

Isolation and molecular characterisation of *Dunaliella tertiolecta* with truncated light-harvesting antenna for enhanced photosynthetic efficiency.

Johansson, AS^{1,a}; Stephenson, PG¹; Edwards, RJ^{3,5,b}; Yoshida, K^{2,c}; Moore, CM¹; Terauchi, R^{2,d}; Zubkov, MV^{4,e}; Terry, MJ^{3,5}; Bibby, TS^{1,5#}

Affiliations: ¹School of Ocean and Earth Science, University of Southampton, Southampton, SO14 3ZH, UK. ²Iwate Biotechnology Research Centre, Japan. ³School of Biological Sciences, University of Southampton, UK. ⁴National Oceanography Centre, Southampton, SO14 3ZH, UK. ⁵Institute for Life Sciences, University of Southampton, UK. Current affiliation: ^aEMBL Heidelberg 69117, Germany. ^bUNSW Sydney, Australia. ^cGraduate School of Agricultural Science, Kobe University, Japan. ^dGraduate School of Agriculture, Kyoto University, Japan and ^eScottish Association for Marine Science.

#Correspondence: tsb@noc.soton.ac.uk

Abstract

Here we report the development of a high-throughput selection protocol using random mutagenesis and live single-cell sorting to isolate cell lines from the algae *Dunaliella tertiolecta* with reduced chlorophyll content, with the aim to optimise the antenna size for increased photosynthetic efficiency. Two promising cell lines (*Ica1* and *Ica2*) have been isolated that display a truncated light-harvesting antenna, and hence improved photosynthetic energy conversion efficiency by increasing the light intensity at which photosynthesis becomes saturated (I_s). *Ica1* and *Ica2* differ significantly: the *Ica2* phenotype retains an ability to alter its antenna size in response to varying light intensity, whereas *Ica1* appears to have lost this ability and is 'locked' to a truncated antenna and high-light phenotype. Despite these clear differences, transcriptomic analysis shows that the expression profiles for differentially expressed nuclear-encoded photosynthetic genes is similar in both *Ica1* and *Ica2*, possibly suggesting underlying mutations in the regulation of photosynthesis are causing the observed changes in phenotype rather than mutations impacting specific components of the photosynthetic apparatus. The combination of approaches presented here offer the capacity to substantially improve photosynthetic efficiency from any microalgal species irrespective of the extent to which it has been characterised genetically or the availability of molecular tools for rational engineering. It thus offers the potential to begin to exploit the huge natural diversity of microalgae for enhanced biomass production.

Keywords: Microalgae, Photosynthesis, Fluorescence-activated cell sorting (FACS), Light Harvesting, Chlorophyll, Transcriptomics

Introduction

The huge natural diversity of single-celled algae (microalgae) offers a massive potential for the generation of an array of sustainable products [1–4]. Realising the potential of microalgae as sustainable sources of both natural and synthetic products, relies on an ability to rapidly select and improve extant cell lines. Partially, this has been limited by: (1) the small number of published genomes, which have grown only slowly from 10 fully sequenced eukaryotic microalgal genomes in 2010 [5], to 28 listed as published as of April 2019 by the U.S. Department of Energy Joint Genome Institute [6]; and (2) the limited (although rapidly developing) transformation protocols and genetic tools for targeted genetic manipulation in relevant species [7–11]. However, techniques involving selection and characterisation of cells from liquid culture such as fluorescent-activated cell sorting (FACS¹) combined with random mutagenesis approaches offer the ability to rapidly improve cell lines from any algal source, irrespective of the availability of developed genetic tools [12–15] or published genomes or requirement for growth on plates. Moreover, compared to targeted approaches, random approaches have the advantage that resultant cell lines are not classed as genetically modified (GM), which may be considered commercially undesirable [16]. Linking these random approaches with high-throughput ‘*omic*’ technologies enables

¹ Abbreviations: FACS, fluorescent-activated cell sorting; FRRf, fast repetition rate fluorometer; *lca*, low chlorophyll antenna; ETR, electron transport rate; NPQ, non-photochemical quenching; MDS, multi-dimensional scaling; SDE significantly differentially expressed; GO, gene ontology; OEC, oxygen evolving complex

characterisation of diverse mutations and the potential to identify '*cryptic*' genes that are important in the regulation of a phenotype of interest but which are currently unconsidered by rational targeted approaches. Such techniques can be applied specifically to marine microalgae species, which are of interest as they: (1) do not place demands on fresh water resources when grown on large scales; and (2) have huge natural genetic potential owing to the highly diverse marine microalgae species that exploit every sunlit niche in the marine environment.

Optimizing the efficiency of photosynthesis is central to the commercial realisation of any developing biotechnological application using microalgae, irrespective of the selected species and/or target product [17, 18]. To a first order, the photosynthetic energy conversion efficiency (the efficiency at which absorbed photons are converted into biomass) is the primary bottleneck for downstream energy supply and has been identified as a fundamental target for the commercial realisation of the myriad of potential biotechnological applications of microalgae [19–22]. Photosynthetic energy conversion efficiency is affected by how well cells absorb photons and how effectively this potential is stored in reduced carbon bonds. A significant loss term in overall efficiency can result from the over-absorption of photons by light-harvesting pigment-binding antenna complexes, which serve to transfer excitation energy to reaction centres that are the sites of the initial photochemical reactions of photosynthesis [23, 24]. In natural environments, cells are continually mixed in a water column and exposed to variations in light intensities that can span over many orders of magnitude, the photosynthetic apparatus is then regulated to some integrative characteristic of the light intensity the cell experiences [25, 26]. When exposed to supra-optimal light, cells

absorb more energy than the downstream processes can handle and dissipate this excess excitation energy as fluorescence or heat. The resulting loss of potential can be significant (up to 90% of the absorbed light) [17] and, when considering cells grown in mass culture, results in the rapid attenuation of available light. This, in turn, causes light-limitation at depth, through self-shading within the culture. Therefore, cell lines selected for a small light-harvesting system will theoretically be more productive than wild-type (WT) when grown in mass culture. Targeting this characteristic thus has the potential to improve the productivity of any cell line selected for growth rate, robustness, lipid content etc. Previous work has shown improvements in productivity (both biomass and hydrogen) of cell lines selected for small light-harvesting antenna size using both random and targeted genetic approaches [27]. In 2002 it was shown that a *C. reinhardtii* mutant (*alb3*) had strongly reduced LHCs and reduced amounts of PSII [28], while the *stm3* mutant was shown to have increased levels of chlorophyll per cell [29]. The first truncated light-harvesting chlorophyll antenna (*tla*) strain, *tla1*, generated in *C. reinhardtii* using random DNA insertion followed by screening for low chlorophyll fluorescence or high Chl a:b ratio was reported in 2003 [24], followed by *tla2* and *tla3* in 2012 [30, 31], all three strains were shown to have higher photosynthetic activity at saturating light intensities or above in WT.

In this paper, we report the application of a high-throughput pipeline and characterisation protocol for the selection of random mutants from *Dunaliella tertiolecta* (a motile, single-celled green algae, lacking a cell wall, for which there is currently no published genome). *D. tertiolecta* is a suitable candidate for commercial algal biotechnology development as it is halo-tolerant and grows well in a wide range of

conditions, helping to reduce the impact on freshwater supplies and the environment [32]. Here we isolated and characterised two new cell lines of *D. tertiolecta* with truncated light-harvesting antenna. We complemented this approach with next-generation transcriptomics and bioinformatics to characterise the molecular basis of gene regulation that determines the selected phenotype. This approach, that does not rely on the availability of a published genome or require growth on plates, provides a protocol that can potentially be applied to the selection and improvement of any microalgae, enabling the development of previously poorly characterised marine microalgae and enabling exploitation of the huge genetic potential these cells lines offer.

Results

Isolation of cell lines with reduced pigment content

Previous studies have established a linear relationship between the functional light harvesting antenna size and the light saturated rate of photosynthesis (P_{max}), such that strains with truncated antenna typically have a higher maximum photosynthetic efficiency per unit pigment [23, 33]. Further, cell lines with truncated antennae have previously been shown to have an overall reduced chlorophyll content [24, 34]. It therefore seemed reasonable to assume that selection for cells with a reduced total chlorophyll content could be exploited as a proxy for cells with truncated antennae and increased photosynthetic efficiency.

WT *D. tertiolecta* were exposed to chemical mutagenesis and then screened using flow cytometry (see methods), resulting in 3502 sorted cells with reduced chlorophyll content (based on chlorophyll autofluorescence) without a significant change in cell size (SI Figure 1). Cultures that maintained WT growth rate at 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ while displaying a reduced chlorophyll fluorescence (SI Figure 2) were retained and subjected to in-depth photophysiological characterisation using fast-repetition-rate fluorometry (FRRf). FRRf was used as a real-time *in vivo* tool to measure both the functional size of the light-harvesting antenna associated with photosystem II (σPSII), thus confirming that the reduced chlorophyll content had resulted in a truncated antenna of PSII while the apparent photosynthetic efficiency of PSII (F_v/F_m) was maintained (SI Table 1). From this analysis, two cell lines were successfully established that showed a dramatic reduction in σPSII . These lines, termed *low chlorophyll antenna 1 (lca1)* and *lca2*, displayed σPSII values that were 56.7% and 66.6% of WT, respectively.

Physiological characterisation of *lca1* and *lca2*

A detailed analysis of *lca1* and *lca2* compared to WT cells grown in batch culture under varying light levels (high 1000, medium 100 and low 20 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) was conducted to assess the physiology of the newly isolated cell lines. Growth rate (μ) increased with irradiance in all three cell lines and there was no significant difference between cell lines at any specific light intensity (Figure 1A and Table 1) although *lca1* showed a reduced growth rate under low light possibly reflecting the lower chlorophyll content in this cell-line. The photosynthetic efficiency (F_v/F_m) decreased with irradiance but was comparable to WT under most light levels (a significantly higher F_v/F_m was

observed in *lca1* under medium light) (Figure 1B, Table 1). In contrast, in *lca1*, both chlorophyll *a* per cell and σ PSII were significantly smaller than WT at all light levels (Figure 1C-D and Table 1). σ PSII of *lca2* was smaller than that of WT at medium- and high-light levels, but comparable to WT at low light; chlorophyll per cell was only significantly less than WT at low light (Figure 1C-D and Table 1). Measurements of the rate of electron transport from PSII and the rate of oxygen evolution for medium light grown cells over an imposed light gradient (0 - 2000 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) provides further insights into the physiology of photosynthesis. In both *lca1* and *lca2*, the average rate of oxygen evolution normalised to chlorophyll *a* (Figure 2A) and the electron transport rate (ETR) per reaction centre (Figure 2B) were comparable to or higher than in WT. The calculated maximum rate of photosynthesis (P_{max}) for chlorophyll normalised oxygen evolution and PSII ETR were both higher than WT for both *lca1* and *lca2*, with values significantly higher in *lca1* (Table 2). Correspondingly the calculated light intensity at which photosynthesis was saturated (I_s) was significantly higher than WT in *lca1*, while the initial slope of the curve (α) was significantly lower in both *lca1* and *lca2* for ETR (Table 2). Both WT and *lca2* displayed a drop in both the minimum (F'_o) and maximum fluorescence (F'_m) and a reduction in the functional cross section (σ PSII), (Figure 2C-E) with substantial non-photochemical quenching (NPQ) (Figure 2F) when exposed to actinic light above saturating levels over the duration of the progressive light exposure experiments (~45 min). Combined with the calculated ETR, this suggests activation of an NPQ mechanism alongside the increased closure of active PSII at an irradiance around 1000 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$ in both strains. No similar response was observed in *lca1*, with F'_o and F'_m as well as σ PSII remaining near constant at an irradiance above 100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$.

Consequently, while ETR plateaued out in both WT and *lca2* at around 1000 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, it continued to increase in *lca1* throughout the imposed light gradient, resulting in the significantly higher observed P_{max} values (Table 2).

Pigment composition and protein abundance

Accessory pigment abundance normalised to chlorophyll *a* (Figure 3A) revealed the presence of significantly increased amounts of violaxanthin in *lca1* under low (187%) and medium (481%) light compared to WT, while some other pigments were below detection limits (especially in *lca1* grown under low light). The amount of chlorophyll *b* per cell was significantly reduced in *lca1* and *lca2* compared to WT at all light levels, and could not be detected at all in *lca1* under low light (Figure 3B). The chlorophyll *a/b* ratio was therefore significantly higher than WT in both *lca1* and *lca2* under medium (8.82 ± 1.10 and 9.94 ± 0.12 vs 4.55 ± 0.43) and high light (8.91 ± 0.06 and 12.6 ± 1.55 vs 4.43 ± 0.73).

The abundance of core photosynthetic complexes PSII, PSI and Rubisco was estimated using Western blots and antibodies against the PSBA, PSAC and RBCL peptides in cultures acclimated to medium light [35, 36]. The results (SI Figure 3A-B) showed a significant (ANOVA, Tukey HSD, $p < 0.01$) reduction of PSAC relative to WT in both *lca1* (23.0%) and *lca2* (61.7%), possibly suggesting greater capacity for cyclic electron flow in this strain. The level of PSBA relative to WT was comparable for *lca1* (99.2%) and reduced in *lca2* (72.1); the difference was not significant (ANOVA, Tukey HSD, $p > 0.05$) in either strain. The amount of RBCL was not significantly (ANOVA, Tukey HSD, $p > 0.05$) reduced in either *lca1* or *lca2* relative to WT.

Transcriptome analysis

Transcriptomes in the form of 27 Super-SAGE libraries (see methods) from triplicate WT, *Ica1* and *Ica2* cultures grown under three different light levels, were analysed for differential expression and gene enrichment. The use of SuperSAGE enabled multiplexing of all libraries within a single sequencing run with maintained precision in terms of transcript representation [37]. The short-read SAGE libraries consisting of 26 base pair (bp) mRNA sequences representing 382846 unique tags (unitags) from 32331173 sequences (tags), with an average of approximately 1.2 M tags per library, were mapped against previously published expressed sequence tags (ESTs) from *D. tertiolecta* [56] to extend the effective read-length. These extended reads were annotated and filtered to include only nuclear encoded genes, resulting in a final set of 9065 unitags with extended annotations that contained 16.6% of the original tags. Further filtering of this set for explicit high-confidence annotations with tag counts passing the cut-off filters for the statistical analysis (> 1 count per million in 9 libraries) resulted in a small list of 1312 unitags covering 14.4% of the total reads, which were in turn used for the higher-level analysis and to assign gene functions. A subset of genes with functions related to photosynthetic processes were selected for quantitative-PCR validation of the SuperSAGE results (*FER2*, *CSP*, *FSA*, *LHCII-2.1*, *LHCB5*, *LHCII-3*, *PHOA* and *RPI*). A significant positive correlation in differential gene expression was observed (Pearson $r=0.7$; $p=1.7 \times 10^{-27}$) (SI Figure 4).

Multi-dimensional scaling of TMM-normalised (see methods) tag expression separated the replicate groups by light and strain (Figure 4A). The *Ica* strains were less separated

on the first dimension compared to WT, indicating a similar reduction in light response. There was a clear divergence on the second dimension for *lca2*, indicating a different mode of light response. In contrast, *lca1* scaled with WT on the second dimension and medium/high light on the first dimension, suggesting a low degree of light response that largely overlaps with WT under higher light. This pattern shows a strong qualitative agreement with the observed phenotypic differences across strains and light levels (SI Figure 5A and 5D). Global expression profiles (Figure 4B), demonstrated a marked difference in the number of significantly differentially expressed (SDE) tags with false discovery rates (FDR) below 0.05 between WT and the *lca* strains over different light conditions. In WT as many as 26% of all tags are SDE when comparing low to high light, this was reduced to 6.6% in *lca1* and 7.8% in *lca2*.

This overall reduction in light regulation of genes in *lca1* and *lca2* was further defined by mapping the function of SDE tags using gene ontology (GO). A strong reduction in the number of SDE tags with GO terms assigned to cellular components was observed in both *lca* strains (Figure 4C), indicating a general reduction in the response to light in both *lca* mutants and specifically a dramatic reduction in the regulation of chloroplast (plastid and thylakoid) components. In *lca1*, none of the tags with GO terms assigned to plastid components were SDE when comparing low to high light. In *lca2* the number was reduced by 50% for thylakoid and 73% for plastid components compared to WT. This further clarifies that the reduction in light regulation in the *lca* mutants is closely linked to processes associated with the chloroplast.

Regulation of photosynthetic genes

Genes identified as involved in photosynthesis by the assigned KEGG KO were investigated for average fold-change and numbers of low versus high light SDE genes (FDR < 0.05). There was a reduction in the average fold-change from 3.0 in WT to 1.6 and 2.0 in *lca1* and *lca2*, respectively, and a reduction in the number of SDE genes from 15 in WT to 3 and 9, respectively (Figure 5A).

The expression profiles for genes assigned to PSII, the oxygen-evolving complex (OEC), PSI, and the light harvesting complex (LHC) of each photosystem (LHCII and LHCI) show few significant changes between *lca* strains and WT under high light (Figure 5B-C). Expression profiles then diverge at lower light levels (Figure 5B-C). The majority of these SDE photosynthetic genes show reduced expression in *lca* strains compared to WT. The most significant reduction in gene expression was seen when comparing *lca1* to WT under low light (Figure 5B), with a ~2 to 4-fold reduction in many genes assigned to PSII, OEC, PSI and LHCII gene expression, with further ~2-fold reductions in many genes under medium light. The expression profile changes were more modest when comparing *lca2* to WT (Figure 5C), with a lower number of SDE genes under low and medium light than between *lca1* and WT, and reduced variation in gene expression with light. The reduced expression of *LHC* genes compared to WT, particularly *LHCII* linked genes, in both *lca1* and *lca2* under low and medium light compared to WT has the potential to result in a decreased amount of chlorophyll *a/b*-binding proteins and a reduced chlorophyll binding capacity under low and medium light, mirroring the results observed in the *lca* phenotypes with a reduced chlorophyll content per cell and reduced σ PSII.

Discussion

Reductions in the size of LHCs in microalgae have previously been shown to result in increased photosynthetic efficiency or productivity through reduced energy loss via fluoresced light or heat dissipation under saturating light conditions [38]. Compared to WT the two *lca* strains characterized in this study have an increase in chlorophyll a:b ratio, (Figure 3B) (indicative of a reduction of pigment associated with LHC) [39], reduced total chlorophyll per cell (SI Figure 5B-C), a decrease in the measured functional size of PSII antenna (Figure 1D) and a reduction in expression of genes associated with light harvesting proteins (Figure 5). Combined, these characteristics indicate that our two novel *lca* strains have truncated light harvesting antenna that result in an increase in the light intensity (I_s) at which photosynthesis saturates (P_{max}) in both strains; in *lca1*, this increase was significant compared to WT (Table 2). While results from other truncated antenna phenotypes from different algal species in large-scale experiments have been variable [40, 41], the phenotypes of *lca1* and *lca2* could theoretically result in increased biomass if grown on a large scale. Indeed, it is conceivable that randomly selected stable mutants in which maintenance of growth rate is a selection criteria will outperform WT cell lines; this is in contrast to reports with rationally engineered strains in which growth rate may have been compromised [42]. Ort et al. have previously suggested that four different parameters, namely the size of the LHC, the concentration of functional photosystems, the quantum yield of photosynthesis and its relation to the functional antenna size, are useful for assessing the potential for truncated light antenna (*tla*) strains to outperform WT strains [43, 44]. While *lca1* meets many of these criteria (Table 3), with the exception of the relative

reduction in PSI to PSII, *lca2* does so only at certain light intensities, which reflects differences in the underlying physiology of the selected strains.

lca1 had reduced chlorophyll content and smaller antenna size at all measured light intensities. *lca2* displays a more dynamic physiology, increasing the chlorophyll content and antenna size at low light in a manner which more closely resembles WT physiology (Table 3). Therefore, while *lca1* seems to be largely 'locked' into a truncated antenna and pseudo-'high-light' phenotype (although see below), *lca2* has retained the ability to more fully photoacclimate to altered irradiance. Further differences in phenotype can be seen in the trend of the size of the functional antenna of PSII and in the pigment profiles of *lca1* and *lca2* at different light intensities (SI Figure 5). WT and *lca2* both increase the size of σ PSII (SI Figure 5A) and the amount of chlorophyll *a* and *b* per cell (SI Figure 5B-C) in response to growth under low light; however, while the increase is proportional in WT, the increase in chlorophyll *b* in *lca2* is reduced compared to chlorophyll *a*, which results in an increase in the chlorophyll *a/b* ratio (SI Figure 5D). The chlorophyll *b* levels in *lca1* are low under medium and high light and indeed below detection at low light. The proportional increase in chlorophyll *a* and *b* in WT and *lca1* and the lack of a change in σ PSII suggest that this strain primarily adjusts to changes in light by increasing or decreasing the number of photosystems (n-type acclimation) [45]; this is in contrast to *lca2*, which changes the size of antennae (σ -type acclimation). When grown in mass culture, the ability to photoacclimate and increase the size of σ PSII and pigment content under low light and sub-saturating light as seen in *lca2* may be beneficial when under relatively constant low light conditions. However, mixing rate, depth and cell (pigment) density within the culture will all combine to dictate

the frequency of light fluctuations, which may be rapid in many cases. Indeed, within a rigorously mixed mass culture with rapid to strong fluctuations in light intensity, the consistently small antenna of (*lca1*) may be beneficial. However, the lower apparent capacity for rapidly inducible NPQ within this strain (Figure 2) may be disadvantageous in such a situation.

At a gene expression level, the transcriptomic analysis shows that both *lca* strains display a reduced ability to regulate photosynthesis-associated genes (photoacclimate) in response to changes in light compared to WT. More than three times as many tags are differentially expressed in WT versus the *lca* strains when comparing the expression under low to high light (Figure 4B) and fewer tags linked to the chloroplast and the thylakoid membrane were significantly (FDR < 0.05) differentially expressed in cultures grown under low light versus high light in *lca1* and *lca2* compared to WT (Figure 4C); therefore, a far lower number of tags linked to the chloroplast were upregulated in response to low light in either *lca1* or *lca2*. In *lca1* not a single tag assigned to the plastid was SDE, while in *lca2* the number was reduced by 75%. The WT, as expected, displayed a strong upregulation of gene expression when grown under low light. The greater overall similarity of expression profiles for *lca2* and *lca1* compared to WT, even given the differences in phenotype would also suggest that both *lca1* and *lca2* mutants share a common factor associated with the overall ability to either sense or regulate photosynthesis in regard to changes in light, rather than for example a simple knockout of a *LHC* gene. This could suggest that the mutations in *lca1* and *lca2* target a generic upstream mechanism for altering the

antenna size of photosynthetic species, which, if identified, could be targeted using sophisticated gene editing technologies such as CRISPR [9, 46].

Several mutations affecting the composition of the LHC have previously been reported. The transcriptomes were searched against these genes where possible to determine if the observed phenotypes resulted from known mutations. Several matches were detected, including the genes associated with the *tla1*, *tla2* (*CpFTSY*) and *tla3* (*CpSRP43*) [24, 30, 31] truncated antenna mutants as well as the *stm3* mutant (increased chlorophyll per cell) [29]. These genes were found in low or moderate numbers, but neither were found to be SDE compared to WT with the exception of *CpSRP43* that was found to be moderately upregulated in *Ica2* under low and medium light and in *Ica1* under medium light, in contrast to the *tla3* mutant that result from a deletion of *CpSRP43*. Hence, while the genetic basis of *Ica1* and *Ica2* may be different to that in the *tla* mutants [24, 28, 29], from a physiological perspective *Ica1* shares the phenotype (a strong reduction in the size of the functional antenna of PSII) with the *tla1* mutant [15], as well as the *D. salina* mutant *dcd1* [34]. In *tla1*, the amount of CP26 protein was strongly reduced [24]. This protein is encoded by the *LHCB5* gene, which was significantly (FDR < 0.05) down-regulated in both of the *Ica* strains, consistent with the similarity of the phenotypes. Similar molecular phenotypes may thus be achievable through the genetic alteration of multiple different genes. The *dcd1* of mutant *D. salina* [25] displayed a similarly irregular xanthophyll profile to *Ica1* (Figure 3A); however, while *dcd1* accumulates zeaxanthin under high light and stores normal amounts of violaxanthin under low light, the opposite was observed in *Ica1*. Interestingly, CP26 is also thought to play a role in the qI (slower photoinhibition) component of NPQ [47],

linked to conformational changes after alternative binding of zeaxanthin or violaxanthin to the inner allosteric site L2, which also result in changes to the fluorescent yield [48]. This suggests that the observed xanthophyll profile in *lca1* with specific changes to these two pigments, and potentially the reduced NPQ capacity, could result from a mutation that has affected this binding site or simply from a reduction in the amount of available CP26 through the down-regulation of *LHCB5*. The *PSBS* gene, also associated with the interface [49, 50] between PSII and the LHCII, thought to play a key role in qE (the fast component of NPQ) [51, 52], was detected and was either equal to or increased in both *lca* mutants compared to *WT* under medium light (Figure 5B-C), while no matches were found for genes encoding LHCSR, the protein typically associated with qE in green algae [53].

The pattern of gene regulation in *lca2* is indicative of a suppressed light response, which results in less increase in PSII LHC genes compared to *WT* under decreasing light as well as a lack of up-regulation of PSI LHC under low light. The photoacclimation response in *lca1* is further limited, with a large reduction in expression of genes related to PSII LHC and PSI LHC compared to *WT*. This gene regulation matches the observed phenotype closely. While direct changes in regulatory genes may be involved, another possibility is that the strains have fully or partially lost the ability to sense changes in light conditions. In *D. salina* it has been proposed that photoacclimation due to changes in irradiance results from two different signal transduction pathways that regulate the expression of *CAO* and *LHCB* [54]. The gene encoding *CAO* was not detected in the extended annotation presented here due to the lack of a matching EST. Also, the *C. reinhardtii cbs3* mutant, which lacked a functional

CAO gene [55], was fully chlorophyll *b* deficient, which again may suggest that the mutations have targeted a regulatory mechanism rather than the actual gene. Based on the similar phenotype and expression profile in the *lca* strains presented here a gene involved in the regulation of photosynthesis may well be the basis of one or both of the mutations responsible for the *lca1* and *lca2* phenotypes.

Conclusions

The approach presented in this study, using random mutagenesis of a microalgal strain coupled with high-throughput selection for truncated antenna strains via a chlorophyll proxy, efficiently selected for the targeted phenotype. Detailed evaluation of two selected strains showed that the observed phenotypes displayed differences that were qualitatively paralleled by changes in the expression of nuclear-encoded genes directly involved in photosynthesis. The method described can conceivably be applied to any culturable microalgal species without the need for detailed characterisation of the individual genome, meaning that new species, more suitable for a specific niche, can be exploited for commercial or scientific purposes.

Methods

Strains, culturing and long-term maintenance

Samples of *Dunaliella tertiolecta* CCMP364 were ordered from the NCMA culture collection. Cultures were maintained in triplicate under 100-200 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ of white light under a 12hr light, 12 hr dark cycle at 22-24°C in duplicate locations, both in liquid f/2 - Si media as well as on f/2 - Si agar plates [56, 57]. Sub-culturing was performed every second week for liquid cultures and every second month for plated cultures. During experiments, 25 mL of fresh liquid media were inoculated to a concentration of 25000 cells per mL in 40 mL canted and vented cell culture flasks, from starter cultures pre-acclimated to the experimental light conditions and incubated in AG130 growth chambers (PSI, Drasov, Czech Republic). The bottles were shaken and repositioned on a daily basis to increase aeration and minimize potential variation in light exposure. In addition, WT and mutant strains were cryopreserved following the techniques of Tanniou et al. by encapsulation in sodium alginate beads and successfully resurrected after long-term (> 6 month) suspended storage in liquid nitrogen vapour [58].

Mutagenesis and flow-cytometry based cell sorting

Pigment mutants were generated using chemical mutagenesis using ethyl methanesulfonate (EMS). WT cells grown to exponential phase under 100 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ of white light (12-hour light, 12-hour dark cycle) in f/2 - Si media were incubated with 0.01% of EMS for 2 hr in the dark after which the cells were pelleted by centrifugation and washed three times with fresh media. Washed cells were then transferred into 20 mL of liquid media and regrown under the conditions described above for 10 days. Fluorescence-activated sorting of surviving cells was done by a MoFlo cytometer (DakoCytomation Ltd.) at a pressure of 60 psi of sheath fluid (0.1 %

NaCl w/v in MilliQ water) using a 70 μm nozzle. The sorter was aligned following a standard manufacturer procedure using 3.0 μm yellow-green microspheres (Polysciences, Germany). The laser (Innova 90C-A4 UV, Coherent) was tuned to 488 nm and aligned through the first pinhole. Light scatter at 90° (SSC) of analysed cells was detected using the 457 \pm 25 nm optical filter and cellular chlorophyll red autofluorescence was detected using the 670 \pm 15 nm optical filter. The Summit 5.3 software was used to sort the cells with the lowest relative chlorophyll fluorescence while maintaining cell-size (SI Figure 1). Sorted cells were collected in 96-well plates containing 100 μL of media using the CyCLONE component of the MoFlo Sorter. Following incubation any viable cell lines with high cell density and low chlorophyll (SI Figure 2) were inoculated into larger 25 ml culture volumes and then further screened for photophysiological parameters as described below.

Growth rate/cell count

Cell density and size were monitored daily on fresh culture samples with a Beckman Coulter counter (Multisizer 3, Beckman Coulter) using 100-500 μL of sample and 9.5-9.9 mL of 3% (w/v) NaCl solution as diluent and electrolyte. The specific growth rate (μ) was calculated by fitting a linear regression to the natural log of the measured cell count versus the time for the exponential phase (with the slope corresponding to the growth rate) as suggested in [59]. The doubling time was calculated as the natural log of 2 divided by the specific growth rate (μ).

Pigment analysis

Chlorophyll *a* content was measured by filtering 1.0 mL of liquid culture using a hand-held pump on to MF 100 microfiber filters (Fisher Scientific UK Ltd, Loughborough, UK). The filtered content was re-suspended in 4 mL of 90% acetone in plastic tubes under low light, immediately covered in aluminium foil and kept in the cold overnight (4°C). The extracted dissolved chlorophyll was analysed on a Turner Designs TD-700 fluorometer (Turner Designs, Sunnyvale, CA, USA) [60]. Total pigment analysis was achieved using high-performance-liquid-chromatography (HPLC) following the methods of Barlow et al., yielding information on the absolute photosynthetic pigment complement of cell lines [61]. Samples were prepared by filtering down the equivalent of 1 mL of cells onto MF 100 microfiber filters (Fisher Scientific UK Ltd, Loughborough, UK), followed by re-suspension in 1 mL of 90% acetone. The cells were then disrupted using sonication for 30 s using an Ultrasonics W-380 sonicator with a C3 probe (Heat Systems Ultrasonics). Samples were kept on ice and carefully protected from light during all stages of the preparation.

Fluorescent kinetics

Variable chlorophyll fluorescence was measured using a Chelsea Scientific Instruments FastTracka™ Mk II Fast Repetition Rate fluorometer (FRRf) integrated with a FastAct™ Laboratory system. All samples were dark acclimated for 30 min and FRRf measurements were blank corrected using filtered growth media. Rapid light curves were acquired at 14 light intervals ranging from 7 to 2046 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$, every 4 sec over a 45 min period, with 24 sequences per acquisition, each sequence consisting of 150 flashlets in the saturation phase (1 μs duration) and 25 flashlets in the relaxation phase (84 μs duration). Data from the FRRf were analysed to derive

relative values of the minimal and maximal fluorescence (F_0 and F_m) and hence F_v/F_m , as well as the functional absorption cross-section of PSII (σ_{PSII}) [62, 63]. Minimum fluorescence yield under actinic light (F'_o) was calculated using the measured F_o , F_m and F'_m [64], NPQ was defined using the normalised Stern-Volmer coefficient [65], while the electron transport rate was calculated based on Kolber and Falkowski (1993) [66] and Gorbunov et al. (2001) [67] following Fujiki et al. (2007) [68].

Western blot analysis

Photosynthetic protein abundances were quantified using techniques similar to those described elsewhere [69–71]. Samples for protein extraction were collected by concentrating 35 mL of liquid culture down to 1 mL using centrifugation. Processing of the sample was performed according to the protocol described by Brown et al. (2008) [69]. Quantification was performed using primary antibodies designed against peptide tags from protein subunits that are both representative of the functional photosynthetic complex and are conserved across all oxygenic photosynthetic species namely: PSBA (PSII), PSAC (PSI), RBCL (RuBisCO) [72]. Samples were run alongside peptide standards (Agrisera) following the procedure described by Brown et al. (2008) [69]. Protein levels on immunoblots were quantified using QuantityOne™ and Image Lab™ software using independent standard curves specific to each blot (Brown et al. 2008) [69]. Results were only used when samples fell within the linear range of loaded standards.

PE curves

The rate of oxygen evolution (P) was measured as a function of light intensity (I) using an Oxylab unit and a DW1 electrode chamber (Hansatech Instruments, Norfolk, UK). To measure the oxygen evolution rate, a 2 mL sample was adjusted to a concentration of approximately 1 million cells to assure absence of cell self-shading and transferred to the electrode chamber maintained at room temperature. Change in dissolved oxygen concentration was first measured as a function of time in the absence of light (dark respiration) and then at 10 different light intensities ranging from 25 to 2000 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$. Oxygen evolution rates were determined in sequence from low to high light intensity and each measurement was taken for a period of 3 min after allowing for stabilization of the signal for 2 min.

SAGE library preparation

RNA was extracted using the cetyltrimethylammonium bromide (CTAB) method [73] using 35 mL of samples from cultures of WT, *Ica1* and *Ica2* grown under low, medium and high light (20, 100 and 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). The cells were disrupted on a Tissue Lyser II bead beater (Qiagen, Valencia, CA, USA) using 250 mL of beads at 20 RPM for 5 min with 1.5 mL of CTAB added per sample. The extracted RNA was transcribed into cDNA using the ImProm-II Reverse Transcription System (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. SAGE libraries were prepared and sequenced at Iwate Biotechnology Research Centre, Kitakami, Iwate, Japan, following the SuperSAGE method [74] using a more recent protocol (HT-SuperSAGE) as described in [75]. The cDNA was digested using the NlaIII restriction enzyme, ligated with adapter sequences and further cleaved using EcoP15I linked to a second set of adapter sequences, including 4-bp index sequences.

Pooled samples were sequenced on an Illumina Genome Analyzer II. The resulting reads were processed using a Perl script [75] to extract and assign tags to libraries based on the index sequences. The final data in the form of a combined count table and the original sequence reads were registered in the NCBI Gene Expression Omnibus under accession number GSE129614.

Statistical analysis

Results reported as significant are based on the results from the analysis of variance (ANOVA) and the pairwise means from Tukey HSD using the AOV and TukeyHSD functions in R respectively, together with an interaction model for the two experimental conditions, strain and light or by strain only (single light level). The homogeneity of variances and normality assumptions was tested using Levene's test (`leveneTest`) in the R `car` package and by inspecting the residuals in a Q-Q plot.

Transcriptome analysis

To improve the annotation of the short tags, all sequences were mapped against 409789 previously published expressed sequence tags (EST's) from *D. tertiolecta* [76] using GFESSA [77]. Each of the mapped EST's were subsequently re-annotated using BlastX (e-value < 1E-3, BLOSUM-62 matrix) running on the open cluster on the Oslo Bioportal against NCBI's non redundant (NR) protein database as well as the Swiss-Prot database [78]. Processing of the resulting files were performed using original and modified scripts from the guide by De Wit et al., 2012 [79], with final removal of any hits with an e-value above 1E-6. In cases where multiple ESTs were matched against

a unitag, the EST with the lowest e-value from the NCBI annotation was chosen as the representative sequence.

The statistical analysis was carried out using the edgeR packet [80]. Tags not meeting a criterion of at least one count per million in a minimum of nine libraries were filtered and removed from the analysis. Scaling factors for the library sizes of the remaining filtered tags were calculated using the weighted trimmed mean of the log expression ratios (trimmed mean of M values (TMM)) [81] as part of the package. Testing for differential expression was done using generalized linear models (GLM) likelihood ratio tests. Further validation of individual genes and associated tags was done by manual alignments against downloaded sequences from described genes.

Filtering for nuclear-encoded genes

The GI identifier for all proteins identified as either chloroplast or mitochondrial in GenBank or Swiss-Prot was downloaded from NCBI (for a total of 2162128 identifiers) and then searched against the mapped ESTs to remove any matches. Follow-up filtering was also done using a separate annotation against the *Arabidopsis thaliana* TAIR 10 protein database.

Quantitative polymerase chain reaction

Real-time PCR was carried out on an Opticon qPCR system (Bio-Rad, Hercules, California, USA) using Precision 2x real-time PCR Master Mix with SYBR green (Primerdesign Ltd, Southampton, United Kingdom) using the following thermal profile: initial denaturation at 95°C for 10 min followed by 40 cycles of denaturation for 15 s at

95°C and combined annealing/elongation for 1 min at 60°C. This was followed by a final extension for 10 min at 72°C. Dissociation curves were recorded from 60 to 95°C using a ramp speed of 2°C s⁻¹. Samples were analysed using duplicate technical repeats together with triplicate biological repeats for a total of six measurements per strain and light level. The samples were analysed using the Livak ($\Delta\Delta\text{CT}$) method using the *D. tertiolecta* *S11 rRNA* gene as the reference and WT low-light sample as the calibrator [82, 83].

Declarations

Acknowledgements: Authors would like to thank Ms Nicola Pratt (University of Southampton) for technical assistance and maintenance of cultures.

Author Contributions: SAJ, PS, CMM, TB, MT conceived and planned the experiments. SAJ and PS carried out the experiments. MZ planned and conducted FCM analysis and SAJ, RE and RT planned and conducted expression analysis. RT and KY provided sequence data. SAJ, PS, CMM, TB, MT analysed and interpreted the data and wrote the manuscript.

Funding: The authors would like to thank the Graduate School of the National Oceanography Centre (GSNOCS) for funding the PhD studentship to SAJ.

Competing interests: The authors declare that they have no competing interests.

Availability of data and materials: The datasets used and/or analysed during the current study are available from the corresponding authors on reasonable request.

Consent for publication: Not applicable.

Ethics approval and consent to participate: Not applicable.

References

1. Guiry MD. How Many Species of Algae Are There? *J Phycol.* 2012;48:1057–63. doi:10.1111/j.1529-8817.2012.01222.x.
2. Engelen S, Hingamp P, Sieracki M, Vargas C, Audic S, Henry N, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science* (80-). 2015;348:1261605-1/11.
3. Berepiki A, Gittins JR, Moore CM, Bibby TS. Rational engineering of photosynthetic electron flux enhances light-powered cytochrome P450 activity. *Synth Biol.* 2018;3.
4. Berepiki A, Hitchcock A, Moore CM, Bibby TS. Tapping the Unused Potential of Photosynthesis with a Heterologous Electron Sink. *ACS Synth Biol.* 2016;5:1369–75.
5. Brooijmans RJW, Siezen RJ. Genomics of microalgae, fuel for the future?: Genomics update. *Microb Biotechnol.* 2010;3:514–22.
6. U.S. Department of Energy Joint Genome Institute. Algae. 2018. <https://genome.jgi.doe.gov/algae/algae.info.html>.

7. Jackson HO, Berepiki A, Baylay AJ, Terry MJ, Moore CM, Bibby TS. An inducible expression system in the alga *Nannochloropsis gaditana* controlled by the nitrate reductase promoter. *J Appl Phycol*. 2018;:1–11.
8. Scaife MA, Smith AG. Towards developing algal synthetic biology. *Biochem Soc Trans*. 2016;44:716–22.
9. Jeon S, Lim JM, Lee HG, Shin SE, Kang NK, Park Y II, et al. Current status and perspectives of genome editing technology for microalgae. *Biotechnol Biofuels*. 2017;10:1–18.
10. Esland L, Larrea-Alvarez M, Purton S. Selectable Markers and Reporter Genes for Engineering the Chloroplast of *Chlamydomonas reinhardtii*. *Biology (Basel)*. 2018;7:46.
11. Doron L, Segal N, Shapira M. Transgene Expression in Microalgae—From Tools to Applications. *Front Plant Sci*. 2016;7 April:1–24.
12. Thai T, Doan Y, Obbard JP. Enhanced intracellular lipid in *Nannochloropsis* sp. via random mutagenesis and flow cytometric cell sorting. *Algal Res*. 2012;1:17–21. doi:10.1016/j.algal.2012.03.001.
13. Chaturvedi R, Fujita Y. Isolation of enhanced eicosapentaenoic acid producing mutants of *Nannochloropsis oculata* ST-6 using ethyl methane sulfonate induced mutagenesis techniques and their characterization at mRNA transcript level. *Phycol Res*. 2006;54 December 2004:208–19.
14. Mendoza H, de la Jara A, Freijanes K, Carmona L, Ramos AA, de Sousa Duarte V, et al. Characterization of *Dunaliella salina* strains by flow cytometry: a new approach to select carotenoid hyperproducing strains. *Electron J Biotechnol*. 2008;11:0–0. doi:10.2225/vol11-issue4-fulltext-2.

15. Manandhar-Shrestha K, Hildebrand M. Development of flow cytometric procedures for the efficient isolation of improved lipid accumulation mutants in a *Chlorella* sp. microalga. *J Appl Phycol*. 2013;25:1643–51.
16. Rodolfi L, Chini Zittelli G, Bassi N, Padovani G, Biondi N, Bonini G, et al. Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol Bioeng*. 2009;102:100–12. <http://www.ncbi.nlm.nih.gov/pubmed/18683258>.
17. Stephenson PG, Moore CM, Terry MJ, Zubkov M V, Bibby TS. Improving photosynthesis for algal biofuels: toward a green revolution. *Trends Biotechnol*. 2011;29:615–23. doi:10.1016/j.tibtech.2011.06.005.
18. Gimpel JA, Specht EA, Georgianna DR, Mayfield SP. Advances in microalgae engineering and synthetic biology applications for biofuel production. *Curr Opin Chem Biol*. 2013;17:489–95. doi:10.1016/j.cbpa.2013.03.038.
19. Dyo YM, Purton S. The algal chloroplast as a synthetic biology platform for production of therapeutic proteins. *Microbiol (United Kingdom)*. 2018;164:113–21.
20. Purton S, Economou CK, Wannathong T, Waterhouse JC, Young REB. New tools for chloroplast genetic engineering allow the synthesis of human growth hormone in the green alga *Chlamydomonas reinhardtii*. *Appl Microbiol Biotechnol*. 2016;100:5467–77. doi:10.1007/s00253-016-7354-6.
21. Chiaiese P, Corrado G, Colla G, Kyriacou MC, Roupheal Y. Renewable sources of plant biostimulation: Microalgae as a sustainable means to improve crop performance. *Front Plant Sci*. 2018;871 December:1–6.
22. Gaignard C, Gargouch N, Dubessay P, Delattre C, Pierre G, Laroche C, et al. New horizons in culture and valorization of red microalgae. *Biotechnol Adv*.

2019;37:193–222. doi:10.1016/j.biotechadv.2018.11.014.

23. Polle JEW, Kanakagiri S, Jin E, Masuda T, Melis A. Truncated chlorophyll antenna size of the photosystems - A practical method to improve microalgal productivity and hydrogen production in mass culture. *Int J Hydrogen Energy*. 2002;27:1257–64.

24. Polle JEW, Kanakagiri S-D, Melis A. tla1, a DNA insertional transformant of the green alga *Chlamydomonas reinhardtii* with a truncated light-harvesting chlorophyll antenna size. *Planta*. 2003;217:49–59. doi:10.1007/s00425-002-0968-1.

25. Kolber Z, Wyman K V, Falkowski PG. Natural variability in photosynthetic energy conversion efficiency: A field study in the Gulf of Maine. *Limnol Oceanogr*. 1990;35:72–9. doi:10.4319/lo.1990.35.1.0072.

26. Moore CM, Suggett D, Holligan PM, Sharples J, Abraham ER, Lucas MI, et al. Physical controls on phytoplankton physiology and production at a shelf sea front: a fast repetition-rate fluorometer based field study. *Marine Ecology Progress Series*. 2003;259:29–45.

27. Kosourov S, Ghirardi M, Seibert M. A Truncated Antenna Mutant of *Chlamydomonas reinhardtii* Can Produce More Hydrogen than the Parental Strain. *Int J Hydrogen Energy*. 2011;36:2044–8.

28. Bellafiore S, Ferris P, Naver H, Göhre V, Rochaix J-D. Loss of Albino3 leads to the specific depletion of the light-harvesting system. *Plant Cell*. 2002;14:2303–14.

29. Mussgnug J, Wobbe L, Elles I. NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. *Plant Cell* 2005;17 December:3409–21. doi:10.1105/tpc.105.035774.algae.

30. Kirst H, García-Cerdán JG, Zurbriggen A, Melis A. Assembly of the Light-Harvesting Chlorophyll Antenna in the Green Alga *Chlamydomonas reinhardtii* Requires Expression of the TLA2 - CpFTSY Gene. *Plant Physiol.* 2012;158:930–45. doi:10.1104/pp.111.189910.
31. Kirst H, Garcia-Cerdan JG, Zurbriggen a., Ruehle T, Melis a. Truncated photosystem chlorophyll antenna size in the green microalga *Chlamydomonas reinhardtii* upon deletion of the TLA3-CpSRP43 gene. *Plant Physiol.* 2012.
32. Pick U, Avidan O. Triacylglycerol is produced from starch and polar lipids in the green alga *Dunaliella tertiolecta*. *J Exp Bot.* 2017;68:4939–50.
33. Perrine Z, Negi S, Sayre RT. Optimization of photosynthetic light energy utilization by microalgae. *Algal Res.* 2012;1:134–42. doi:10.1016/j.algal.2012.07.002.
34. Jin E, Polle JEW, Melis A. Involvement of zeaxanthin and of the Cbr protein in the repair of photosystem II from photoinhibition in the green alga *Dunaliella salina*. *Biochim Biophys Acta - Bioenerg.* 2001;1506:244–59.
35. Macey AI, Ryan-Keogh T, Richier S, Moore CM, Bibby TS. Photosynthetic protein stoichiometry and photophysiology in the high latitude north atlantic. *Limnol Oceanogr.* 2014;59:1853–64.
36. Ryan-Keogh TJ, Macey AI, Cockshutt AM, Moore CM, Bibby TS. The cyanobacterial chlorophyll-binding-protein IsiA acts to increase the in vivo effective absorption cross-section of PSI under iron limitation. *J Phycol.* 2012;48:145–54.
37. Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, Albertyn Z, et al. High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS One.* 2010;5:e12010. doi:10.1371/journal.pone.0012010.

38. Mussgnug JH, Thomas-Hall S, Rupprecht J, Foo A, Klassen V, McDowall A, et al. Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. *Plant Biotechnol J.* 2007;5:802–14. doi:10.1111/j.1467-7652.2007.00285.x.
39. Tanaka A, Melis A. Irradiance-Dependent Changes in the Size and Composition of the Chlorophyll a-b Light-Harvesting Complex in the Green Alga *Dunaliella salina*. *Plant Cell Physiol.* 1997;38:17–24. doi:10.1093/oxfordjournals.pcp.a029080.
40. Huesemann MH, Hausmann TS, Bartha R, Aksoy M, Weissman JC, Benemann JR. Biomass productivities in wild type and pigment mutant of *Cyclotella* sp. (Diatom). *Appl Biochem Biotechnol.* 2009;157:507–26. doi:10.1007/s12010-008-8298-9.
41. Nakajima Y, Tsuzuki M, Ueda R. Improved productivity by reduction of the content of light-harvesting pigment in *Chlamydomonas perigranulata*. *J Appl Phycol.* 2001;6714:95–101.
42. Murray J. Chapter 14 Redox Active Protein Maquettes: Multi-functional “Green Enzymes.” In: *Molecular Solar Fuels*. The Royal Society of Chemistry; 2012. p. 408–25. doi:10.1039/9781849733038-00408.
43. Ort DR, Zhu X, Melis A. Optimizing Antenna Size to Maximize Photosynthetic Efficiency. *PLANT Physiol.* 2011;155:79–85. doi:10.1104/pp.110.165886.
44. Melis A. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Sci.* 2009;177:272–80. doi:10.1016/j.plantsci.2009.06.005.
45. Falkowski PG, Owens TG. Light-Shade Adaptation : TWO STRATEGIES IN MARINE PHYTOPLANKTON. *Plant Physiol.* 1980;66:592–5.

46. Guihéneuf F, Khan A, Tran L-SP. Genetic Engineering: A Promising Tool to Engineer Physiological, Biochemical, and Molecular Stress Resilience in Green Microalgae. *Front Plant Sci.* 2016;7 March:1–8. doi:10.3389/fpls.2016.00400.
47. Dall'Osto L. A Mechanism of Nonphotochemical Energy Dissipation, Independent from PsbS, Revealed by a Conformational Change in the Antenna Protein CP26. *Plant Cell Online.* 2005;17:1217–32.
48. Moya I, Silvestri M, Vallon O, Cinque G, Bassi R. Time-Resolved Fluorescence Analysis of the Photosystem II Antenna Proteins in Detergent Micelles and Liposomes. *Biochemistry.* 2001;40:12552–61. doi:10.1021/bi010342x.
49. Dekker JP, Boekema EJ. Supramolecular organization of thylakoid membrane proteins in green plants. *Biochim Biophys Acta.* 2005;1706:12–39. doi:10.1016/j.bbabi.2004.09.009.
50. Nield J, Funk C, Barber J, Hideg E, Heber U, Asada K, et al. Supermolecular structure of photosystem II and location of the PsbS protein. *Philos Trans R Soc B Biol Sci.* 2000;355:1337–44.
51. Niyogi KK, Li XP, Rosenberg V, Jung HS. Is PsbS the site of non-photochemical quenching in photosynthesis? *J Exp Bot.* 2005;56:375–82.
52. Holt N, Fleming G, Niyogi K. Toward an understanding of the mechanism of nonphotochemical quenching in green plants. *Biochemistry.* 2004;43. <http://pubs.acs.org/doi/abs/10.1021/bi0494020>.
53. 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;462:518–21. doi:10.1038/nature08587.
54. Masuda T, Tanaka A, Melis A. Chlorophyll antenna size adjustments by

- irradiance in *Dunaliella salina* involve coordinate regulation of chlorophyll a oxygenase (CAO) and Lhcb gene expression. *Plant Mol Biol.* 2003;51:757–71.
55. Tanaka A, Ito H, Tanaka R, Tanaka NK, Yoshida K, Okada K. Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. *Plant Biol.* 1998;95 October:12719–23.
56. Guillard RRL. Culture of Phytoplankton for Feeding Marine Invertebrates. In: Smith W, Chanley M, editors. *Culture of Marine Invertebrate Animals SE - 3.* Springer US; 1975. p. 29–60. doi:10.1007/978-1-4615-8714-9_3.
57. Guillard RRL, Ryther JH. STUDIES OF MARINE PLANKTONIC DIATOMS: I. CYCLOTELLA NANA HUSTEDT, AND DETONULA CONFERVACEA (CLEVE) GRAN. *Can J Microbiol.* 1962;8:229–39. doi:10.1139/m62-029.
58. Tanniou A, Turpin V, Lebeau T. Comparison of cryopreservation methods for the long term storage of the marine diatom *Haslea ostrearia* (simonsen). *Cryobiology.* 2012;65:45–50. doi:10.1016/j.cryobiol.2012.03.011.
59. Andersen RA. *Algal culturing techniques.* Academic Press; 2005.
60. Welschmeyer N a. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnol Oceanogr.* 1994;39:1985–92.
61. Barlow R, Cummings D, Gibb S. Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. *Mar Ecol Prog Ser.* 1997;161:303–7.
62. Kolber Z, Falkowski PG. Use of active fluorescence to estimate phytoplankton photosynthesis in situ. 1993;38:1646–65.
63. Kolber ZS, Prášil O, Falkowski PG. Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and

- experimental protocols. *Biochim Biophys Acta - Bioenerg.* 1998;1367:88–106.
doi:10.1016/S0005-2728(98)00135-2.
64. Oxborough K, Baker NR. Resolving chlorophyll a fluorescence images of photosynthetic efficiency into photochemical and non-photochemical components -- calculation of qP and Fv-/Fm-; without measuring Fo-; *Photosynth Res.* 1997;54:135–42. doi:10.1023/A:1005936823310.
65. Mckew BA, Davey P, Finch SJ, Hopkins J, Lefebvre SC, Metodiev M V., et al. The trade-off between the light-harvesting and photoprotective functions of fucoxanthin-chlorophyll proteins dominates light acclimation in *Emiliana huxleyi* (clone CCMP 1516). *New Phytol.* 2013;200:74–85.
66. Kolber Z, Falkowski PG. Use of active fluorescence to estimate phytoplankton photosynthesis in situ. *Limnol Oceanogr.* 1993;38:1646–65.
67. Gorbunov MY, Kolber ZS, Lesser MP, Falkowski PG. Photosynthesis and photoprotection in symbiotic corals. *Limnol Oceanogr.* 2001;46:75–85.
68. Fujiki T, Suzue T, Kimoto H, Saino T. Photosynthetic electron transport in *Dunaliella tertiolecta* (Chlorophyceae) measured by fast repetition rate fluorometry: Relation to carbon assimilation. *J Plankton Res.* 2007;29:199–208.
69. Brown CM, MacKinnon JD, Cockshutt AM, Villareal TA, Campbell DA. Flux capacities and acclimation costs in *Trichodesmium* from the Gulf of Mexico. *Mar Biol.* 2008;154:413–22.
70. Losh JL, Young JN, Morel FMM. Rubisco is a small fraction of total protein in marine phytoplankton. *New Phytol.* 2013;198:52–8.
71. Hopkinson BM, Xu Y, Shi D, McGinn PJ, Morel FMM. The effect of CO₂ on the photosynthetic physiology of phytoplankton in the Gulf of Alaska. *Limnol Oceanogr.*

2010;55:2011–24.

72. Campbell DA, Cockshutt AM, Porankiewicz-Asplund J. Analysing photosynthetic complexes in uncharacterized species or mixed microalgal communities using global antibodies. *Physiol Plant*. 2003;119:322–7.

73. Pearson G, Lago-Leston A, Valente M, Serrão E. Simple and rapid RNA extraction from freeze-dried tissue of brown algae and seagrasses. *Eur J Phycol*. 2006;41:97–104.

74. Matsumura H, Bin Nasir K, Yoshida K, Ito A, Kahl G, Kruger D, et al. SuperSAGE array: the direct use of 26-base-pair transcript tags in oligonucleotide arrays. *Nat Methods*. 2006;3:469. doi:10.1038/NMETH882.

75. Matsumura H, Yoshida K, Luo S, Krüger DH, Kahl G, Schroth GP, et al. High-Throughput SuperSAGE. *Methods Mol Biol*. 2011;687:135–46. doi:10.1007/978-1-60761-944-4.

76. Rismani-Yazdi H, Haznedaroglu BZ, Bibby K, Peccia J. Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: Pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics*. 2011;12:148.

77. Edwards RJ. GFESSA : Genome-Free EST SuperSAGE Analysis. 2011;:1–11. <http://www.southampton.ac.uk/~re1u06/software/packages/seqsuite/>.

78. Altschul SF, Madden TL, Schäffer a a, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.

79. De Wit P, Pespeni MH, Ladner JT, Barshis DJ, Seneca F, Jaris H, et al. The simple fool's guide to population genomics via RNA-Seq: an introduction to high-

throughput sequencing data analysis. *Mol Ecol Resour.* 2012;12:1058–67.

doi:10.1111/1755-0998.12003.

80. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat Protoc.* 2013;8:1765–86. doi:10.1038/nprot.2013.099.

81. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 2010;11:R25. doi:10.1186/gb-2010-11-3-r25.

82. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif.* 2001;25:402–8. <http://www.ncbi.nlm.nih.gov/pubmed/11846609>.

83. LaRoche J, Wyman K, Falkowski P. A cDNA for *Dunaliella tertiolecta* Cytosol Ribosomal Protein S11. *Plant Physiol.* 1994;104:1447–8.

Tables

Table 1. Photophysiological parameters. Growth rate, chlorophyll *a* content per cell, chlorophyll *a:b* ratio, functional cross section of PSII (σ PSII) and photosynthetic efficiency (F_v/F_m) in WT, *lca1* and *lca2* strains grown under low, medium and high light (20, 100, 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Errors are one standard deviation from biological triplicates.

Growth parameters	Low light			Medium light			High light		
	WT	<i>lca1</i>	<i>lca2</i>	WT	<i>lca1</i>	<i>lca2</i>	WT	<i>lca1</i>	<i>lca2</i>

Growth rate (d ⁻¹)	0.30±0.02	0.22±0.01	0.28±0.01	0.69±0.11	0.62±0.16	0.68±0.10	0.85±0.04	0.83±0.08	0.96±0.02
Chl a per cell (pg cell ⁻¹)	1.86±0.25	0.34±0.04	1.06±0.06	1.26±0.13	0.65±0.05	1.18±0.08	0.97±0.05	0.63±0.02	0.68±0.04
Chl a:b ratio	3.98±0.20	N/A	4.68±0.25	4.55±0.43	8.82±1.10	9.94±0.12	4.43±0.73	8.91±0.06	12.6±1.55
σPSII (nm ²)	0.89±0.02	0.72±0.03	0.89±0.03	1.06±0.02	0.75±0.02	0.86±0.02	0.93±0.02	0.65±0.03	0.66±0.02
F _v /F _m	0.55±0.01	0.56±0.00	0.54±0.00	0.51±0.01	0.56±0.01	0.51±0.00	0.52±0.01	0.53±0.01	0.53±0.00

Table 2. Oxygen evolution and electron transport rate parameters. The calculated P_{max}, α and I_s for the chlorophyll a normalised rate of oxygen evolution and the electron transport rate of PSII (P_f) in WT, *lca1* and *lca2* strains grown under medium light (100 mmol photons m⁻² s⁻¹). Errors are one standard deviation from biological triplicates, values that are significantly different from WT (ANOVA, p < 0.05) are shown with asterisks corresponding to the p-value (***) 0.001, ** 0.01 or * 0.05) from a Tukey HSD post-hoc test.

Strain	Oxygen evolution			Electron transport rate		
	P _{max} mmol O ₂ (mol Chl a) ⁻¹ s ⁻¹	α mmol O ₂ (mol Chl a) ⁻¹ s ⁻¹ × (μmol photons m ⁻² s ⁻¹) ⁻¹	I _s μmol photons m ⁻² s ⁻¹	P _{max} μmol e (μmol RC) ⁻¹ s ⁻¹	α μmol e (μmol RC) ⁻¹ s ⁻¹ × (μmol photons m ⁻² s ⁻¹) ⁻¹	I _s μmol photons m ⁻² s ⁻¹
WT	55.07±3.23	0.115±0.05	523.6 ±205.1	481.4±47.1	0.884±0.034	543.9±33.1
<i>lca1</i>	159.42±22.10 (**)	0.209±0.073	803.33±190.00	811.0±62.1 (***)	0.757±0.010 (***)	1071±69.1 (***)
<i>lca2</i>	71.70±16.01	0.146±0.054	517.96±169.60	533.3±27.2	0.817±0.014 (*)	652.8±34.4

Table 3. Truncated light antenna parameters. The relative percent change compared to WT in functional PSII antenna size, proportion of reaction centres per cell, maximum electron transport rate (P_f (PSII) P_{max}) and the rate of photosynthesis (O₂ evolution P_{max}) suggested to impact the photosynthetic conversion efficiency.

Question	<i>lca1</i> (% of WT)	<i>lca2</i> (% of WT)	Parameter
Reduction in antenna size	70.8	80.6	σPSII
Reaction centres per cell	99.2:23.0	72.1:61.7	PsbA:PsaC
Maximum electron transport rate	168	111	P _f (PSII) P _{max}
Rate of photosynthesis	289	131	O ₂ evolution P _{max}

Figures

Figure 1. Growth rate, chlorophyll a content and PSII parameters. (A) Growth rate, (B) photosynthetic efficiency, (C) chlorophyll a content per cell, and (D) functional cross section of PSII in WT, *lca1* and *lca2* strains grown under low, medium and high light (20, 100, 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Error bars are one standard deviation from biological triplicates, values that are significantly different from WT (ANOVA, $p < 0.05$) within a light level are shown with asterisks corresponding to the p-value (***) 0.001, ** 0.01 or * 0.05 from a Tukey HSD post-hoc test.

Figure 2. Oxygen evolution, electron transport rate and photosynthetic parameters. (A) The rate of oxygen evolution in WT, *lca1* and *lca2* as measured under increasing irradiance from 0 to 2000 $\text{mmol photons m}^{-2} \text{s}^{-1}$ and normalised to chlorophyll *a* and (B) the electron transport rate of PSII (P_f), (C) the minimum fluorescence under actinic light (F'_o), (D) the maximum fluorescence under actinic light (F'_m), (E) the functional cross section under actinic light (σPSII), (F) non-photochemical quenching (NPQ) as measured by rapid light curves from 0 to 2048 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$. Cultures were grown under medium light (100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Error bars are one standard deviation from biological triplicates.

Figure 3. Accessory pigment and chlorophyll ratios. (A) The amount of accessory pigments (violaxanthin, antheraxanthin, zeaxanthin, lutein and β -carotene) and (B) chlorophyll *a:b* ratios in WT, *lca1* and *lca2* strains grown under low, medium and high light (20, 100, 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Error bars are one standard deviation from biological triplicates, values that are significantly different from WT (ANOVA, $P < 0.05$) are shown with asterisks corresponding to the p-value (** 0.001 , ** 0.01 or * 0.05) from a Tukey HSD post-hoc test. Not detected (n.d.).

Figure 4. Overview of libraries after dimensional reduction and SDE tags by light condition or cellular components. (A) The difference in leading (maximum) fold change (FC) between the top 500 differentially expressed tags as shown by multi-dimensional scaling (MDS) plots of the filtered data. Each ellipse encircles three replicates from the same condition with the calculated centre for each replicate group within one strain linked together with dashed lines. The plot shows the results for all 27 libraries, with replicates of WT, *lca1* and *lca2* strains under low, medium and high light). (B) The total number of significantly (FDR < 0.05) up or down-regulated annotated tags in WT, *lca1* and *lca2* between light levels (low vs. medium light, medium vs. high light and low vs. high light). (C) Enrichment in the mapping to cellular components within tags that are significantly differentially expressed (FDR < 0.001) based on the fold change comparing low to high light in WT, *lca1* or *lca2*. GO identifiers linked to tags with a negative fold-change (down-regulated) are shown in red, while tags with a positive fold-change (up-regulated) under the compared conditions are

shown in blue.

Figure 5. Differential expression of core photosystem and light harvesting complex genes. The number of SDE genes (FDR < 0.05) and the mean fold-change in WT, *lca1* and *lca2* strains when comparing the expression under low (20 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) versus high light (1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) (A). Differential gene expression under low, medium and high light (20, 100, 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$) in (B) *lca1* versus WT, and (C) *lca2* versus WT. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying a FDR of <0.001 and a lighter colour a FDR of < 0.05. The expression is shown as the log fold change. Genes linked to the oxygen evolving complex are shaded light green, remaining PSII-related genes are shaded darker green, PSI genes are shaded pink, while LHCII- and LHCI-related genes are shaded purple and yellow, respectively.

Supplemental tables

Supplemental Table 1. Post-selection cell lines with truncated PSII functional antenna. The top ten cell lines with most truncated photosystem II functional antenna (σPSII) compared to WT from post-sorting of mutated cells grown under medium light (100 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). The photosynthetic efficiency (F_v/F_m), and the difference of σPSII and F_v/F_m to WT are also shown. Some strains failed to grow consistently

(including #8) and was therefore removed from in-depth analysis.

Strain	σ PSII (nm ²)	σ PSII (% of WT)	F _v /F _m	F _v /F _m (% of WT)
#1	0.87	78.3	0.51	102.4
#2	0.87	78.3	0.48	95.6
#3	0.87	78.3	0.54	107.8
#4	0.86	77.4	0.54	108
#5	0.86	77.4	0.48	96
#6	0.83	74.7	0.52	104.6
#7 (<i>lca2</i>)	0.74	66.6	0.52	104
#8	0.72	64.8	0.56	112
#9	0.68	61.2	0.51	102
#10 (<i>lca1</i>)	0.63	56.7	0.53	106

Supplemental figures

Supplemental Figure 1. Selection of low chlorophyll mutagenized *D. tertiolecta* cells using FACS. Each sorting event represents a single cell and is marked by a black dot. The rectangular outline in red shows the selection window gated by the chlorophyll fluorescence as measured by the FL3-H channel and the estimated cell size by the SSC-H side scatter channel.

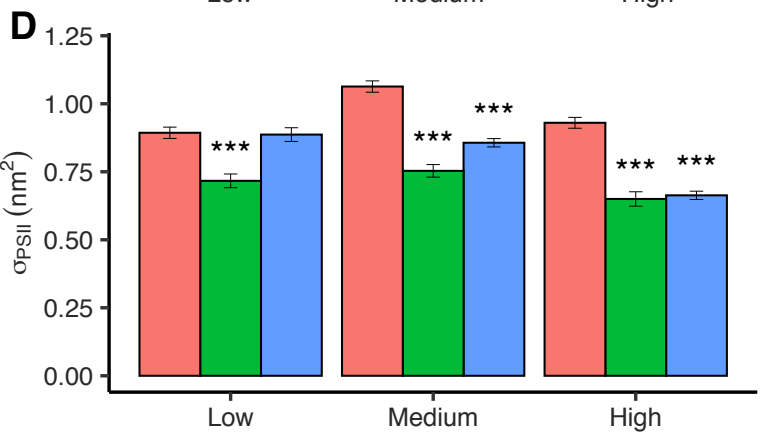
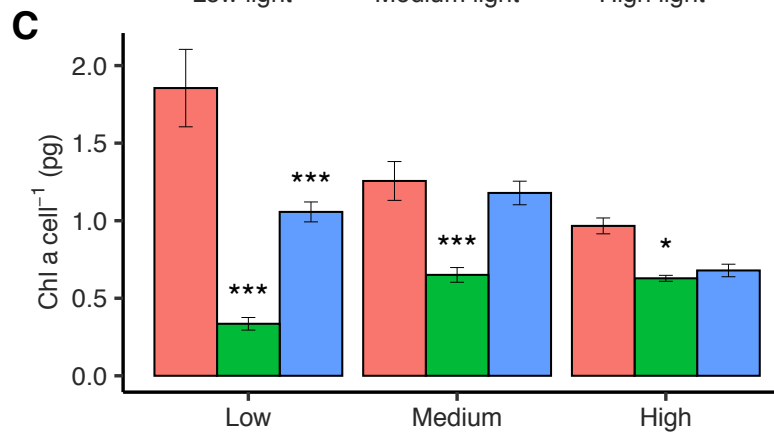
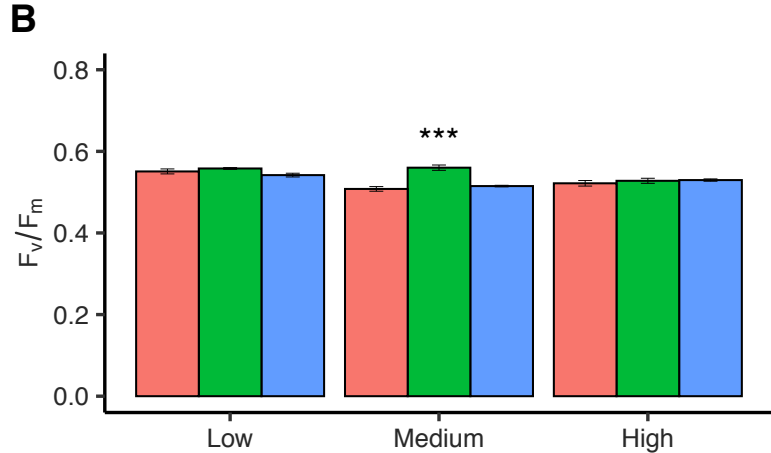
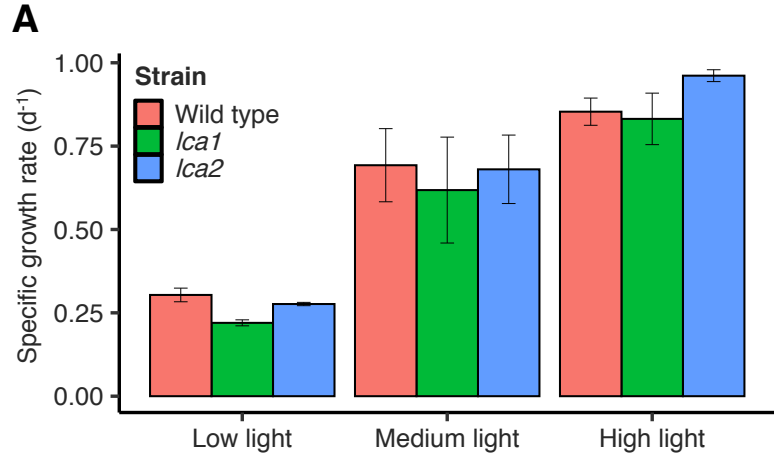
Supplemental Figure 2. Post sorting screening of low chlorophyll cells. The cells were analysed for cell density using the absorbance at 350 nm and chlorophyll fluorescence at 685 nm as a measure of total chlorophyll content using a plate reader two weeks after the original selection. Selected cells are shown as blue triangles, while cells not passing the selection are shown as red circles.

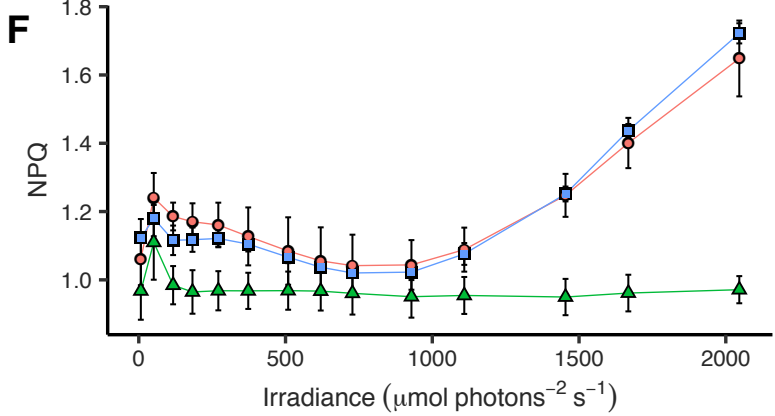
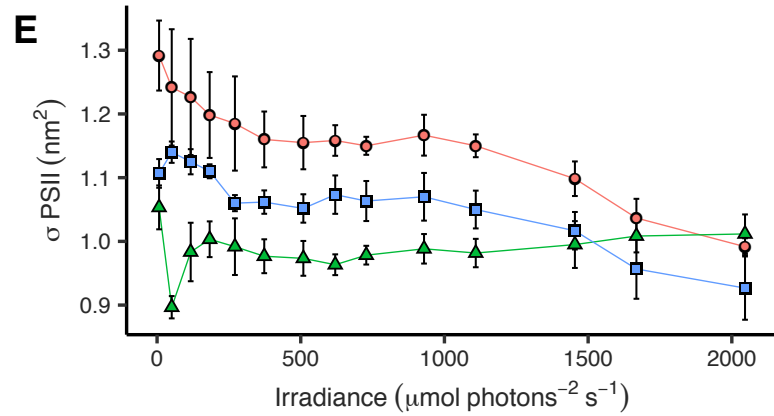
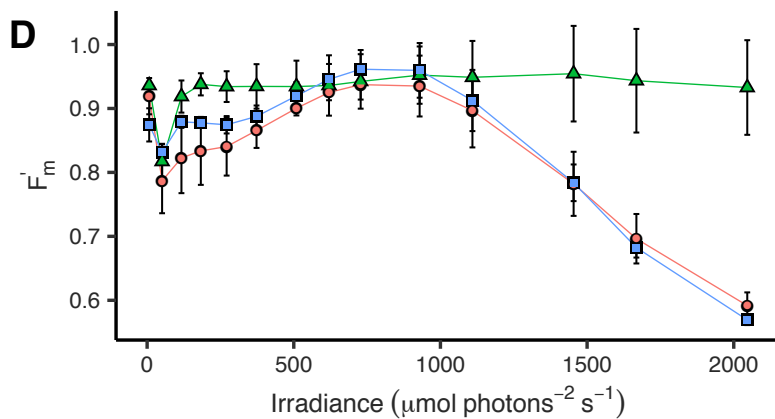
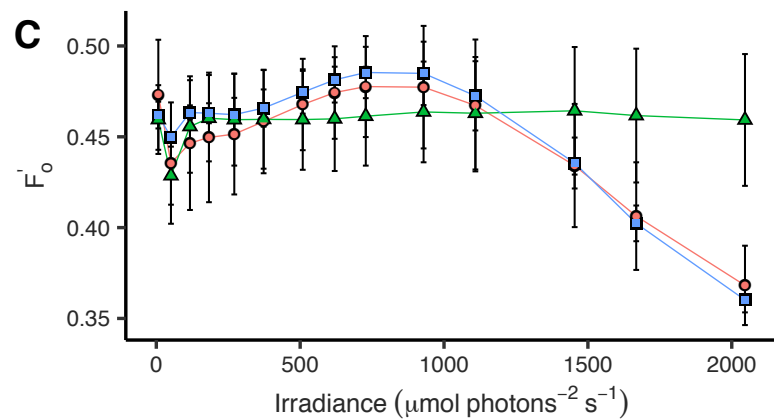
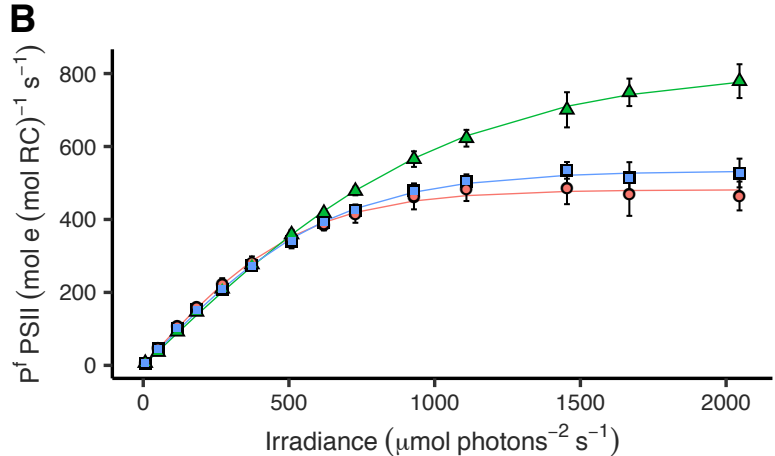
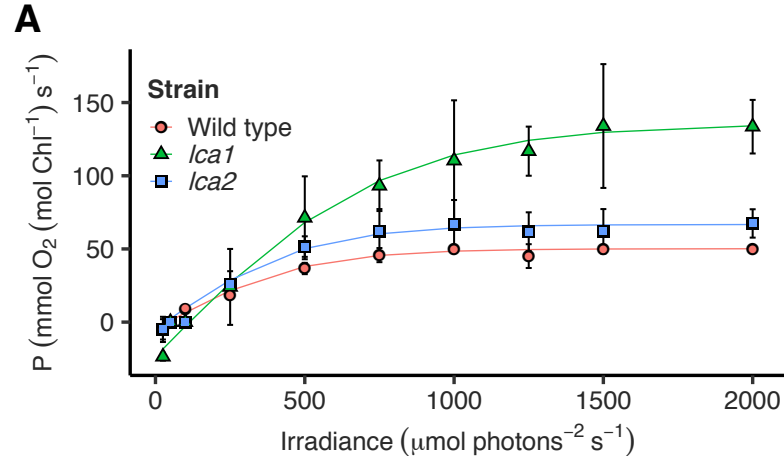
Supplemental Figure 3. Abundance of photosynthetic reaction centres and carbon fixation proteins. The relative proportions (A) of PSBA, PSAC and RBCL proteins in the *lca1* and *lca2* strains compared to WT as measured by Western blots (B) from protein extractions of cultures grown under medium light (100 mmol photons m⁻² s⁻¹). Error bars are one standard deviation from biological triplicates.

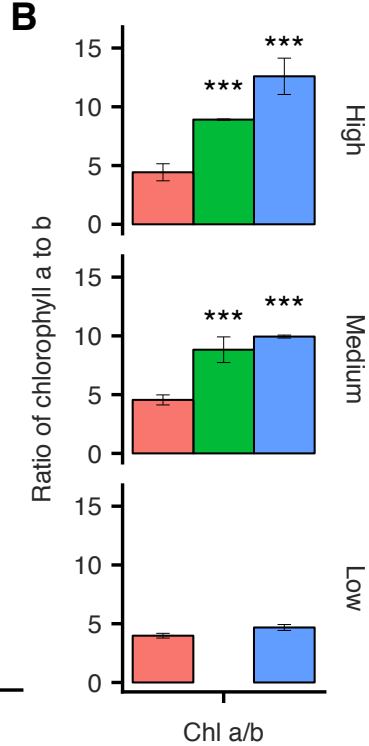
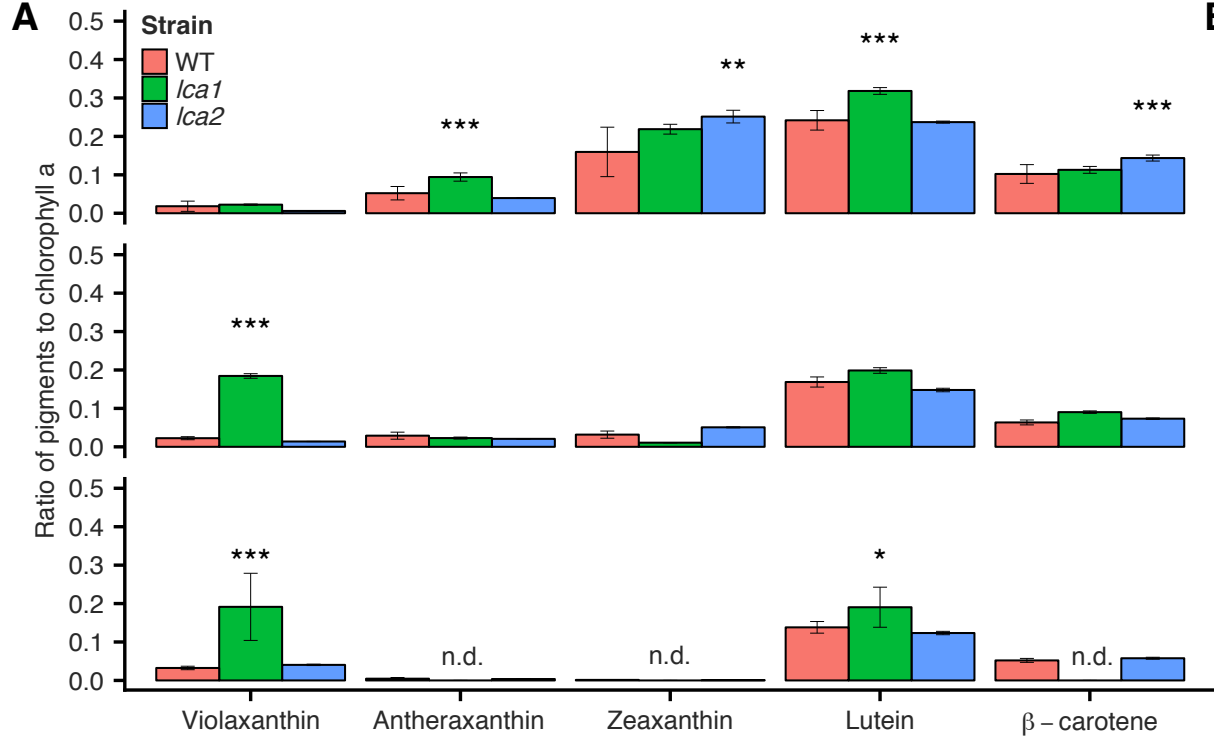
Supplemental Figure 4. The correlation in gene expression as measured using SuperSAGE and quantitative-PCR (qPCR). For each gene, FER2 (Ferredoxin-2), CSP (Chloroplast stem-loop-binding protein), FSA (Fructose-6-P aldolase), LHCII-2.1 (Major light-harvesting chlorophyll a/b protein 2.1), LHCB5 (Chlorophyll a-b binding protein CP26), LHCII-3 (Major light-harvesting chlorophyll a/b protein 3), PHOA (Starch phosphorylase) and RPI (Ribose-5-phosphate isomerase) the average result (N = 3) under low, medium and high light for WT, *lca1* and *lca2* is shown relative to WT under low light. The qPCR results are normalised against the *S11rRNA* gene. The result from a linear (least square) regression to all results is shown as a solid line together with the calculated Pearson correlation coefficient and p-value. The shaded area indicates the 95% confidence interval while the dashed lines indicate the 95% prediction interval.

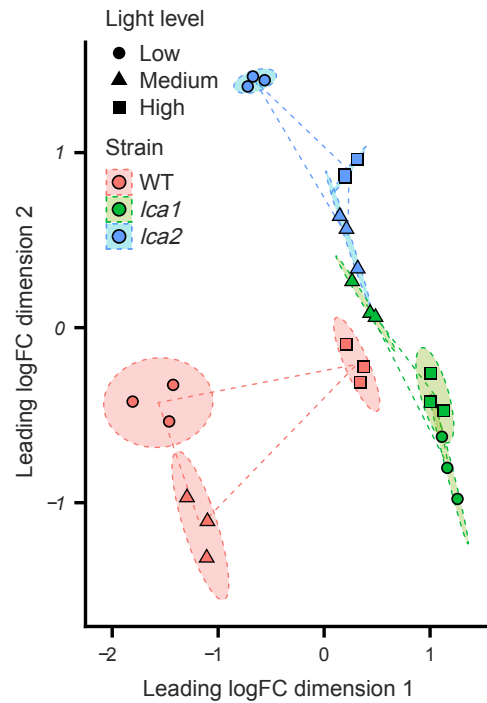
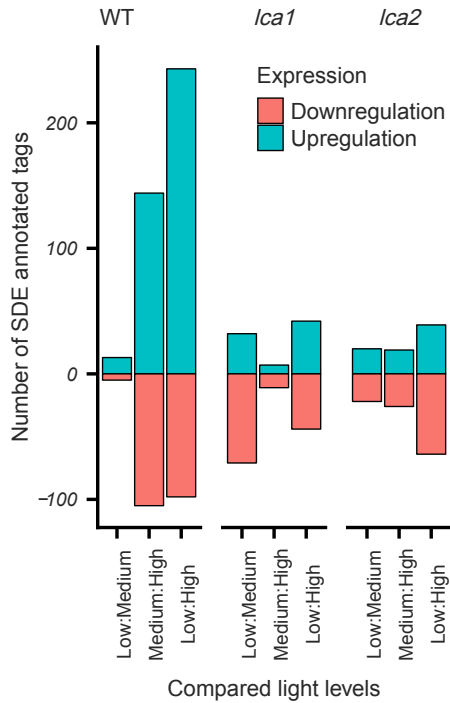
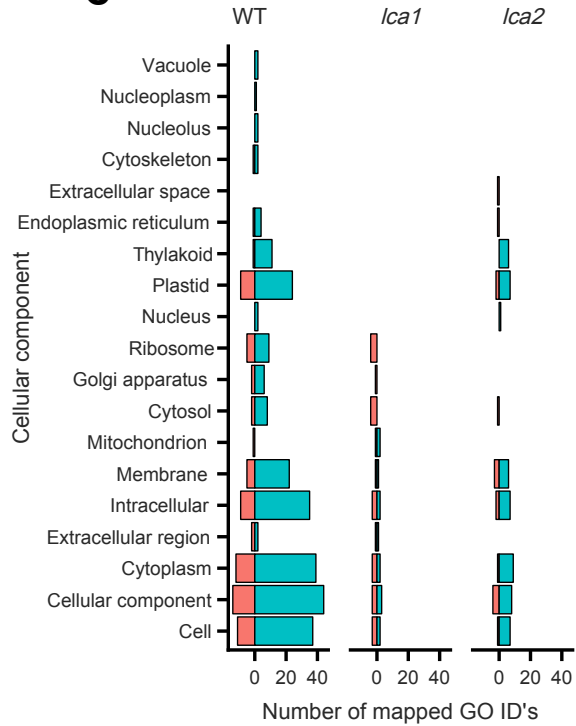
Supplemental Figure 5. Trended PSII functional cross-section and chlorophyll content. (A) Functional cross-section of PSII, (B) chlorophyll a content per cell, (C) chlorophyll b content per cell, (D) chlorophyll a:b ratio in WT, *lca1* and *lca2* strains

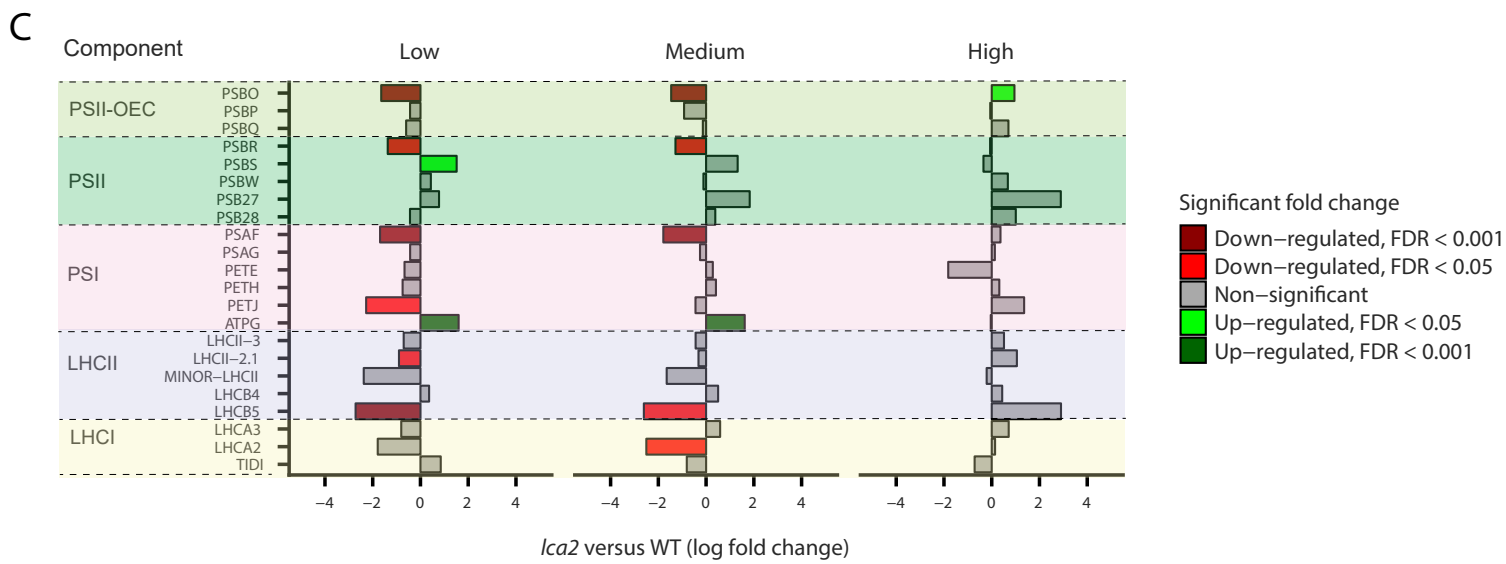
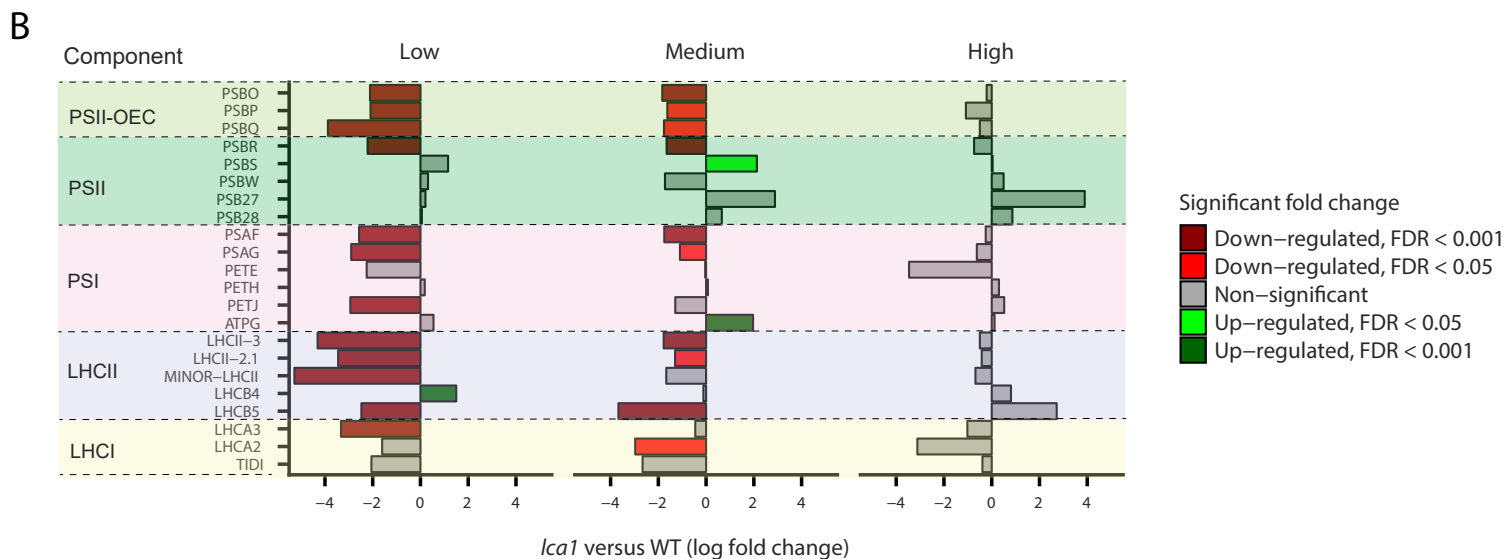
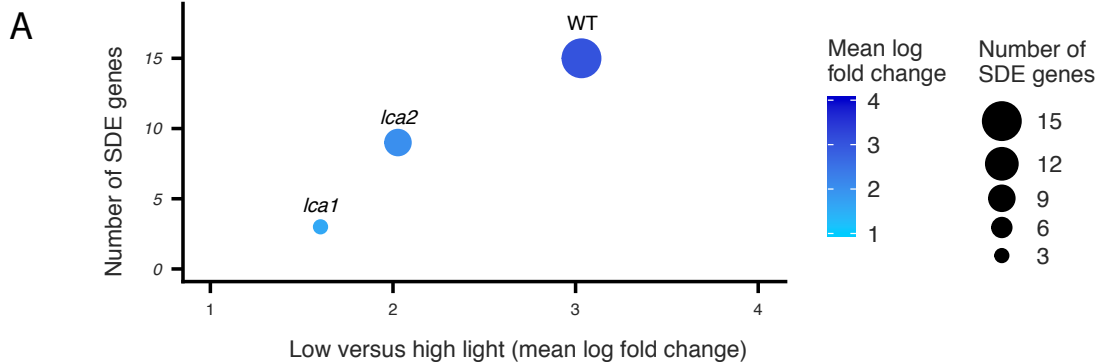
grown under low, medium and high light (20, 100, 1200 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$). Trended changes (dashed line) were calculated from linear regressions to the average result for biological triplicates under each light level. Error bars are one standard deviation.

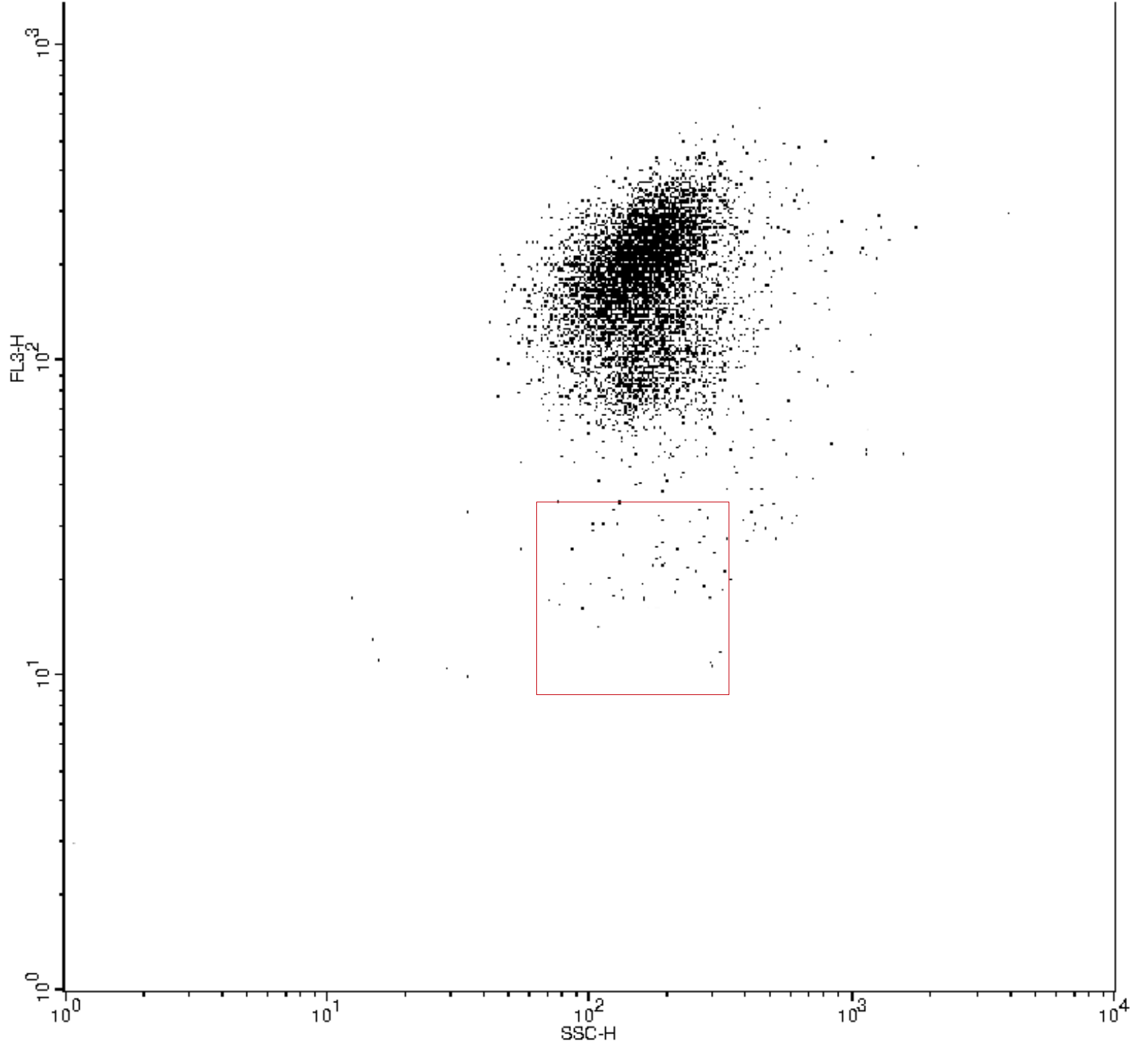


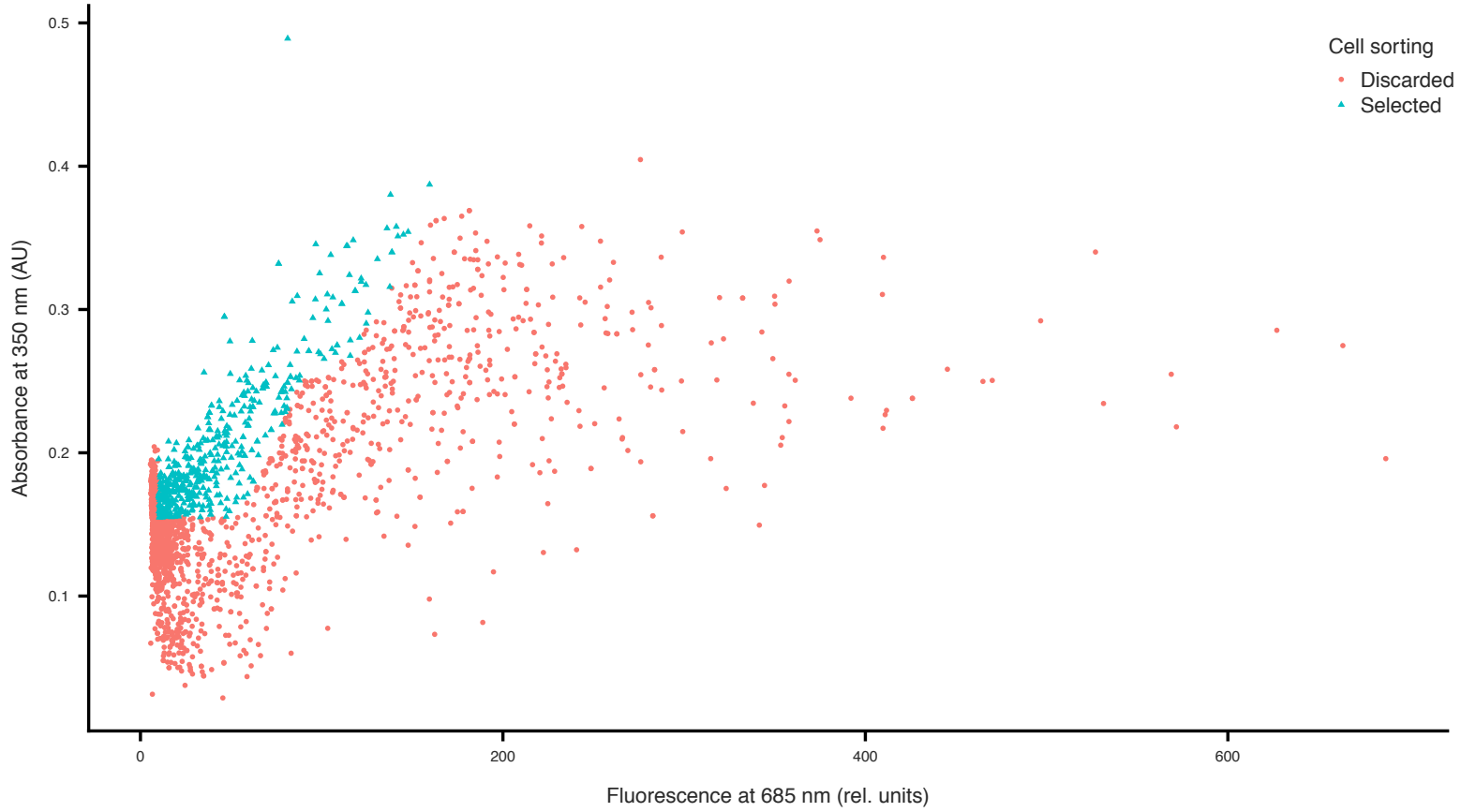


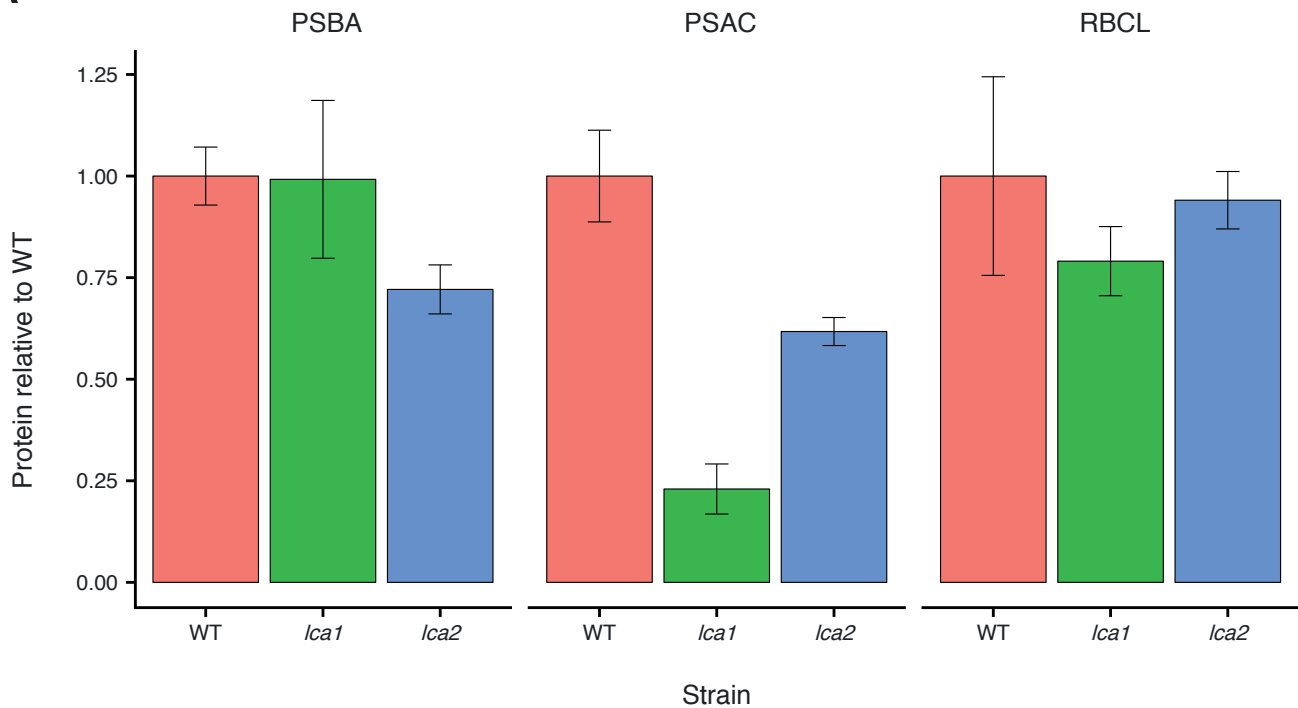


A**B****C**







A**B**