence of larger deletions, 1405 genes were disrupted in 509 ARC mutants. As expected, this list is enriched for genes that encode proteins with annotated functions in photosynthesis and tetrapyrrole synthesis, and it includes 78 GreenCut2 genes (31). We examined the affected genes in each mutant to identify possible causative genes using several criteria, including GreenCut2 membership, existence of protein domains suggestive of a function in photosynthesis, and occurrence of multiple mutant alleles in the ARC. We also searched for overlaps with available photosynthesis mutant datasets, namely CLiP (*Chlamydomonas*), PML (maize), DEPI (*Arabidopsis*), and those found co-expressed with photosynthesis genes (*Chlamydomonas*). The CLiP collection has been used to identify mutants that are defective in photosynthetic growth in pooled cultures (9). This study identified 303 candidate photosynthesis genes. We identified 41 of those 303 genes in our list of 273 higher-confidence genes (Table 2, S4 Table). This overlap is lower than might be expected but could be explained simply by the fact that both the CLiP and ARC mutant collections are based on a total of ~60,000 insertional mutants, which is not sufficient to saturate the *Chlamydomonas* genome for mutations affecting photosynthesis. The maize PML consists of approximately 2100 photosynthesis mutants that contain 50 to 100 *Mu* transposable elements per individual. It is estimated to be a saturated collection with 3-4 mutant alleles for ~600 genes (42). The FSTs of this library were obtained with Illumina sequencing of fragmented gDNA that was enriched for the *Mu* element (22). Our higher-confidence candidate gene list overlapped with 17 genes identified from the maize PML (http://pml.uoregon.edu/photosyntheticml.html). DEPI screening of 300 *Arabidopsis* mutants affecting genes that encode chloroplast-targeted proteins (Chloroplast 2010 project, http://www.plastid.msu.edu/) identified 12 mutants with altered photosynthetic response (43). These mutants likely represent disruption in genes that are conditionally important in acclimation to changing light environments. Two of the 12 genes found through DEPI overlapped with our higher-confidence photosynthesis candidate gene list. The largest overlap (84 genes) was observed between our higher-confidence list and the group of photosynthesis-related genes defined based on co-expression analysis (47).

For two of the higher-confidence photosynthesis genes, *LPA3* and *PSBP4*, we validated the insertion-associated lesions for four of the ARC mutants and demonstrated their requirement for photoautotrophic growth (Figure 5). LPA3 is a GreenCut2 protein (CPLD28) that contains a DUF1995 domain. Insertion mutants containing large or small deletions in *LPA3* (Cre03.g184550) were acetate-requiring and exhibited a severe defect in PSII function even in the dark, as evidenced by F_v/F_m values near zero (Figure 5). Mutants affecting Cre02.g105650 and Cre10.g441650, two *Chlamydomonas* genes coding for proteins similar to *Arabidopsis* LPA2, were not found in the ARC. However, there are two additional genes encoding DUF1995 proteins in the *Chlamydomonas* genome, Cre06.g281800 and Cre08.g369000. The mutant CAL038_02_36 is disrupted in Cre06.g281800. It does not grow photoautotrophically but is able to grow in LL and HL in the presence of acetate. Interestingly, this mutant also has an F_v/F_m of zero in the dark (S1 Table). The severe phenotypes of these mutants in *Chlamydomonas* indicate nonoverlapping functions in PSII assembly of the gene products of *LPA3* and Cre06.g281800.

PSBP (encoded by *PSBP1*/*OEE2* in *Chlamydomonas*) together with PSBO and PSBQ constitute the oxygen-evolving complex (OEC) of PSII (48,49). In green algae and plants, PSBP appears to have expanded into a large family of proteins sharing similar domains beyond the canonical PSBP of the OEC. The *Chlamydomonas* genome contains 13 additional genes encoding proteins with PsbP-like domains whose individual functions are unknown. We showed that PSBP4 is required for photoautotrophic growth in *Chlamydomonas*, ruling out redundancy in its function with other PSBP-like domain-containing proteins. An *Arabidopsis* ortholog of CrPSBP4 (AT4g15510, PPD1) has been shown to play a role in PSI assembly (40,41), which is consistent with the light-sensitivity of our *psbp4-1* mutant. Two other members of the PSBP family, *PSBP3*, and *PSBP9*, were found to be disrupted in the ARC. The large family of PSBP-like domain-containing proteins is speculated to have resulted in divergence of their functions (50), and the availability of mutants in these genes should help to reveal their functions.

Of the 273 higher-confidence candidate photosynthesis genes that we curated based on WGS analysis of the ARC, only 68 have a previously demonstrated function in photosynthesis. This is similar to the results of pooled growth analysis of ~60,000 *Chlamydomonas* insertional mutants by Li et al. (2019), which revealed 303 candidate photosynthesis genes, of which only 65 have previously known roles in photosynthesis (9). Thus, 238 genes in the study of Li et al. (2019) and 205 genes in our study remain to be analyzed experimentally to determine their specific functions in photosynthesis. Moreover, the fact that only 42 genes are shared by these two sets of candidate photosynthesis genes suggests that there are still many more photosynthesis genes that remain to be identified, which highlights the enormous potential for future validation and discovery of new proteins involved in oxygenic photosynthesis.

Material and methods

Strains and culture conditions

Mutants described in this work were generated from wild-type strain 4A+ (CC-4051 in the 137c background. Cells were grown mixotrophically (ac) on Tris-acetate-phosphate (TAP) medium and photoautotrophically (min) on minimal high-salt medium (HS) medium (51) in low light (LL) of 60-80 μmol photons m^{-2} s^{-1} and high light (HL) of 350-400 μmol photons m^{-2} s^{-1}. LL and HL conditions were obtained using GE F25T8/SPX41/ECO and Sylvania F72T12/CW/VHO fluorescent bulbs, respectively.

Genomic DNA preparation and whole-genome sequencing

Chlamydomonas cultures were grown in 20 mL TAP to stationary phase, and genomic DNA was extracted using an alkaline lysis buffer (50 mM Tris-HCl (pH 8), 200 mM NaCl, 20 mM EDTA, 2% SDS, 1% PVP 40,000, 1 mg/mL Proteinase K) followed by phenol-chloroform extraction. DNA was collected, washed and eluted using DNeasy Plant mini-columns (QIAGEN). The resulting quality of the DNA was confirmed to be A_{260}/A_{280} of approximately 1.8 and A_{260}/A_{230} of >2. Plate-based DNA library preparation for Illumina sequencing was performed on the PerkinElmer Sciclone NGS robotic liquid handling system using Kapa Biosystems library preparation kit. 200 ng of sample DNA was sheared to 600 bp using a Covaris LE220 focused ultrasonicator. The sheared DNA fragments were size selected by double-SPRI, and then the selected fragments were end-repaired, A-tailed, and ligated with Illumina-compatible sequencing adaptors from IDT containing a unique molecular index barcode for each sample library. The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed with other libraries, and the pool of libraries was then prepared for sequencing on the Illumina HiSeq sequencing platform utilizing a TruSeq paired-end cluster kit, v4, and Illumina's cBot instrument to generate a clustered flow cell for sequencing. Sequencing of the flow cell was performed on the Illumina HiSeq2500 sequencer using HiSeq TruSeq SBS sequencing kits, v4, following a 2x150 indexed run recipe. The reads were aligned to the reference genome using BWA-mem. To identify plasmid insertion sites, discordant paired-end reads with one end mapping to the plasmid used for mutagenesis and the other to a chromosome location were mapped and manually validated for each mutant using Integrated Genome Viewer (IGV) (http://software.broadinstitute.org/software/igv/home). Putative structural variations unpaired to the plasmid sequence were called using a combination of BreakDancer (filtered to quality 90+) and Pindel and manually validated using IGV. Resulting genome sequences of 79 mutants were not unique (33 were duplicated, three were triplicated and one was quadruplicated). In all cases the mutants sharing similar sequences came from the same agar plate and sequencing plate, suggesting that it could be due to an error at the genome extraction step or in maintenance of the mutant strains; these mutants were not included in further analysis.

Molecular analyses of mutants by PCR and mutant complementation

Deletions predicted from genome sequences were confirmed by using PCR primers that anneal proximal to the borders and within the deletions. The insertion of the plasmid sequence accompanied by a 4 bp-deletion in *lpa3-3* was sequenced from the PCR product from the predicted region. Primers used for PCRs indicated in Figure 4 are listed in Supplemental S4 Table. For complementation of *lpa3-1*, *lpa3-2*, and *lpa3-3*, a 3531 bp genomic fragment containing the full length *CrLPA3* gene (Cre03.g184550) with 1209 bp upstream of the start codon and 719 bp downstream of the stop codon was amplified using primers Comp11F and Comp11R. This fragment was subsequently Gibson cloned into the vector pSP124S using primers PS1362 and PS1363 to inverse PCR around pSP124S. For complementation of mutant *psbp4-1*, a 3246 bp genomic fragment containing the full length *CrPSBP4* gene (Cre08.g362900), including 1209 bp upstream of the start codon and 719 bp downstream of the stop codon, was amplified using primers Comp12F and Comp12R and similarly cloned into vector pSP124S. Primer sequences are listed in supplemental S4 Table. Constructs for complementation were transformed into the respective mutants using the glass bead method (52). Colonies were selected on 10 µM zeocin TAP agar plates and screened for rescued individuals by measuring F_v/F_m as described below.

F_v/F_m measurement

Chlamydomonas strains were grown on agar plates in Dark+ac, LL-min, or HL-min, and F_v/F_m (F_m-F_o/F_m) was measured using a chlorophyll fluorescence video imager (IMAG-MAX/L, WALZ). Plates with the streaks of strains were dark-acclimated for

30 min and exposed to a pulse of saturating light (4000 µmol photons m^{-2} s^{-1}). Fluorescence images of F_m and F_o were captured during saturating pulses, and false-color images of F_v/F_m were generated.

Acknowledgments

We thank Alice Barkan for sharing the data for PML to compare with ARC higher-confidence candidate genes and Sabeeha Merchant, Masakazu Iwai, and Dhruv Patel for critical reading of the manuscript. This work was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under field work proposal 449B. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. K.K.N. is an investigator of the Howard Hughes Medical Institute.

Competing interests

The authors declare no competing interests.

Supporting information

S1 Table. Plasmid-paired and unpaired discordant sites detected in ARC by WGS and mutant phenotypes.

S2 Table. Mutants with deletions unassociated with plasmid insertion.

S3 Table. Total genes affected in ARC and their description.

S4 Table. Higher-confidence candidate genes and corresponding mutants.

S5 Table. List of PCR primers used in this study.

S1 Fig. Proportion of different types of insertions observed in ARC.

S2 Fig. Genetic linkage test of par^{R} and Ac- phenotypes.

S3 Fig. Two mutant alleles in tocopherol cyclase (Cre01.g013801, *VTE1*) in ARC.

S1 Appendix. Citations from S4 Table.

Figures and Tables

Table 1. GreenCut2 genes affected in ARC.

Table 2. Higher confidence photosynthesis candidate genes.

Figure legends

Fig 1. Growth and chlorophyll fluorescence screen pipeline.

Mutants were scored for growth on (A) D+ac, (B) LL+ac, (C) HL+ac, (D) LL+ac+zeocin, (E) LL-min, (F) HL-min. F_v/F_m values were measured on cells grown on (G) D+ac, (H) LL-min, (I) HL-min. FST, flanking sequence tag. A representative plate spotted from a 96-well plate is shown. D, dark; LL, low light; HL, high light; +ac, added acetate; min, minimal media.

Fig 2. Examples of structural variations and the frequency mutants with simple or complex insertions in ARC.

Boxes contain schematic examples of mapped reads as seen in IGV. Black box, mapped reads (concordant and discordant) against plasmid and chromosome. Blue box, examples of “Simple insertions”; Gray box, examples of “Complex insertions”. Gray box shows examples of different complex insertions that are intra- or interchromosomal rearrangements. Second from left in gray box shows a possible translocation between two chromosomes. Pie chart shows frequency of “Simple mutants” containing only simple insertions and “Complex mutants” containing complex insertions.

Fig 3. Structural variation accompanying insertions.

(A) Duplication and deletion sizes and (B) number of mutants grouped by the number of genes affected by two-sided insertions. Only two-sided insertions were included in this analysis.

Fig 4. Genes represented by multiple mutant alleles are more likely to be causative genes. (A) Number of genes among all 1405 genes affected in ARC grouped by the number of mutant alleles that represent the gene. Schematic of mutant alleles disrupted in (B) *cpsfl1* mutants and (C) *lpa3* mutants and the allele frequencies of surrounding genes. Note that not all genes with multiple mutant alleles are causative; some occur among the 1405 affected genes because of their physical proximity to the true causative genes.

Fig 5. Identification of *CrLPA3* and *CrPSBP4* required for photoautotrophic growth. (A) Schematic of loci and deletions indicated from whole-genome sequence data in mutants *lpa3-1* (CAL028_01_27), *lpa3-2* (CAL039_03_42), and *lpa3-3* (CAL040_01_25) that share a disruption in Cre03.g184550, gene encoding a predicted ortholog of Arabidopsis LOW PHOTOSYSTEM II ACCUMULATION 3 (LPA3) and mutant *psbp4-1* (CAL032_04_48) that had a deletion encompassing Cre08.g362900, a gene encoding a protein predicted as PSBP4. Numbered arrowheads indicate the PCR probes used in testing for deletions shown in the agarose gel photos. WT and *lpa3-3* sequences indicate the plasmid insertion site and associated 4 bp-deletion. (B) Growth and chlorophyll fluorescence phenotype of WT, mutants and their complemented lines. Cells were grown with acetate in the dark, without acetate under 400 µmol photons s^{-1} m^{-2} and imaged for growth and F_v/F_m measurements (HL-Ac). F_v/F_m value are represented by false colors as shown in the reference bar. (C) F_v/F_m values of each genotype under different growth conditions. comp, complemented line.

References

1. Harris EH. *Chlamydomonas* as a model organism. Annu Rev Plant Biol. 2001;52:363–406.
2. Salomé PA, Merchant SS. A series of fortunate events: Introducing Chlamydomonas as a reference organism. Vol. 31, Plant Cell. 2019. p. 1682–707.
3. Levine RP. A screening technique for photosynthetic mutants in unicellular algæ. Nature. 1960;188(4747):339–40.
4. Dent RM, Han M, Niyogi KK. Functional genomics of plant photosynthesis in the fast lane using *Chlamydomonas reinhardtii*. Vol. 6, Trends in Plant Science. 2001. p. 364–71.
5. Goodenough UW, Armstrong JJ, Levine RP. Photosynthetic Properties of *ac-31*, a Mutant Strain of *Chlamydomonas reinhardi* Devoid of Chloroplast Membrane Stacking. Plant Physiol [Internet]. 1969 Jul 1;44(7):1001 LP – 1012. Available from: http://www.plantphysiol.org/content/44/7/1001.abstract
6. Sager R, Granick S. Nutritional studies with Chlamydomonas reinhardi. Ann N Y Acad Sci. 1953;56(5):831–8.
7. Dent RM, Sharifi MN, Malnoë A, Haglund C, Calderon RH, Wakao S, et al. Large-scale insertional mutagenesis of *Chlamydomonas* supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants. Plant J. 2015;82(2):337–51.
8. Zhang R, Patena W, Armbruster U, Gang SS, Blum SR, Jonikas MC. High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA.

Plant Cell. 2014;26(4):1398–409.

9. Li X, Patena W, Fauser F, Jinkerson RE, Saroussi S, Meyer MT, et al. A genome-wide algal mutant library and functional screen identifies genes required for eukaryotic photosynthesis. Nat Genet. 2019;51(4):627–35.

10. Li X, Zhang R, Patena W, Gang SS, Blum SR, Ivanova N, et al. An indexed, mapped mutant library enables reverse genetics studies of biological processes in *Chlamydomonas reinhardtii*. Plant Cell. 2016;28(2):367–87.

11. Lin H, Miller ML, Granas DM, Dutcher SK. Whole Genome Sequencing Identifies a Deletion in Protein Phosphatase 2A That Affects Its Stability and Localization in *Chlamydomonas reinhardtii*. PLoS Genet. 2013;9(9).

12. Dutcher SK, Li L, Lin H, Meyer L, Giddings TH, Kwan AL, et al. Whole-genome sequencing to identify mutants and polymorphisms in *Chlamydomonas reinhardtii*. G3 Genes, Genomes, Genet. 2012;2(1):15–22.

13. Tulin F, Cross FR. Patching holes in the *Chlamydomonas* genome. G3 Genes, Genomes, Genet. 2016;6(7):1899–910.

14. Breker M, Lieberman K, Cross FR. Comprehensive discovery of cell-cycle-essential pathways in *Chlamydomonas reinhardtii*. Plant Cell. 2018;30(6):1178–98.

15. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science. 2007;318(5848):245–51.

16. Schierenbeck L, Ries D, Rogge K, Grewe S, Weisshaar B, Kruse O. Fast forward genetics to identify mutations causing a high light tolerant phenotype in

Chlamydomonas reinhardtii by whole-genome-sequencing. BMC Genomics. 2015;16(1).

17. Gabilly ST, Baker CR, Wakao S, Crisanto T, Guan K, Bi K, et al. Regulation of photoprotection gene expression in *Chlamydomonas* by a putative E3 ubiquitin ligase complex and a homolog of CONSTANS. Proc Natl Acad Sci U S A. 2019;116(35):17556–62.
18. Smith HE. Identifying insertion mutations by whole-genome sequencing. Biotechniques. 2011;50(2):96–7.
19. Cao Y, Rui B, Wellems DL, Li M, Chen B, Zhang D, et al. Identification of piggyBac-mediated insertions in *Plasmodium berghei* by next generation sequencing. Malar J. 2013;12(1).
20. Urban M, King R, Hassani-Pak K, Hammond-Kosack KE. Whole-genome analysis of Fusarium graminearum insertional mutants identifies virulence associated genes and unmasks untagged chromosomal deletions. BMC Genomics. 2015;16(1).
21. Lin H, Cliften PF, Dutcher SK. MAPINS, a highly efficient detection method that identifies insertional mutations and complex DNA rearrangements. Plant Physiol. 2018;178(4):1436–47.
22. Williams-Carrier R, Stiffler N, Belcher S, Kroeger T, Stern DB, Monde RA, et al. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy *Mutator* lines of maize. Plant J. 2010;63(1):167–77.
23. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, et al. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science. 2003;301(5633):653–7.

24. Strizhov N, Li Y, Rosso MG, Viehoever P, Dekker KA, Weisshaar B. High-throughput generation of sequence indexes from T-DNA mutagenized *Arabidopsis thaliana* lines. Biotechniques. 2003;35(6):1164–8.

25. Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, et al. A high-throughput *Arabidopsis* reverse genetics system. Plant Cell. 2002;14(12):2985–94.

26. Dent RM. Functional Genomics of Eukaryotic Photosynthesis Using Insertional Mutagenesis of *Chlamydomonas reinhardtii*. Plant Physiol [Internet]. 2005;137(2):545–56. Available from: https://www.ncbi.nlm.nih.gov/pubmed/15653810

27. Heinnickel ML, Alric J, Wittkopp T, Yang W, Catalanotti C, Dent R, et al. Novel thylakoid membrane GreenCut protein CPLD38 impacts accumulation of the cytochrome b_6f complex and associated regulatory processes. J Biol Chem. 2013/01/11. 2013;288(10):7024–36.

28. Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, et al. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature. 2009;

29. Calderon RH, García-Cerdán JG, Malnoë A, Cook R, Russell JJ, Gaw C, et al. A conserved rubredoxin is necessary for photosystem II accumulation in diverse oxygenic photoautotrophs. J Biol Chem. 2013;288(37):26688–96.

30. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25(21):2865–71.

31. Karpowicz SJ, Prochnik SE, Grossman AR, Merchant SS. The GreenCut2

resource, a phylogenomically derived inventory of proteins specific to the plant lineage. J Biol Chem. 2011;286(24):21427–39.

32. Li Z, Keasling JD, Niyogi KK. Overlapping photoprotective function of vitamin E and carotenoids in *Chlamydomonas*. Plant Physiol. 2011/11/15. 2012;158(1):313–23.

33. Moseley J, Quinn J, Eriksson M, Merchant S. The *Crd1* gene encodes a putative di-iron enzyme required for photosystem I accumulation in copper deficiency and hypoxia in *Chlamydomonas reinhardtii*. EMBO J. 2000;19(10):2139–51.

34. Spicher L, Almeida J, Gutbrod K, Pipitone R, Dörmann P, Glauser G, et al. Essential role for phytol kinase and tocopherol in tolerance to combined light and temperature stress in tomato. J Exp Bot. 2017;68(21–22):5845–56.

35. Dorp K Vom, Hölzl G, Plohmann C, Eisenhut M, Abraham M, Weber APM, et al. Remobilization of Phytol from Chlorophyll Degradation is Essential for Tocopherol Synthesis and Growth of Arabidopsis. Plant Cell. 2015;27(10):2846–59.

36. García-Cerdán JG, Schmid EM, Takeuchi T, McRae I, McDonald KL, Yordduangjun N, et al. Chloroplast Sec14-like 1 (CPSFL1) is essential for normal chloroplast development and affects carotenoid accumulation in *Chlamydomonas*. Proc Natl Acad Sci U S A. 2020;117(22):12452–63.

37. Hertle AP, García-Cerdán JG, Armbruster U, Shih R, Lee JJ, Wong W, et al. A Sec14 domain protein is required for photoautotrophic growth and chloroplast vesicle formation in *Arabidopsis thaliana*. Proc Natl Acad Sci U S A. 2020;117(16):9101–11.

38. Cai W, Ma J, Chi W, Zou M, Guo J, Lu C, et al. Cooperation of LPA3 and LPA2 Is Essential for Photosystem II Assembly in Arabidopsis. Plant Physiol. 2010;154(1):109–20.

39. Retraction to Cooperation of LPA3 and LPA2 Is essential for photosystem II assembly in Arabidopsis (Plant Physiol, (2010) 154, (109-120), 10.1104/pp.110.159558). Vol. 173, Plant Physiology. 2017. p. 1526.

40. Roose JL, Frankel LK, Bricker TM. The PsbP domain protein 1 functions in the assembly of lumenal domains in photosystem I. J Biol Chem. 2014;289(34):23776–85.

41. Liu J, Yang H, Lu Q, Wen X, Chen F, Peng L, et al. PSBP-DOMAIN PROTEIN1, a Nuclear-Encoded thylakoid lumenal protein, is essential for photosystem I assembly in Arabidopsis. Plant Cell. 2013;24(12):4992–5006.

42. Belcher S, Williams-Carrier R, Stiffler N, Barkan A. Large-scale genetic analysis of chloroplast biogenesis in maize. Vol. 1847, Biochimica et Biophysica Acta - Bioenergetics. 2015. p. 1004–16.

43. Cruz JA, Savage LJ, Zegarac R, Hall CC, Satoh-Cruz M, Davis GA, et al. Dynamic Environmental Photosynthetic Imaging Reveals Emergent Phenotypes. Cell Syst. 2016;2(6):365–77.

44. Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugière S, et al. Predalgo: A new subcellular localization prediction tool dedicated to green algae. In: Molecular Biology and Evolution. 2012. p. 3625–39.

45. Emanuelsson O, Nielsen H, Heijne G Von. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein

Sci. 1999;8(5):978–84.

46. Pollock S V., Mukherjee B, Bajsa-Hirschel J, Machingura MC, Mukherjee A, Grossman AR, et al. A robust protocol for efficient generation, and genomic characterization of insertional mutants of *Chlamydomonas reinhardtii*. Plant Methods. 2017;13(1).

47. Salomé PA, Merchant SS. Co-Expression Networks in the Green Alga Chlamydomonas reinhardtii Empower Gene Discovery and Functional Exploration. bioRxiv. 2020.

48. Rova M, Franzén LG, Fredriksson PO, Styring S. Photosystem II in a mutant of *Chlamydomonas reinhardtii* lacking the 23 kDa psbP protein shows increased sensitivity to photoinhibition in the absence of chloride. Photosynth Res. 1994;39(1):75–83.

49. de Vitry C, Olive J, Drapier D, Recouvreur M, Wollman FA. Posttranslational events leading to the assembly of photosystem II protein complex: a study using photosynthesis mutants from *Chlamydomonas reinhardtii*. J Cell Biol. 1989;109(3):991–1006.

50. Ifuku K, Ishihara S, Shimamoto R, Ido K, Sato F. Structure, function, and evolution of the PsbP protein family in higher plants. Vol. 98, Photosynthesis Research. 2008. p. 427–37.

51. Harris EH. The *Chlamydomonas* Sourcebook Volume1: Introduction to *Chlamydomonas* and Its Laboratory Use. Journal of Chemical Information and Modeling. 2013.

52. Kindle KL. High-frequency nuclear transformation of *Chlamydomonas*

reinhardtii. Methods Enzymol. 1998;297:27–38.

Table 1. GreenCut2 proteins within genes affected in ARC.

Gene ID	Gene name	Description	Comments
Cre01.g000850	CPLD38	Required for cyt b6f accumulation	
Cre01.g009650	BUG25	Basal body protein and putative AP2 domain transcription factor	
Cre01.g013801		Tocopherol cyclase	
Cre01.g016500		Dihydrolipoamide dehydrogenase	Not in Table 2
Cre01.g016514	DLD2	Dihydrolipoamide dehydrogenase	
Cre01.g027150		DEAD/DEAH-box helicase	
Cre01.g033763		D-Amino acid aminotransferase-like PLP-dependent enzymes superfamily	
Cre01.g033832		DEAD-box ATP-dependent RNA helicase 39	
Cre01.g043350	CAO1	Chlorophyllide a oxygenase	
Cre01.g049000		Pterin dehydratase	
Cre01.g049600	CGLD22	Expressed protein similar to ATP synthase I	
Cre02.g084350	CGLD1	Predicted protein	
Cre02.g084500		Zinc finger MYND domain containing protein 10	
Cre02.g084550	NAT10	Acyl-CoA N-acyltransferase-like protein	Not in Table 2
Cre02.g086550	CGL122	23S rRNA (adenine2503-C2)-methyltransferase	
Cre02.g105450	CGL141	F7O18.3 PROTEIN	
Cre02.g114750	CDPK5	MAP kinase activated protein kinase 5	Not in Table 2
Cre02.g120100	RBCS1	RubisCO small subunit 1, chloroplast precursor	
Cre02.g120150	RBCS2	RubisCO small subunit 2	
Cre03.g158900	DLA2	Dihydrolipoamide acetyltransferase	
Cre03.g160300	RAM1	Stress associated endoplasmic reticulum protein SERP1/RAMP4	Not in Table 2
Cre03.g173350	ANK22	Predicted protein with ankyrin repeats	Not in Table 2
Cre03.g182551	PCY1	Pre-apoplastocyanin	
Cre03.g182600	CPL1	Histone deacetylation protein Rxt3	
Cre03.g184550	CPLD28	LPA3, Predicted protein	
Cre03.g185200		Metallophosphoesterase/metallo-dependent phosphatase	
Cre05.g246800	GUN4	Tetrapyrrole-binding protein	
Cre05.g243800	CPLD45	PSB27	
Cre05.g242400	PGR5	Proton Gradient Regulation 5, Chloroplastic	
Cre05.g242000	CHLD	Magnesium chelatase subunit D	
Cre05.g238332	PSAD	Photosystem I reaction center subunit II	
Cre06.g278212	CGL46	Predicted protein	
Cre06.g280650	CGL59	Predicted protein	
Cre07.g315150	RBD1	Rubredoxin	
Cre07.g318200	CGLD34	ET and MYND domain-containing protein DDB	

Cre08.g362900	PSBP4	Lumenal PsbP-like protein	
Cre08.g372000	CGLD11	Predicted protein	
Cre08.g382300	CCB4	CGLD23 protein	
Cre09.g387000	CGL34	Predicted protein	Not in Table 2
Cre09.g394325	ELI3	Early light-inducible protein	
Cre09.g411200	TEF5	Rieske [2Fe-2S] domain containing protein	Not in Table 2
Cre10.g420350	PSAE	Photosystem I 8.1 kDa reaction center subunit IV	
Cre10.g435850	CPLD24	Predicted protein	Not in Table 2
Cre10.g440450	PSB28	Photosystem II subunit 28	
Cre10.g445100	CGL50	Predicted protein	
Cre10.g466500	CPL12	Glyoxylase family protein (yaeR)	
Cre11.g467689	PETC	Rieske iron-sulfur subunit of the cytochrome b6fcomplex, chloroplast precursor	
Cre11.g467754		Solute carrier protein, UAA transporter family	Not in Table 2
Cre11.g467700	UPD1	Uroporphyrinogen-III decarboxylase	
Cre11.g468750	CPLD48	Predicted protein	
Cre11.g469450	CGL124	Adhesion regulating molecule 110kDa cell membrane glycoprotein	
Cre12.g494000	CGL82	Predicted protein	
Cre12.g510050	CTH1	Copper target 1 protein	Not in Table 2
Cre12.g509050	PSBP3	OEE2-like protein of thylakoid lumen	
Cre12.g517700		Short-chain dehydrogenase/reductase, probably chlorophyll b reductase	
Cre12.g524300	CGL71	Predicted protein	
Cre12.g524350	HUS1	DNA damage checkpoint protein	Not in Table 2
Cre12.g554800	PRK1	Phosphoribulokinase	
Cre13.g562475		ER lumen protein retaining receptor family protein-related	Not in Table 2
Cre13.g563150	CGLD8	Predicted protein	
Cre13.g575000	CCS1	Protein required for cytochrome c synthesis/biogenesis	
Cre13.g577850		Peptidyl-prolyl cis-trans isomerase, FKBP-type	Not in Table 2
Cre13.g578650		Similar to complex I intermediate-associated protein 30	Not in Table 2
Cre13.g579550	CGL27	Predicted protein	Not in Table 2
Cre14.g618050	PLP3	Plastid lipid associated protein	Not in Table 2
Cre14.g624201		Thioredoxin-like protein CDSP32, chloroplastic	Not in Table 2
Cre16.g660000	CPLD63	GDT1-like protein 2, chloroplastic	Not in Table 2

Cre16.g665250	APE1	Thykaloid associated protein, Acclimation of Photosynthesis to Environment1	
Cre16.g666050	CPLD49	(Saccharopine) Dehydrogenase	
Cre16.g687450	CPLD54	K(+) Efflux Antiporter 3, chloroplastic (KEA3)	Not in Table 2
Cre16.g675100	CPLD53	Zinc finger protein Constans-related	
Cre16.g674950	POD2	Prolycopene isomerase / CRTISO	
Cre17.g702150	HCF164	Thioredoxin-like protin HCF164, chloroplastic	
Cre17.g702500	TAB2	PsaB RNA binding protein	
Cre17.g710800	NFU3	Iron-sulfur cluster assembly protein	
Cre17.g717350	TRI1	tRNA dimethylallyltransferase / tRNA prenyltransferase	Not in Table 2
Cre17.g717400	TRIT1	tRNA dimethylallyltransferase (miaA, TRIT1)	
Cre17.g731100	CPL14	DUF2358	

Table 2. Higher confidence photosynthesis candidate genes.

Cre ID	Gene name	Description	Subcellular localization[1]	Green Cut2[2]	Other mutant libraries[3]	Multiple candidates[4]
Cre01.g000850	CPLD38	DUF3007	C	G		
Cre01.g013801		Tocopherol cyclase	C	G		
Cre01.g016514	DLD2	Dihydrolipoyl dehydrogenase/Lipoyl dehydrogenase	C	G		
Cre01.g027150	CPLD40, HEL5	DEAD/DEAH-box helicase	C	G	Cr	
Cre01.g033763		D-Aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein	C	G		
Cre01.g033832		DEAD-box ATP-dependent RNA helicase 39	C	G		
Cre01.g043350	CAO1	Chlorophyllide a oxygenase	C	G		
Cre01.g049000	CGL31,PTD1	Pterin dehydratase	C	G	Zm	
Cre01.g049600	CGLD22	Expressed protein similar to ATP synthase I	C	G		
Cre02.g086550	CGL122	23S rRNA (adenine2503-C2)-methyltransferase (rlmN)	C	G		
Cre02.g120100	RBCS1	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 1, chloroplast precursor	C	G		
Cre02.g120150	RBCS2	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2	C	G		
Cre03.g158900	DLA2	Dihydrolipoamide acetyltransferase	C	G		
Cre03.g182551	PCY1	Pre-apoplastocyanin	C	G	Cr	
Cre03.g185200	CPL3, MPA6	Metallophosphoesterase/metallo-dependent phosphatase	C	G	Cr	M
Cre05.g238332	PSAD	Photosystem I reaction center subunit II, 20 kDa	C	G	Cr	
Cre05.g242000	CHLD	Magnesium chelatase subunit D	C	G	Cr	
Cre05.g242400		PGR5	C	G		
Cre05.g243800	CPLD45	Predicted protein	C	G	Cr	

Cre05.g246800	GUN4	Tetrapyrrole-binding protein	C	G	Zm	
Cre06.g278212	CGL46	Predicted protein	C	G		
Cre06.g280650	CGL59	Predicted protein	C	G	Cr,Zm	
Cre07.g315150	RBD1	Rubredoxin	C	G		M
Cre08.g362900	PSBP4	Lumenal PsbP-like protein	C	G	Zm	
Cre08.g372000	CGLD11	Predicted protein	C	G		
Cre08.g382300	CCB4	CGLD23 protein, required for Cyt b6 assembly	C	G	Zm	
Cre09.g394325	ELI3	Early light-inducible protein	C	G		
Cre09.g411200		Rieske domain-containing protein	C	G	At	M
Cre10.g420350	PSAE	Photosystem I 8.1 kDa reaction center subunit IV	C	G	Cr	
Cre10.g440450	PSB28	Photosystem II subunit 28	C	G		
Cre10.g445100	CGL50	Predicted protein	C	G		
Cre10.g466500	CPL12	Glyoxylase family protein (yaeR)	C	G	Cr	
Cre11.g467689	PETC	Rieske iron-sulfur subunit of the Cytochrome b6fcomplex, chloroplast precursor	C	G	Cr	
Cre11.g467700	UPD1	Uroporphyrinogen-III decarboxylase	C	G		M
Cre11.g468750	CPLD48, LPA3	Predicted protein	C	G		
Cre12.g509050	PSBP3	OEE2-like protein of thylakoid lumen	C	G		
Cre12.g524300	CGL71	Tricopentapeptide repeat, Protein O-GlcNac transferase	C	G	Cr,Zm	
Cre12.g554800	PRK1	Phosphoribulokinase	C	G	Cr	M
Cre13.g563150	CGLD8	Predicted protein	C	G	Zm	
Cre16.g665250	APE1	Thykaloid associated protein required for photosynthetic acclimation to variable light intensity	C	G		
Cre16.g666050	CPLD49, SCD1	Saccharopine dehydrogenase	C	G	Cr	M
Cre16.g675100	CrCO	Zinc finger protein CONSTANS-related	C	G		
Cre17.g702150	TRX20,HCF164	Thioredoxin-like protein HCF164, chloroplastic	C	G	Cr	
Cre17.g702500	TAB2	DUF1092, PsaB RNA binding protein	C	G	Zm	
Cre17.g710800	NFU3	Iron-sulfur cluster assembly protein	C	G		
Cre17.g731100	CPL14	Uncharacterized conserved protein	C	G		
Cre01.g018600	BAP31	B-cell receptor-associated protein 31-like	C			

Cre01.g034600		WD-40 domain	C		
Cre01.g049350		Zinc metalloprotease EGY2, chloroplastic-related	C		M
Cre01.g050500	PPR1	Pentatricopeptide repeat protein	C	Cr	M
Cre02.g076600		Peptidyl-tRNA hydrolase, PTH1 family	C	Zm	
Cre02.g087900		Mitogen-activated protein kinase kinase kinase/MLTK	C		
Cre02.g105650			C	Cr	
Cre02.g120250	CDPK7, STT7	Calcium/calmodulin-dependent protein kinase	C		
Cre02.g142146		Divinyl chlorophyllide a 8-vinyl-reductase/[4-vinyl]chlorophyllide a reductase	C	Zm	
Cre03.g145347			C		
Cre03.g149450		Ion channel pollux-related	C		
Cre03.g154550	PCR1	Pyrroline-5-carboxylate reductase	C		
Cre03.g155250			C		
Cre03.g159851		I-kappa-b-like protein IKBL	C		
Cre03.g172500	PTO2/PTOX2	Plastid terminal oxidase	C		
Cre03.g185550	SBP1	Sedoheptulose-1,7-bisphosphatase	C	Cr	
Cre03.g194200	PDH2	Pyruvate dehydrogenase E1 beta subunit	C		
Cre03.g206369		Tyrosine kinase specific for activated (GTP-bound)//Serine/Threonine protein kinase	C	Cr	
Cre03.g207153			C		
Cre03.g211633		Similar to Flagellar Associated Protein FAP165	C		M
Cre03.g213201			C		
Cre05.g232200	NDA3	Mitochondrial NADH dehydrogenase	C		
Cre05.g238322		Tryptophan--tRNA ligase/Tryptophanyl-tRNA synthetase	C		
Cre05.g238500		23S rRNA (adenine2503-C2)-methyltransferase	C		
Cre05.g241900			C		
Cre06.g259100			C	Cr	
Cre06.g262650	OPR22, TAA1	RAP domain (RAP)	C		
Cre06.g271200		NADH oxidase (H2O2-forming)	C		
Cre06.g280150	PSBP9	PsbP-like protein	C		

Cre06.g281800		Domain of unknown function (DUF1995)	C	Cr	
Cre06.g284100	RHP1	Rh protein, CO2-responsive	C		
Cre06.g284150	RHP2	Rh protein	C		
Cre07.g331450	NAT19		C		
Cre07.g344950	LHCA9	Light-harvesting protein of photosystem I	C		
Cre07.g349800			C		
Cre07.g356350	DXS1	1-Deoxy-D-xylulose 5-phosphate synthase, chloroplast precursor	C		
Cre08.g358250	MCA1	PPR repeat/Maturation/stability factor for petA mRNA	C	Zm	
Cre08.g358350	TDA1, OPR34	FAST Leu-rich domain-containing	C	Cr	M
Cre08.g361250		Protein O-GlcNAc transferase/OGTase (DUF563)	C		
Cre09.g388356	TBC2	Translation factor for chloroplast psbC mRNA/Translation factor for chloroplast PsbC mRNA	C	Cr	M
Cre09.g390060			C	Cr	
Cre09.g392729		Methionyl-tRNA formyltransferase/transformylase	C		
Cre09.g394150	RAA1	FAST kinase-like protein, subdomain 1	C	Cr	
Cre09.g398919			C		
Cre10.g417750		Neuropathy target esterase/Swiss cheese D.melanogaster	C		M
Cre10.g419900			C		
Cre10.g421150		Glycosyltransferase 14 Family Member	C		
Cre10.g431950		Dual-specificity kinase	C		
Cre10.g448950		Endonuclease/Exonuclease/Phosphatase family	C	Cr	
Cre10.g452800	LCIB	Low-CO2-inducible protein	C	Cr	
Cre11.g467712		Structural maintenance of chromosomes SMC family member	C	Cr	
Cre11.g476100			C	Cr	M
Cre11.g477625	(CHLH2)	Magnesium chelatase subunit H	C	Zm	
Cre12.g486750			C		
Cre12.g487500	CGL61, NYE1	Stay green 1 protein, predicted protein	C		
Cre12.g494550	RNP10	RNA binding protein	C		M
Cre12.g496250			C		

Cre12.g508850	GST8	Glutathione S-transferase, GST, superfamily, GST domain containing	C		
Cre12.g510650	FBP1	Fructose-1,6-bisphosphatase	C	Cr	
Cre12.g510750			C		
Cre12.g517681			C	Cr	
Cre12.g522000			C		M
Cre12.g524250			C	Cr	
Cre12.g531050	RAA3	PsaA mRNA maturation factor 3	C	Cr	
Cre12.g538650	HEM4	Uroporphyrinogen-III synthase	C		
Cre12.g549500		Pyrimidodiazepine synthase	C		
Cre13.g569700			C	Cr	M
Cre13.g573000		Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit N-methyltransferase I-related	C		
Cre13.g574200	PAP2	Poly(A) polymerase/Topoisomerase related protein	C		
Cre13.g578750	TBA1	PsbA translation factor	C		
Cre13.g580650		Serine/Threonine-protein phosphatase 2A activator (PPP2R4, PTPA)	C		
Cre13.g580850		Chloroplast 50S ribosomal protein L22-related	C		
Cre13.g584950			C		
Cre14.g621650		Malonyl-CoA acyl carrier protein transacylase (fabD)	C		
Cre14.g624350	VTE6	MPBQ/MSBQ methyltransferase	C		
Cre16.g658950			C	Cr	
Cre16.g662150	CCB1, CPLD51	CPLD51 protein, required for Cyt b6 assembly	C		M
Cre16.g665800	SSS4	Soluble starch synthase	C		
Cre16.g670754		Voltage and ligand gated potassium channel	C		
Cre16.g677050		Adenylate and guanylate cyclase catalytic domain//Bacterial extracellular solute-binding protein	C		
Cre16.g684250			C		
Cre16.g684300		3-Hydroxyisobutyrate dehydrogenase-related	C	Zm	
Cre16.g684900			C		M
Cre16.g686510			C		
Cre16.g687966	FAP5	Tetratricopeptide repeat, Flagellar associated protein	C		

Cre16.g689150	SQD3	Sulfolipid synthase	C			
Cre16.g692228	MARS1	Serine/Threonine protein kinase	C		Cr	
Cre17.g704000		Polyvinyl-alcohol oxidase/PVA oxidase	C			
Cre17.g719450		Ca2+/calmodulin-dependent protein kinase, EF-Hand protein superfamily//Serine/threonine protein kinase	C			
Cre17.g724600	PAO2	Pheophorbide a oxygenase, Rieske iron-sulfur cluster protein	C			
Cre17.g724700	PAO1	Pheophorbide a oxygenase, Rieske iron-sulfur cluster protein	C			
Cre17.g734548	PPD2	Pyruvate phosphate dikinase, chloroplastic	C		Zm	
Cre12.g509001	RPK2	Mitogen-activated protein kinase	n/a		Cr	
Cre01.g009650	BUG25	Basal body protein and putative AP2 domain transcription factor	O	G		M
Cre02.g084350	CGLD1	Predicted protein (GDT1 like protein 1, chloroplastic)	O	G		
Cre03.g184550	CPLD28, LPA3	Predicted protein	O	G		
Cre07.g318200	CGLD34	SET and MYND domain containing protein DDB	O	G		
Cre11.g469450	CGL124	Adhesion regulating molecule 1 110 kDa cell membrane glycoprotein	O	G		
Cre12.g494000	CGL82	Predicted protein/BRCA1-associated protein	O	G		
Cre12.g517700	NYC1, SDR21	Short-chain dehydrogenase/reductase, probably chlorophyll b reductase	O	G		
Cre13.g575000	CCS1	Protein required for Cytochrome c synthesis/biogenesis, chloroplastic	O	G	Zm	
Cre17.g717400	miaA, TRIT1	tRNA dimethylallyltransferase	O	G		M
Cre01.g016570		Mitogen-activated protein kinase kinase kinase 19	O			
Cre01.g019700	PAP7	Non-canonical poly(A) polymerase	O			
Cre01.g030700	PTK14	Protein tyrosine kinase	O			
Cre01.g032450	GLG1	Golgi apparatus protein 1	O			
Cre01.g033450		Sphingomyelin phosphodiesterase 2	O			
Cre01.g040150		WNK lysine deficient protein kinase (WNK, PRKWNK)	O			
Cre01.g043850		Serine/Threonine protein kinase	O			
Cre01.g044850		Sacsin (SACS)	O			
Cre01.g053900	NGLY1, PNG1	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase	O			
Cre02.g080700	BIP1	Endoplasmic reticulum associated HSP70 protein	O			
Cre02.g084250	PPP7	Protein phosphatase 1K, mitochondrial	O			

Cre02.g088650		Phosphatidylinositol N-acetylglucosaminyltransferase/glucosaminyltransferase	O	
Cre02.g099601		Androgen induced inhibitor of proliferation AS3/PDS5-related	O	
Cre02.g099850	PDC2	Pyruvate dehydrogenase, E1 component, alpha subunit	O	
Cre02.g100300		Phosphatidylinositol 3-kinase-related protein kinase	O	
Cre02.g106250	LAL2	La-like RNA-binding protein	O	
Cre02.g110500			O	
Cre02.g142750			O	
Cre02.g143400		3',5'-cyclic-nucleotide phosphodiesterase	O	
Cre03.g145387	FAP239	Flagellar associated protein	O	
Cre03.g145987			O	
Cre03.g156150		ATP-dependent RNA helicase DDX10/DBP4	O	
Cre03.g160250			O	
Cre03.g160400	SAC1	Sulfur acclimation 1 protein, sodium/sulfate co-transporter	O	M
Cre03.g164900		Serine/Threonine protein kinase OSR1	O	
Cre03.g168100			O	
Cre03.g173600		Ubiquitin and ubiquitin-like proteins	O	
Cre03.g175700		CobW-related	O	
Cre03.g179650		BTB/POZ domain (BTB)	O	
Cre03.g182550	PNO3	Ferredoxin-NAD(+) reductase	O	
Cre03.g182900		PNAS-related	O	
Cre03.g197450		Winged helix dna-binding domain-containing protein	O	
Cre03.g199250	CYG51	Adenylate/guanylate cyclase	O	
Cre03.g207400		von Willebrand factor type A domain	O	
Cre03.g209505		Serine/Threonine-protein kinase SRK2	O	
Cre03.g210961		Phosphatidylinositol transfer protein PDR16-related	O	
Cre04.g212401		Baculoviral IAP repeat-containing protein 6 (apollon) (BIRC6, BRUCE)	O	
Cre05.g232150	GDH2	Glutamate dehydrogenase	O	
Cre05.g245550	PIK1	Phosphatidylinositol 4-kinase	O	

Cre06.g264100			O	
Cre06.g268750	MME1	Malate dehydrogenase, decarboxylating	O	
Cre06.g278094	ELG14	Exostosin-like glycosyltransferase	O	
Cre06.g280050	XRN1	Single-stranded RNA 5'->3' exonuclease	O	
Cre06.g281250	CFA1	Cyclopropane fatty acid synthase	O	
Cre06.g282300			O	M
Cre06.g289600			O	
Cre06.g300250	TTL10	Tubulin polyglutamylase TTLL2	O	
Cre06.g302305			O	
Cre06.g308100		Enoyl-CoA hydratase 2/ECH2	O	
Cre06.g308150	DNJ23	DnaJ-like protein	O	
Cre07.g336150			O	
Cre07.g342920		Xaa-Pro dipeptidase/X-Pro dipeptidase	O	
Cre07.g348550	TGL13	Protein T08B1.4, Isoform B-related (lipase related)	O	
Cre07.g355750		F-box and WD40 domain protein	O	
Cre07.g356450		Leucine-rich repeat-containing protein	O	
Cre07.g357876			O	
Cre08.g359100		tRNA (guanine(10)-N(2))-methyltransferase	O	
Cre08.g365200			O	
Cre08.g365550			O	
Cre08.g370550		D-2-Hydroxyglutarate dehydrogenase	O	
Cre08.g375000		Actin-fragmin kinase, catalytic	O	
Cre08.g382515		WD repeat-containing protein 26	O	
Cre08.g385300		ET and MYND domain-containing protein DDB	O	
Cre09.g386450			O	
Cre09.g391356		Mitogen-activated protein kinase kinase kinase/MLTK	O	
Cre09.g393136		Clathrin assembly protein	O	
Cre09.g397956	FAP201	Flagellar associated protein (Exotosin family)	O	
Cre09.g399650			O	

Cre09.g410000		DC12-Related	O		
Cre10.g419250			O		
Cre10.g420537		Sphingomyelin phosphodiesterase 2	O		
Cre10.g427950		Leucine-rich repeat-containing protein	O		
Cre10.g429400	MCG1	FAST Leu-rich domain-containing, stabilize petG mRNA	O	Cr	
Cre10.g429601		Cell death-related nuclease 2	O		
Cre10.g433350		Squamosa promoter-binding-like protein 10-related	O		
Cre10.g433900		E3 ubiquitin-protein ligase HUWE1 (HUWE1, MULE, ARF-BP1)	O		
Cre10.g448051		Sec14p-like phosphatidylinositol transfer family protein	O		M
Cre10.g457900			O		
Cre11.g467644	CLPB1	ClpB chaperone, Hsp100 family ClpB chaperone, Hsp100 family	O		M
Cre11.g467690		Glutathione transferase/S-(hydroxyalkyl)glutathione lyase	O		
Cre12.g483650		Serine/Threonine-protein kinase STN7, chloroplastic	O		M
Cre12.g494350		Endomembrane family protein 70	O		
Cre12.g499500	SAC3	Sulfur acclimation protein, Snf1-like Ser/Thr protein kinase	O		M
Cre12.g502000	FAP253	Flagellar associated protein	O		
Cre12.g510034		Tetratricopeptide repeat protein 33, Osmosis responsive factor	O		
Cre12.g511400		Cyclin-related protein with PPR domain	O	Zm,At	M
Cre12.g511650		Auxilin/cyclin G-associated kinase-related	O		
Cre12.g524500	RMT2	Rubisco small subunit N-methyltransferase	O	Cr	
Cre12.g524700		Pyrimidine and pyridine-specific 5'-nucleotidase (SDT1)	O	Zm	
Cre12.g527600		Polyglutamine-binding protein 1 (PQBP1, NPW38)	O		
Cre12.g528250		WASP-interacting protein VRP1/WIP, contains WH2 domain	O		
Cre12.g543100		tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit (TRM6, GCD10)	O		
Cre12.g549050	STR1	Strictosidine synthase	O		
Cre12.g559050		BCDNA, fatty acid metabolism, transport	O		
Cre13.g579450	CST1	Chlamydomonas-specific membrane transporter of unknown function	O		
Cre13.g583650		Non-specific Serine/Threonine protein kinase/Threonine-specific protein kinase	O		

Cre13.g584350			O		
Cre13.g586750		Transportin 3 and Importin 13	O	Cr	M
Cre13.g588650			O		
Cre13.g605650		Betaine aldehyde dehydrogenase/oxidase	O		
Cre13.g607000		Cytosol nonspecific dipeptidase/Prolylglycine dipeptidase	O		
Cre14.g608652			O		
Cre15.g635450			O		
Cre16.g656000		Sphingomyelin phosphodiesterase 2	O		
Cre16.g656200		IQ calmodulin-binding motif (IQ)//Tetratricopeptide repeat (TPR_12)	O		
Cre16.g657979		Kinesin Family Member C2/C3	O		
Cre16.g661250		Thioredoxin peroxidase	O		
Cre16.g663050		Guanylate-binding family protein	O		
Cre16.g663600		MFS transporter, ACS family, solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter)	O		
Cre16.g665400		Small nuclear ribonucleoprotein SmD1	O		
Cre16.g666150	ODA1	Flagellar outer dynein arm-docking complex protein 2	O		
Cre16.g668700			O	Cr	M
Cre16.g678808		U4/U6 small nuclear ribonucleoprotein Prp4 (contains WD40 repeats)	O		
Cre16.g679950	RFC3	DNA replication factor C complex subunit 3	O		
Cre16.g682100		Tropinone reductase I	O		
Cre16.g687500	ARP2	Actin-related protein	O		
Cre17.g704350		Glyoxalase domain-containing protein 4	O		M
Cre17.g711150	FAD2	omega-6 Fatty acid desaturase (delta-12 desaturase)	O		
Cre17.g712850	TRX23	Thiol-disulfide isomerase and thioredoxin	O	Cr	M
Cre17.g721350	GST13	Glutathione S-transferase	O		
Cre17.g721950		E3 UBIQUITIN-PROTEIN LIGASE ARI2-RELATED	O		
Cre17.g722300			O		
Cre17.g725750	SSA2	60 kDa SS-A/Ro ribonucleoprotein	O		
Cre17.g728800	IDH1	Isocitrate dehydrogenase, NAD-dependent	O		

Cre17.g742400	PTK17	Protein tyrosine kinase	O	M

1 C, predicted to be chloroplast targeted by Predalgo or ChloroP; O, other; n/a, not analyzed.

2 G, GreenCut2.

3 Identified in other photosynthesis mutant library studies Chlamydomonas (Cr), Maize (Zm), Arabidopsis (At).

4 M, Multiple strong candidates in this mutant. See S4 Table for further detail.