



## High-throughput insertional mutagenesis reveals novel targets for enhancing lipid accumulation in *Nannochloropsis oceanica*

Christian Südfeld<sup>a,\*</sup>, Michal Hubáček<sup>a</sup>, Daniel Figueiredo<sup>a</sup>, Mihris I.S. Naduthodi<sup>a</sup>, John van der Oost<sup>a</sup>, René H. Wijffels<sup>a,b</sup>, Maria J. Barbosa<sup>a</sup>, Sarah D'Adamo<sup>a</sup>

<sup>a</sup> Wageningen University, Bioprocess Engineering, PO Box 16, 6700 AA, Wageningen, Netherlands

<sup>b</sup> Faculty of Biosciences and Aquaculture, Nord University, N-8049, Bodø, Norway

### ARTICLE INFO

#### Keywords:

Microalgae  
Nannochloropsis  
Insertional mutagenesis  
FACS  
Neutral lipid content  
APETALA2

### ABSTRACT

The microalga *Nannochloropsis oceanica* is considered a promising platform for the sustainable production of high-value lipids and biofuel feedstocks. However, current lipid yields of *N. oceanica* are too low for economic feasibility. Gaining fundamental insights into the lipid metabolism of *N. oceanica* could open up various possibilities for the optimization of this species through genetic engineering. Therefore, the aim of this study was to discover novel genes associated with an elevated neutral lipid content. We constructed an insertional mutagenesis library of *N. oceanica*, selected high lipid mutants by five rounds of fluorescence-activated cell sorting, and identified disrupted genes using a novel implementation of a rapid genotyping procedure. One particularly promising mutant (HLM23) was disrupted in a putative APETALA2-like transcription factor gene. HLM23 showed a 40%-increased neutral lipid content, increased photosynthetic performance, and no growth impairment. Furthermore, transcriptome analysis revealed an upregulation of genes related to plastidial fatty acid biosynthesis, glycolysis and the Calvin-Benson-Bassham cycle in HLM23. Insights gained in this work can be used in future genetic engineering strategies for increased lipid productivity of *Nannochloropsis*.

### 1. Introduction

Microalgae have recently emerged as a promising platform for sustainable production of lipids, pigments and other bioactive compounds. In the last decade, the mostly marine microalgal genus *Nannochloropsis* has received considerable scientific and industrial attention as a suitable candidate for the production of high-value lipids and biofuel feedstocks (Liu et al. 2017; Chini Zittelli et al., 1999; Archambault et al., 2014; San Pedro et al., 2015; Ma et al., 2016). *Nannochloropsis oceanica*, in particular, exhibits robust outdoor growth at relatively high growth rates and it can reach neutral lipid contents of up to 50% of the dry cell weight (Ma et al., 2014; Li et al., 2014; Meng et al., 2015). However, these exceptionally high neutral lipid contents are only reached when *N. oceanica* is exposed to a stress condition that is unfavorable for cell division, for instance nitrogen (N) deficiency (Rodolfi et al., 2009). This poses an obstacle for achieving economically viable lipid productivities because biomass production rates are impaired under N deficiency. One way to reach high neutral lipid production without impairing growth is by genetic engineering (Naduthodi et al., 2021), but progress is

hampered by limited functional genome annotations, as research into the polyphyletic group of microalgae is relatively young (Rumin et al., 2020; Blaby-Haas and Merchant, 2019; Hanschen and Starkenburg 2020). About 50% of all predicted genes of the model strain *N. oceanica* IMET1 lack a functional annotation because the majority of the predicted proteins does not display sufficient sequence similarities with known proteins (Vieler et al., 2012; Wang et al., 2014). This poses a serious limitation for targeted modification of the metabolic network of this organism.

One well-established way to investigate the metabolic functions of unknown proteins in an unbiased manner *in vivo* is the use of forward genetic screens (Moresco et al., 2013; Beutler et al., 2006). This approach couples random mutagenesis to a phenotypic screening for traits of interest. After this screening, mutants with the desired characteristics are analyzed genetically and causative mutations are identified. This allows the researcher to map a phenotype and metabolic function to a protein. Typical mutagenesis strategies include chemical- or radiation-induced mutagenesis which lead to nucleotide substitutions in the host DNA. These mutations can induce changes in primary protein

\* Corresponding author.

E-mail address: [christian.sudfeld@wur.nl](mailto:christian.sudfeld@wur.nl) (C. Südfeld).

<https://doi.org/10.1016/j.ymben.2021.04.012>

Received 3 March 2021; Received in revised form 7 April 2021; Accepted 18 April 2021

Available online 7 May 2021

1096-7176/© 2021 The Authors. Published by Elsevier Inc. on behalf of International Metabolic Engineering Society. This is an open access article under the CC

BY license (<http://creativecommons.org/licenses/by/4.0/>).

sequences and gene expression patterns (Stadler 1928; Auerbach and Robson 1944). An advantage of these mutagenesis strategies is a desirably broad phenotypic diversity. However, these strategies rely on time-consuming and costly genotyping procedures, such as whole genome resequencing (Shelake et al. 2019). Moreover, detecting the causative mutations for a selected trait can be difficult due to several nucleotide substitutions in multiple different genes.

An alternative mutagenesis approach is insertional mutagenesis, which is the introduction of a foreign DNA molecule called insertion cassette (IC) into the genome of an organism at random positions (Austin et al., 2004; Ram et al., 2019). This approach has been the gold standard for elucidating gene functions in the context of forward genetics for a variety of organisms including microalgae (Greco et al., 2001; Alonso et al., 2003; Zhang et al., 2014). The insertion of the IC can be facilitated by enzymes such as transposases, or it can occur through nonhomologous end joining during DNA repair events at the site of double-stranded breaks. If the IC is inserted into a gene, the encoded protein will be either truncated or knocked out and thereby lose its metabolic function. The insertion site in strains with interesting characteristics can be identified through modified PCR procedures such as genome walking (Siebert et al., 1995) or RESDA-PCR (González-Ballester et al., 2005). Although this obviates the need for whole genome resequencing, the commonly employed methodologies for tracing insertions are time-consuming. Moreover, they rely on unpredictable factors, such as the presence of a specific restriction endonuclease (REN) recognition sequence in close proximity to the insertion site. Therefore, these methodologies will likely be too cumbersome to unravel the functions of the myriad of unknown genes in *N. oceanica*. An elegant variation of insertional mutagenesis was first reported for *Escherichia coli* (Goodman et al., 2009) and later adapted for *Chlamydomonas reinhardtii* (Zhang et al., 2014; Li et al., 2015). It involves using an IC that carries recognition sequences for a type IIS REN at both of its termini. Because this type of enzyme cuts in a defined distance outside of its recognition site, the IC can be precisely extracted together with a short defined stretch of adjacent genomic DNA from the genome of a transformed strain. This alleviates any uncertainty during genotyping PCRs.

Phenotypical screenings in forward genetics of microbial organisms often include high-throughput single cell methodologies such as fluorescence-activated cell sorting (FACS), which has gained an outstanding importance in the field of microalgal strain development (Hyka et al., 2013; Pereira et al., 2018). Based on flow cytometry, this technique allows quantification of cellular characteristics such as size and chlorophyll content through the light scattering or fluorescent properties of different cell constituents. Fluorescent probes such as BODIPY dye can be utilized to further quantify non-autofluorescent components of cells, for example lipid bodies. Lipid-rich microalgal strains have successfully been isolated by FACS for a variety of microalgal genera such as *Chlamydomonas* (Terashima et al., 2015), *Dunaliella* (Yao et al., 2017), *Chlorococcum* (Cabanelas et al. 2016) and *Nannochloropsis* (Doan and Obbard, 2012; S. Wang et al., 2016). Recently, first studies have shown that coupling insertional mutagenesis and FACS methodology can lead to the isolation of high lipid-producing *Nannochloropsis* strains (Osorio et al., 2019; Ryu et al., 2020). This has opened the door to sophisticated forward genetics screenings that will aid in unravelling gene functions in this organism at a larger scale. These screenings would greatly benefit from building upon a straightforward and reliable genotyping procedure.

The goal of this study was to identify novel genes that are associated with increased lipid accumulation phenotypes in *N. oceanica*. To achieve this, we developed a forward genetic screen based on insertional mutagenesis, featuring high-throughput phenotype screening using FACS, and a rapid and improved genotyping procedure using the type IIS REN MmeI.

## 2. Results and discussion

### 2.1. Construction and screening of a random insertional mutant library

We created a random insertional mutant library of *N. oceanica* by transforming 120 independent samples with an insertion cassette (IC). This IC (Fig. 1A) contains a zeocin resistance gene ( $zeo^R$ ) driven by the endogenous promoter  $P_{VCP}$  (Wang et al., 2016; Naduthodi et al., 2019). The IC further carries two transcriptional terminators in head-to-head orientation to ensure that genes at the insertion site will be knocked out or truncated regardless of cassette orientation. Terminal barcodes were added along with MmeI restriction endonuclease (REN) recognition sites, to enable rapid tracing of the insertion site by cassette PCR. Microalgal transformants were selected for resistance to Zeocin on agar plates, and ~25,000 colonies were pooled into a single flask to create the insertional mutant library (Fig. 1B).

The mutant library was repeatedly screened and enriched for high lipid mutant (HLM) strains by iterative sorting using FACS for 5x in total (Fig. 1C). Consecutive rounds of sorting can help to enrich a population of cells within a random mutant population (Cabanelas et al. 2016). Screening for HLM strains was achieved using the fluorescent BODIPY dye, which specifically and quantitatively stains neutral lipids (NLs) without affecting cell viability (Südfeld et al., 2020). The mutant library was grown to mid-exponential stage, stained with BODIPY, and cells with high NL content were selected by applying a 2-dimensional “high-NL” gate in the channels capturing BODIPY fluorescence and forward scatter (Fig. S1). The forward scatter was taken into account as a proxy for cell size to ensure that smaller HLM cells would be sorted as well. These cells may have only average absolute BODIPY fluorescence but higher BODIPY fluorescence than most cells of similar size. In the fifth round of sorting, single cells were deposited onto an agar plate to isolate strains from the enriched library.

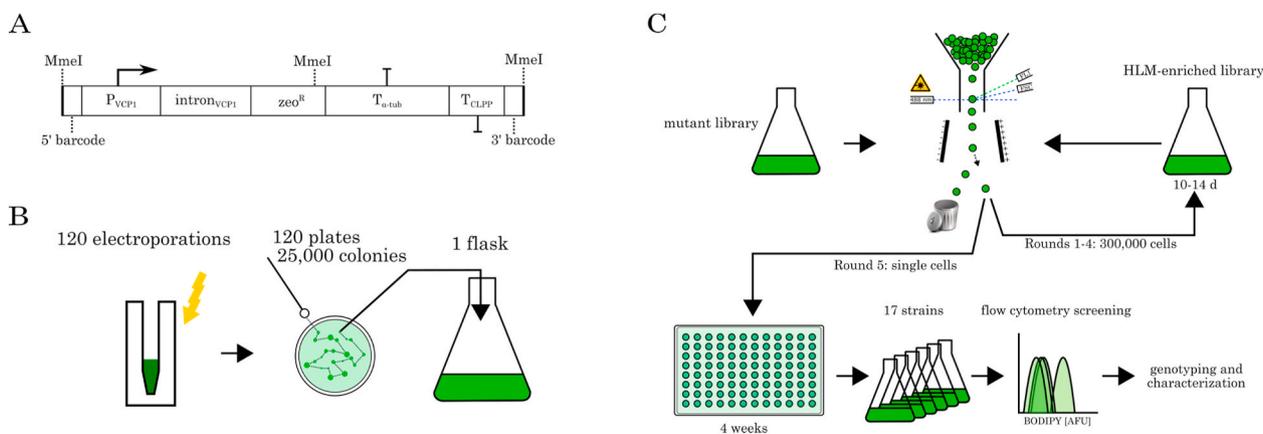
### 2.2. Screening and characterization of isolated high lipid mutant strains

After strain isolation, 17 colonies were further screened for NL content  $\text{cell}^{-1}$  to confirm that isolated strains were higher in NL content and to select candidates for in-depth analyses. Cultures were grown to mid exponential growth phase, stained with BODIPY and analyzed by flow cytometry. The medians of the BODIPY fluorescence distributions were used as an estimation of the average NL content  $\text{cell}^{-1}$  ( $N \geq 2$ , Fig. 2). Almost all strains showed an increase in average BODIPY fluorescence levels. Five strains with particularly high BODIPY fluorescence (highlighted red in Fig. 2) were selected for physiological and biochemical characterization.

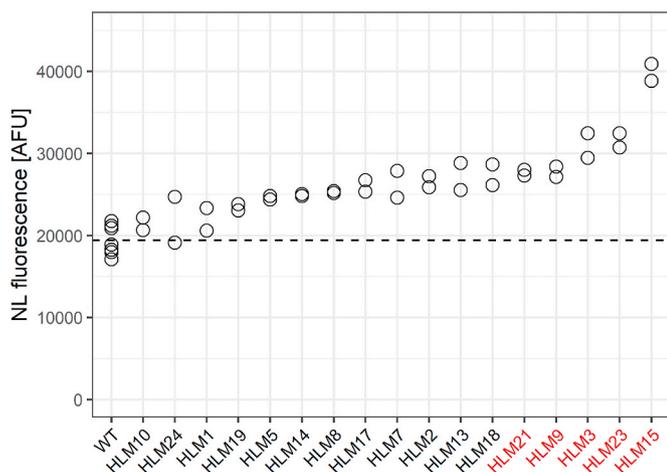
### 2.3. Characterization of promising strains

We compared the growth of five transformant strains and wild type in a batch cultivation with diurnal light cycles at  $600 \mu\text{mol m}^{-2} \text{s}^{-1}$  illumination intensity, using a parallel screening photobioreactor. All strains grew exponentially for 3 d and continued to grow linearly thereafter (Fig. 3A,  $N \geq 3$ ). HLM15, frequently collapsed after >3 d of cultivation and showed a 21% reduction in maximum specific growth rate compared to the wild type (Fig. 3B, Tab. S1). HLM9 showed a 10% reduction in maximum specific growth rate compared to the wild type, whereas growth rate of the remaining mutants was unaffected. HLM21 and HLM23 showed a higher maximum quantum efficiency of photosystem II (PSII) photochemistry ( $F_v/F_m$ ), compared to the wild type and HLM15 (Fig. 3C, Tab. S1). We observed that all HLM strains had larger cells compared to the wild type, and all but HLM9 showed increased chlorophyll *a* autofluorescence per cell (Fig. S2).

We quantified the neutral lipid (NL) and polar lipid (PL) contents and characterized the fatty acid composition in exponentially growing cultures of all strains (Fig. 4,  $N \geq 3$ ). NL contents  $\text{DCW}^{-1}$  were increased by



**Fig. 1. Creation and screening of an insertional mutant library.** (A) Schematic of the IC used to transform *N. oceanica*. The resistance gene  $zeo^R$  is driven by the endogenous VCP gene promoter and flanked by 2 transcriptional terminators in head-to-head orientation. Terminal barcode sequences and recognition sequences for the REN MmeI are needed for genotyping. (B) Wild type *N. oceanica* cells were transformed by electroporation by using 120  $\mu\text{g}$  of DNA in 120 separate reactions. Transformants were selected on zeocin-containing agar plates and  $\sim 25,000$  colonies were pooled into a single flask for screening. (C) Screening pipeline for enrichment and isolation of high lipid mutant (HLM) cells. The initial mutant library was harvested during exponential growth phase, stained with neutral lipid dye BODIPY, and 300,000 cells were sorted into a new flask by gating for high BODIPY fluorescence (Fig. S1). The HLM-enriched library was recovered for 10–14 d before the next sorting step. The library was enriched for HLM cells 4x before the fifth and final sorting during which strains were isolated by deposition of single cells onto agar plates. 17 colonies were selected for expansion and NL content screening by BODIPY staining and flow cytometry.



**Fig. 2. Cellular NL content of 17 mutant strains compared to the wild type.** Cellular NL contents were quantified using lipophilic fluorophore BODIPY and flow cytometry. Each point represents the median of single cell BODIPY fluorescence emission for an independent culture. The mean BODIPY fluorescence of wild type cultures is indicated by a dashed line. Mutant strains that were chosen for further analysis are highlighted red. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

40% in HLM9 ( $mean (M) = 0.302 \text{ g g}^{-1}$ ,  $standard deviation (SD) = 0.017$ ) and HLM23 ( $M = 0.302 \text{ g g}^{-1}$ ,  $SD = 0.016$ ) and by 58% in HLM15 ( $M = 0.341 \text{ g g}^{-1}$ ,  $SD = 0.047$ ) (Fig. 4A, Tab. S1) relative to the wild type ( $M = 0.215 \text{ g g}^{-1}$ ,  $SD = 0.014$ ). Increased NL accumulation in these strains did not occur at the expense of PL contents, which were similar for all strains (Fig. 4B, Tab. S1).

In addition to the increase in NL content  $\text{DCW}^{-1}$ , HLM23 and HLM15 showed an altered fatty acid composition (Fig. 4C). HLM23 had a 12% decreased fraction of monounsaturated fatty acids (MUFAs) in NLs, concomitant with a 55% increase in polyunsaturated fatty acids (PUFAs). The increased PUFA content was mainly due to an increase in the fraction of C20:4 n6 and C20:5 n3, known as eicosapentaenoic acid (EPA). EPA is considered a high-value molecule because it is an essential dietary compound and has potential in disease prevention (Calder 2006;

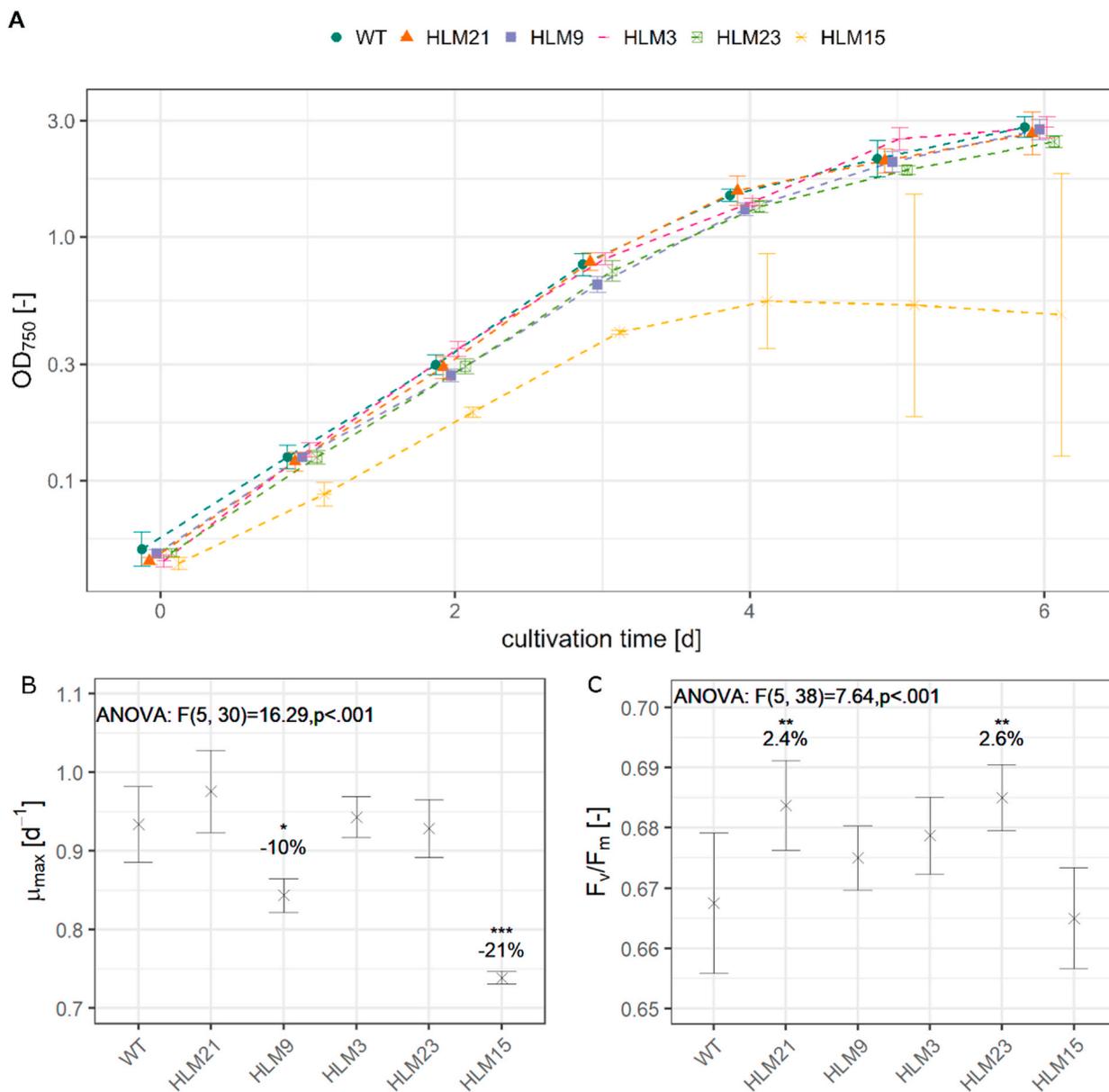
Mozaffarian and Wu 2011). In *Nannochloropsis*, EPA is mainly found in lipid classes such as MGDG or DGDG that are associated with the thylakoid membranes of the chloroplast (Dolch et al., 2017), whereas the fraction of EPA in NLs is negligible under favorable growth conditions. However, translocation of EPA to NLs can be observed during nitrogen stress in wild type *Nannochloropsis* (Janssen et al., 2019).

HLM15 showed increases of 24 and 116% for C14:0 and C18:1 (oleic acid), respectively, in the NL fraction. Oleic acid was further increased by 99% in the PL fraction, compared to the wild type. Oleic acid is a suitable candidate for enrichment in biodiesel due to its chemical properties (Knothe 2008).

#### 2.4. Tracing of cassette insertion sites in selected transformants

We traced the insertion sites in the nuclear genome of the five *N. oceanica* HLM mutants discussed above, by using a novel implementation of MmeI-based cassette PCR, which was inspired by the previously reported ChlaMmeSeq technique (Zhang et al., 2014; Lu et al., 2021). The procedure produces MmeI-restriction fragments that are ligated to an adapter molecule and amplified by cassette PCR (Fig. S3). Using a single REN and obtaining fragments with a constant length offer a substantial improvement over other versions of cassette PCR that rely e.g. on the presence of REN recognition sequences in the genomic DNA around the insertion site (Siebert et al., 1995; González-Ballester et al., 2005).

The previously reported method (Zhang et al., 2014) was altered to eliminate spurious amplification of unspecific templates (Fig. S3). These modifications include: 1) the use of a non-symmetrical adapter (Fig. S3D, blue), inspired by vectorette PCR (Arnold and Hodgson 1991), which prevents background noise due to amplification of genomic DNA fragments that have adapter molecules ligated to both ends; 2) An amine modification at the 3'-end of the shorter adapter strand that prevents the formation of an AP-template strand by extension of the adapter on DNA fragments without a barcode; 3) Use of synthetic barcode sequences at the cassette termini, to allow design of cassette-specific primers that have low similarity with endogenous genome sequences; 4) Nested PCR protocols (Fig. S3F) with touchdown temperature settings (Don et al., 1991; Zhang and Gurr, 2000) in both PCR cycles that further increase the stringency of target amplification. Using our adapted version of MmeI-based insertion site tracing, we were able to map the insertion sites of HLM21, HLM9, HLM3, HLM23 and HLM15 to five distinct



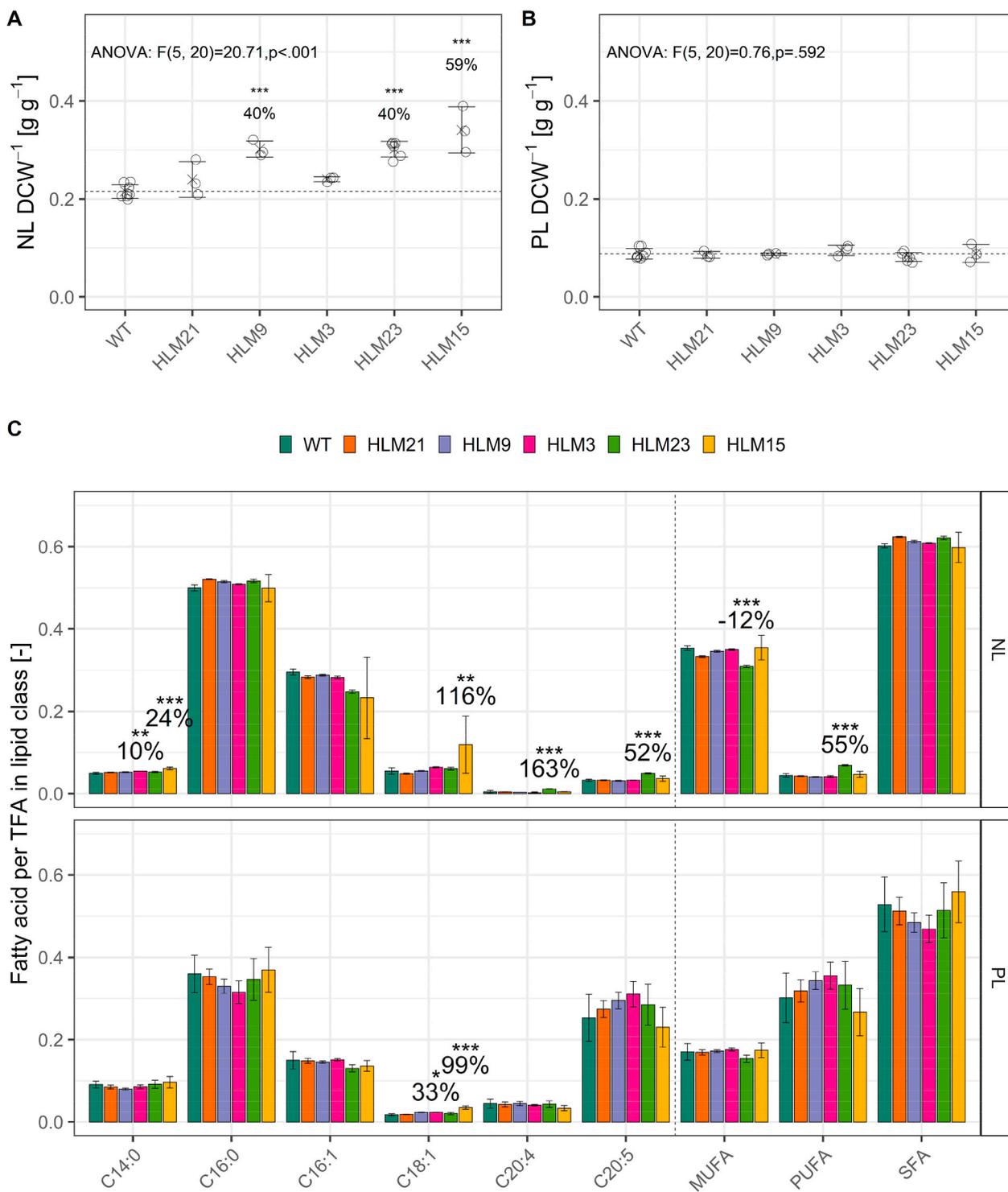
**Fig. 3. Physiological characterization of selected transformant lines.** (A) Growth curves of five transformants and wild type on a logarithmically-scaled y-axis ( $N \geq 3$ ). All cultures showed exponential growth for 3 d, until they reached an  $OD_{750}$  of  $\sim 0.8$ . HLM15 cultures were prone to collapsing after this period whereas all other strains displayed linear growth until they reached an  $OD_{750}$  of  $\sim 5$ –6 (graph truncated). Final biomass concentrations were comparable between cultures. Points are dodged horizontally for better distinction between samples. (B) Maximum specific growth rates ( $\mu_{max}$ ) of all analyzed strains were calculated for the exponential growth phase. The cross and errors denote the mean  $\pm$  SD ( $N \geq 3$ ). Growth rates of HLM3, HLM21, HLM23 were similar to the wild type, whereas growth rates of HLM9 and HLM15 were decreased by 10 and 21% respectively. (C) Maximum quantum efficiency of PSII photochemistry ( $F_v/F_m$ ) was comparable between the wild type and HLM3, HLM9 and HLM15. Increased photosynthetic performance was seen for HLM21 and HLM23. (B–C) Significant changes relative to the wild type are indicated above each group. Significance levels were assessed by Tukey’s HSD test in case of a significant ANOVA outcome. (\*):  $p < 0.05$ ; (\*\*):  $p < 0.01$ ; (\*\*\*) :  $p < 0.001$ .

genomic loci (Fig. 5). Putative insertion sites were confirmed by PCR with primers (arrows with half arrowheads in Fig. 5) complementary to the genome sequence of *N. oceanica* near the insertion sites.

#### 2.5. Cassette insertion in HLM21 occurred in a tetratricopeptide repeat domain-containing protein

In HLM21, the insertion cassette was integrated inside intron 4 of *NO18G02330* (Fig. 5A). Due to the transcriptional terminators present in the insertion cassette, the exons 5–8 of *NO18G02330* are likely not transcribed in HLM21, resulting in a loss of the 190 C-terminal amino acids in the corresponding protein. The uncharacterized *NO18G02330* has insufficient similarity to any known protein to predict its metabolic

function. We bioinformatically analyzed the *NO18G02330* primary protein sequence to predict protein localization and identify conserved domains. The protein does not contain any predicted target peptide, and it contains a tetratricopeptide repeat (TPR) domain. TPRs are found in proteins involved in a variety of metabolic functions, such as cell cycle coordination and transcriptional regulation (Hirano et al., 1990), protein folding (Lamb et al. 1995), transport across membranes (Brocard and Hartig 2006; Baker et al., 2007; Fransen et al., 2008), membrane assembly (Gatsos et al., 2008), vesicle fusion (Young et al. 2003) and bio-mineralization (Zeytuni et al., 2011). The TPR domain usually serves as an interface for interactions between proteins and multiprotein complex assemblies (Zeytuni and Zarivach 2012). Thus, it is likely that *NO18G02330* interacts with one or more proteins in *N. oceanica* and that

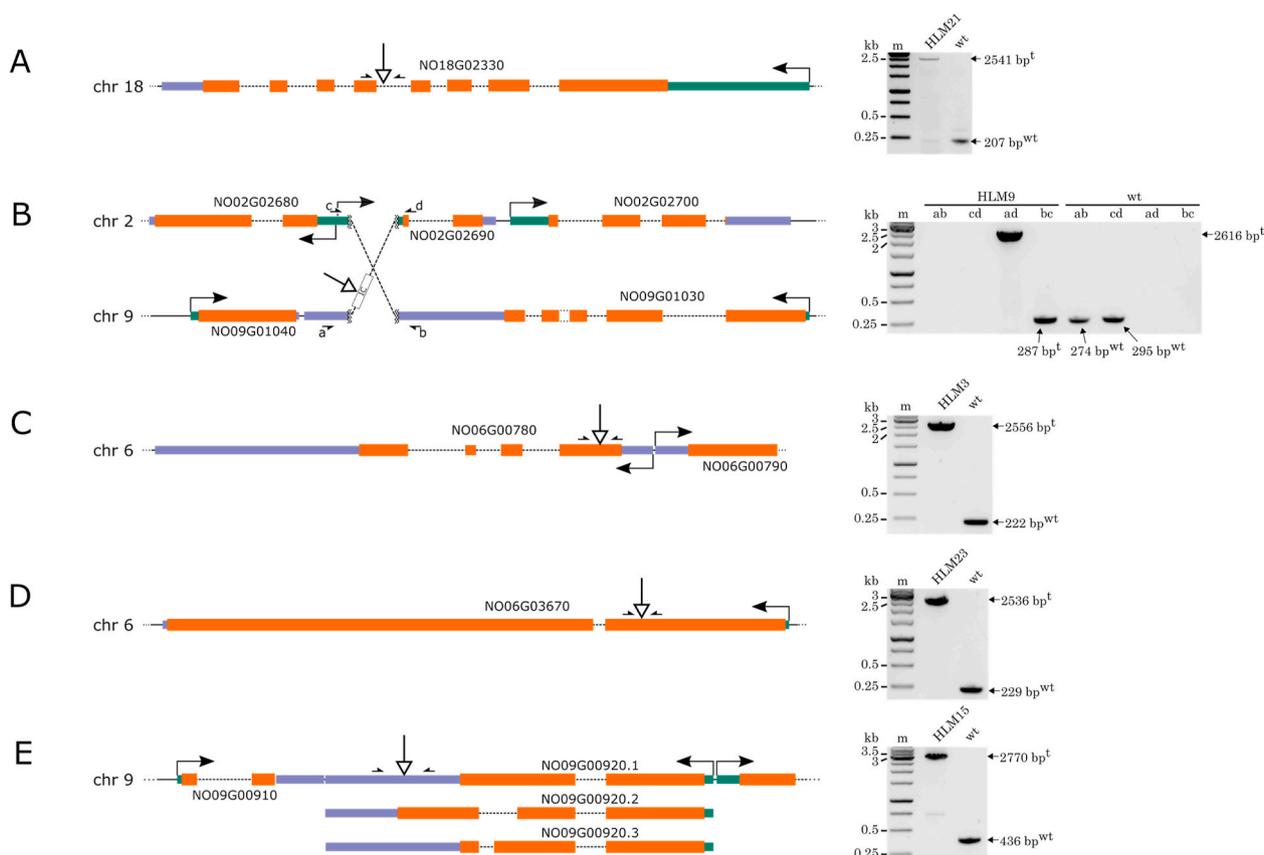


**Fig. 4. Biochemical characterization of selected transformant lines.** (A,B) The amount of neutral (A) and polar (B) lipids DCW<sup>-1</sup>. The superimposed cross and error bars denote the mean  $\pm$  SD. Changes in lipid content relative to the mean wild type lipid content (dashed lines) and significance levels are given above each group. All strains had either increased or unchanged NL content DCW<sup>-1</sup> compared to the wild type. Two-way ANOVA did not suggest a main effect of microalgal strain on PL content DCW<sup>-1</sup>. (C) Fatty acid composition of neutral lipids (NL) and polar lipids (PL) are shown for selected transformants. Bars and error bars denote the mean  $\pm$  SD of  $N \geq 3$  biological replicates. Changes relative to the wild type are shown for significant increases or decreases. (A–C) Significance levels were assessed by Tukey's HSD test. (\*):  $p<0.05$ ; (\*\*):  $p<0.01$ ; (\*\*\*):  $p<0.001$ . MUFA: monounsaturated fatty acids. PUFA: polyunsaturated fatty acids. SFA: saturated fatty acids.

this interaction is disrupted in HLM21. More detailed studies are necessary to verify this hypothesis, for instance by screening for possible interaction partners with the well-established yeast two-hybrid system (Cowell 1997; Erfelink et al., 2018).

#### 2.6. Cassette insertion in HLM9 occurred during chromosomal rearrangement

In HLM9 we traced the IC insertion to two different chromosomes for the 5' and 3' sequencing reactions. PCR analyses showed that a



**Fig. 5. Schematic of insertion sites in selected transformant strains and PCR verification of tracing.** (A) Insertion in HLM21 occurred inside of the fourth intron of *NO18G02330*. (B) Insertion in HLM9 occurred during reciprocal translocation of chromosomes 2 and 9. The IC (white-filled box labelled IC) was ligated between the two shorter fragments. The 5'-UTR of *NO02G02690* and the 3'-UTR of *NO09G01030* were disrupted during chromosomal rearrangement. Exon three of *NO09G03010* was shortened for clarity. a-d denote PCR primers that were used in combinations a+b (ab), c + d, a+d and b + c for verification of rearrangement. (C) Cassette insertion in HLM3 occurred inside of the first exon of *NO06G00780*. (D) Cassette insertion in HLM23 occurred inside of the first exon of *NO06G03670*. (E) Cassette insertion in HLM15 occurred in *NO09G00920*. The last exon of isoform 2 and the 3'-UTR of isoforms 1 and 3 were disrupted by the insertion. Gene exons are depicted in red, putative 5' and 3'-UTR sequences in green and blue respectively. Dashed lines denote introns. Solid kinked arrows show putative transcription start sites. Arrows with white-filled arrowheads show the position of cassette insertion. Arrows with solid, half arrowheads show primer binding sites during PCR reactions that were done to verify correct tracing. Kilobase (kb) sizes of marker (m) bands are indicated on the left side of gel images. Markings on the right and bottom side of gel images indicate the expected band size for transformant (t) and wild type (wt) samples. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

reciprocal translocation t(2; 9) had occurred after double-stranded breaks at position 739,793 of chromosome 2 (total length 1,674,082 bp) and position 349,677 of chromosome 9 (total length 1,290,044 bp). The longer fragments of both chromosomes were each ligated to one side of the IC (Fig. 5B), creating a new, larger chromosome, named here derivative chromosome 9 (total length 1,876,990 bp). The two shorter fragments were ligated to each other creating a new, shorter chromosome, named here derivative chromosome 2 (total length 1,089,470 bp). The chromosome number of a derivative chromosome would normally be assigned based on the chromosome number that donated the centromere, but the positions of centromeres are not known for *N. oceanica*. The designated derivative chromosome numbers are therefore putative. Whereas chromosomal translocations can affect global gene expression in an unpredictable manner, e.g. in mammalian cells (Harewood et al., 2010), they have a more concrete effect on expression of the genes located at the breakage, for instance by causing gene knockouts or knockdowns or by creating novel fusion genes.

In HLM9, the putative promoter region and/or the 5'-UTR of *NO02G02690* and the 3'-UTR of *NO09G01030* are disrupted by the translocation and IC insertion (Fig. 5B). According to transcriptomic analyses (Li et al., 2014; Shi et al., 2020; Wei et al., 2019), *NO02G02690* is a low expressed gene encoding a protein of unknown function that has no homologues outside of the *Nannochloropsis* genus. Orthologues in

*N. gaditana* (*Naga\_100026g34*, E value =  $2 \times 10^{-16}$ , 55% sequence identity on 55% coverage using BLASTp) and *N. salina* (*NSK\_001029*, E value =  $3 \times 10^{-11}$ , 60% sequence identity on 41% coverage using BLASTp) are hypothetical proteins. *NO09G01030* encodes a fungal-type Zn(II)<sub>2</sub>Cys<sub>6</sub> DNA-binding, Zinc finger (ZF) domain protein, which has high amino acid sequence similarity with the *Nannochloropsis gaditana* transcription factor *Naga\_100104g18* (E value = 0.0, 65.59% identity on 72% sequence coverage using BLASTp, 44.46% identity on full length sequences using CLUSTALΩ). *Naga\_100104g18* was previously identified as a control hub for lipid content in *N. gaditana* (Ajjawi et al., 2017). Ajjawi and co-workers found that a knockdown of *Naga\_100104g18* increased lipid content by ~2-fold and resulted in a 5–15% reduction in growth compared to the wild type. Similarly, HLM9 displayed a 9% reduction in growth and 40% increase in NL content compared to the wild type in our experiments, which suggests that *NO09G01030* and *Naga\_100104g18* may be functional homologues and that *NO09G01030* expression may be decreased in HLM9. In their study, Ajjawi and co-workers inserted an antibiotic resistance cassette into the 3'-UTR of *Naga\_100104g18*, which reduced transcript abundance by 50%. This drastic decrease exemplifies the importance of an intact 3'-UTR for gene expression in *Nannochloropsis*. Ajjawi and colleagues inserted the antibiotic resistance cassette at a position 30 nucleotides downstream of the translational STOP codon, which replaced ~98% of the exceptionally

long 3'-UTR of *Naga\_100104g18* with heterologous sequence. Opposed to this, the truncation of the 3'-UTR of *NO09G01030* in HLM9 occurred much further away from the translational STOP codon of the gene, leaving ~70% of the endogenous 3'-UTR sequence intact (Fig. 5B).

## 2.7. Cassette insertion in HLM3 occurred in a Zn(II)-finger type protein of unknown function

Cassette insertion in HLM3 disrupted *NO06G00780* in the first exon (Fig. 5C). *NO06G00780* is a relatively short protein (190 AAs) without predicted subcellular compartmentalization that carries a ZF-HIT domain at its C-terminus. Proteins containing this domain are mostly nuclear proteins associated with gene regulation, chromatin remodeling or pre-snoRNP assembly in multimodal RNA/protein complexes (Bragantini et al., 2016). Their precise functions, however, remain largely unknown. *NO06G00780* has moderate-high similarity (E value  $\geq 2 \times 10^{-41}$ ) with a variety of proteins mostly from filamentous protists belonging to the subphylum Oomycota. Oomycota is a sister-group to the *Nannochloropsis*-containing subphylum Ochrophyta, and consists of non-photosynthetic organisms including the extensively researched parasite *Phytophthora infestans* (Cavalier-Smith, 2018; Derelle et al., 2016). A few of these proteins are putatively classified as  $\text{Ca}^{2+}$ :cation antiporters such as the *Phytophthora palmivora* protein POM74018.1 (E value =  $2 \times 10^{-33}$ , 43% sequence identity on 94% coverage using BLASTp). However, none of the putative homologues that were identified by BLASTp analysis are functionally characterized.

## 2.8. Cassette insertion in HLM15 occurred in a DnaJ (HSP40) gene

The cassette insertion in HLM15 occurred in *NO09G00920*, which encodes 3 isoforms of a DnaJ type B protein. DnaJ proteins, otherwise known as HSP40, are part of a molecular chaperone system that is present in all eukaryotic cells. This chaperone system, consisting of HSP40, HSP70 and nucleotide exchange factors, plays an essential role in protein homeostasis by governing protein folding, unfolding, translation, degradation, translocation and protein disaggregation (Youker and Brodsky 2007; Young 2010; Ahmad et al., 2011). DnaJ/HSP40 family members are obligate co-chaperones to HSP70 as they deliver substrate to HSP70 and stimulate its ATPase activity (Qiu et al., 2006; Ghazaei 2017). As molecular chaperones, DnaJ/HSP40 proteins are pivotal for protein homeostasis especially under stress conditions. DnaJ loss of function mutants of *E. coli* e.g. fail to grow at elevated and reduced temperatures but show less drastic phenotypes at optimal growth conditions (Genevaux et al., 2002). Severe growth defects, temperature sensitivity or lethal phenotypes were also observed for knockouts of different cytosolic DnaJ variants in *Saccharomyces cerevisiae* (Sahi and Craig 2007). In *Arabidopsis thaliana*, the knockout of a nuclear-encoded chloroplastic DnaJ family member induced altered photosynthetic behaviour and triggered a global stress response in transgenic plants including expression of genes related to ROS detoxification. A chloroplastic DnaJ protein was also disrupted in a chemically mutagenized *N. gaditana* strain that showed a light-dependent increase in lipid content of up to ~60%  $\text{DCW}^{-1}$  and a changed proton motive force at the thylakoid membrane (Cecchin et al., 2020). However, this *N. gaditana* strain carried mutations in 233 additional genes, which prevents drawing of reliable conclusions on a direct link between the DnaJ mutation and the observed phenotype.

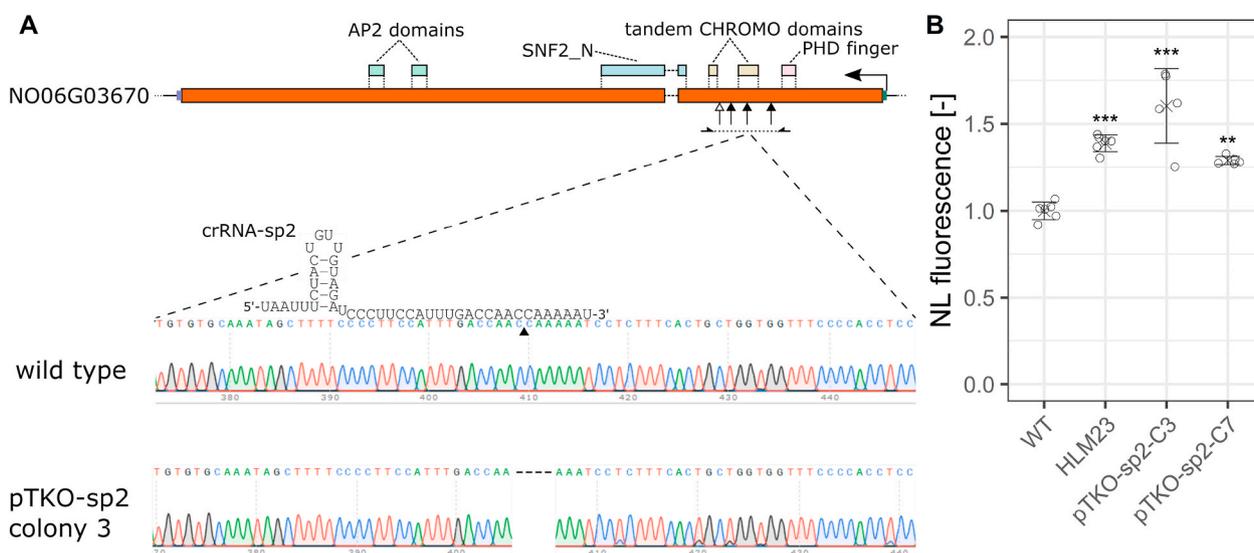
Based on the high functional conservation of DnaJ proteins, *NO09G00920* likely aids protein homeostasis in *N. oceanica* and it may play a crucial role in stress response. A functional impairment of *NO09G00920* in HLM15 would explain the 21% decrease in growth rate observed for this strain, as substantial growth defects were also observed for DnaJ mutant strains of other organisms. A function of *NO09G00920* in stress response is further suggested by HLM15's propensity for collapsing after prolonged incubation under high light conditions

(Fig. 3A). Generally, high light conditions cause an overreduction of the photosynthetic electron transport chain in phototrophs. This leads to the formation of reactive oxygen species (ROS) (Mittler 2002), which can oxidize proteins. Oxidized proteins must be degraded by the cell to prevent toxic effects (Xiong et al., 2007). An impaired activity of *NO09G00920* in HLM15 could jeopardize the strain's ability to cope with protein aggregation or unfolded protein stress, which offers a potential explanation for its decrepit growth and increased high light susceptibility. Unfolded protein stress is further linked to lipid biosynthesis. Mitigation of unfolded protein stress involves activation of the unfolded protein response (UPR) pathway that is tied to TAG production and lipid droplet formation in *S. cerevisiae* (Fei et al., 2009), animal cells (Basseri and Austin 2012), higher plants (Shank et al., 2001) and microalgae (Yamaoka et al., 2019). The increased NL content and decreased growth rate of HLM15 may therefore be the consequence of an elevated level of unfolded protein stress due to the decreased chaperone activity. Similarly, *C. reinhardtii* mutants with an impaired UPR grew slower and contained more TAG than the parental strain when treated with the unfolded protein stress inducer tunicamycin (Yamaoka et al., 2019).

According to the representative version of the *N. oceanica* IMET1 genome (Gong et al., 2020), *NO09G00920* is expressed in 3 different splice variants (*NO09G00920.1*, *NO09G00920.2* and *NO09G00920.3*) in *N. oceanica*. The existence of *NO09G00920.2* is strongly supported by high similarity of this protein with known DnaJ proteins of other organisms over the full amino acid sequence length. However, the C-termini of the predicted *NO09G00920* isoforms 1 and 3 show no similarities with any known protein in the NCBI database, suggesting that they may be misannotated. *NO09G00920* contains an N-terminal J domain, which is responsible for the interaction of DnaJ/HSP40 proteins with HSP70, whereas the C terminus of the protein may be responsible for substrate specificity and functional diversity (Vos et al., 2008). Depending on the splice variant, the insertion of the IC in HLM15 happened in either the 3'-UTR (for the putative isoforms *NO09G00920.1* and *NO09G00920.3*), or, in case of *NO09G00920.2*, inside the last putative exon of the CDS, 30 nucleotides away from the translational STOP codon. This results in a substitution of the 10 C-terminal amino acids (AAs) of *NO09G00920.2* with the dipeptide Trp-Ile. It is unclear whether this mutation would have a detrimental effect on the activity of the 413 AA-containing protein, but a conserved domain of unknown function (DUF1977) is present at the very C-terminus of isoform *NO09G00920.2* and this domain is truncated by the insertion. The 3' end of the IC further replaces the original 3'-UTR of *NO09G00920.2* and the biggest part of the original 3'-UTR of *NO09G00920.1* and *NO09G00920.3* in HLM15 (Fig. 5E). Gene expression may be severely affected for all putative isoforms because the 3'-UTR plays a crucial role in modulating gene expression by regulating transcript localization, stability and translational efficiency (Huang and Teeling 2017). The substantial increase in NL content in HLM15 suggests that DnaJ proteins are potential targets for genetically engineering increased lipid productivities in microalgae. Despite this, we must point out that interfering with mechanisms that ensure protein homeostasis can negatively influence biomass productivity and culture stability under stress conditions, which is a disadvantage for industrial production processes.

## 2.9. Cassette insertion in HLM23 occurred in an APETALA2-like transcription factor

In HLM23, IC insertion occurred in the first exon of *NO06G03670*. We identified conserved domains in *NO06G03670* by querying the conserved domain database (Lu et al., 2020). *NO06G03670* contains two AP2 domains (Fig. 6A), which are conserved domains involved in DNA binding in the plant superfamily of transcription factors AP2/EREBP, and an SNF2 domain and adjacent tandem chromo domains that are together found in ATP-dependent chromatin remodelers. Moreover, an N-terminal plant homeodomain (PHD) type Zn-finger may aid in interaction with histones or other proteins. In HLM23, *NO06G03670* is



**Fig. 6.** CRISPR/Cas-mediated knockout of *NO06G03670* in wild type *N. oceanica* re-instates the HLM23 phenotype. (A) Schematic of the domain architecture of *NO06G03670* and illustration of knockout approach using Cas12a. The insertion site (white-headed arrow) of the IC in HLM23 lies between tandem CHROMO domains within the first exon of the gene. An SNF2\_N and AP2 domains that are required for recognition of target DNA sequences by the transcription factor, are located between the insertion site and the C-terminal end of the protein and are not present in the truncated *NO06G03670* of HLM23. Three crRNA protospacers were designed to target the first *NO06G03670* exon at different positions (black-headed arrows) between the HLM23 insertion site and the translation start site. The homology between crRNA-sp2 and the target region of wild type *NO06G03670* nucleotide sequence is shown exemplarily. pTKO-sp2 colonies 3 and 7 had a four nucleotide deletion at the expected position for Cas12a-mediated cleavage (solid black triangle), which causes a shift of the genetic reading frame. Chromatograms illustrate Sanger sequencing data of PCR products obtained from genomic DNA of the wild type strain and pTKO-sp2 colony 3, using primers located in the first exon of *NO06G03670* (solid black half-headed arrows). (B) NL contents of HLM23 and two CRISPR/Cas-induced *NO06G03670* knockout mutants were increased similarly relative to the wild type. NL contents were measured by BODIPY staining and flow cytometry. Medians of the single cell BODIPY fluorescence distributions were normalized to the average wild type value. Crosses and error bars denote mean and SD ( $N \geq 5$ ). Significance levels were assessed by Tukey's HSD test. (\*):  $p < 0.05$ ; (\*\*):  $p < 0.01$ ; (\*\*\*)  $p < 0.001$ .

rendered non-functional because the protein is truncated after the PHD domain and therefore misses the domains responsible for DNA-binding and chromatin remodeling.

In a computational study, *NO06G03670* has been identified as one of three putative orthologues of the transcription factor AtWRI1, which is related to lipid metabolism in *A. thaliana* (Hu et al., 2014). This gene encodes a transcription factor belonging to the family of APETALA2-like proteins from higher plants and falls into the transcription factor superfamily AP2/EREBP (Kong et al. 2019). In plants, AP2/EREBP family members are reported to be involved in stress response, sugar metabolism, hormone signaling, and the coordination of developmental processes such as floral organ identity determination and seed germination (Riechmann and Meyerowitz 1998; Ohto et al., 2005; Xie et al., 2019). In their pioneering work, Hu and co-workers identified numerous transcription factor genes and transcription factor binding sites in *Nannochloropsis* by *in silico* analyses, and constructed a preliminary global regulatory network (Hu et al., 2014). *NO06G03670* was predicted to modulate transcription of 74 and 69 genes positively and negatively respectively, by binding to the DNA motifs 5'-CGCGCCAW-3', 5'-TCCGCCCM-3' and 5'-GCCSATCC-3' in promoter sequences. These motifs are enriched in genes related to chromosome organization, protein modification and cofactor metabolic processes, respectively. Among the genes predicted to be regulated by *NO06G03670* are several genes related to fatty acid biosynthesis.

Table 1 summarizes the phenotypes and genotypes associated with the HLMs that were investigated more closely in this study.

### 2.10. Targeted knockout of *NO06G03670* in wild type *N. oceanica* confirms the HLM23 phenotype

Because mutant HLM23 showed the highest potential for lipid-overproduction in the here presented experiments (Table 1), we selected this strain to further confirm the connection between its

**Table 1**

**Summary of transformant pheno- and genotypes.** All values are increases relative to the wild type strain for a one-step batch cultivation. The NL productivity increase is a projection based on the maximum specific growth rate ( $\mu_{max}$ ) and NL content during exponential growth phase. Significance levels in parentheses were calculated by Tukey's HSD test. (ns): not significant,  $p > 0.05$ ; (\*):  $p < 0.05$ ; (\*\*):  $p < 0.01$ ; (\*\*\*)  $p < 0.001$ .

Strain	$\mu_{max}$	NL DCW <sup>-1</sup>	NL productivity	Affected gene	Protein type
HLM21	6% (ns)	12% (ns)	19%	NO18G02330	unknown function
HLM9	-10% (*)	40% (***)	26%	NO09G01030/ NO02G02690	Zn(II) <sub>2</sub> Cys <sub>6</sub> fungal type TF
HLM3	1% (ns)	12% (ns)	13%	NO06G00780	Zn finger type
HLM23	-1% (ns)	40% (***)	39%	NO06G03670	APETALA2-like TF
HLM15	-21% (***)	59% (***)	26%	NO09G00920	DnaJ type II

genotype and phenotype. To do so, we employed a plasmid-based CRISPR-Cas mutagenesis strategy, which was recently developed for *Nannochloropsis* (Poliner et al. 2018) [Mihris Naduthodi, personal communication]. Using this strategy, genes can be knocked out in a highly controlled manner, with the help of a plasmid that encodes a constitutively-expressed Cas12a protein and a crRNA sequence. Upon plasmid delivery into cells, this system creates double-stranded breaks in a 5'-proximal gene exon. Repair of double-stranded breaks by non-homologous end joining can lead to frameshift mutations that can be identified by DNA sequencing (Seol et al. 2018). Frameshift mutations are considered a simple and robust strategy to produce targeted genetic knockouts (Van Campenhout et al., 2019).

By employing this approach, we designed three crRNAs with

protospacer sequences (sp1, sp2 and sp3) that are complementary to different positions of the first exon of *NO06G03670*. The positions were chosen upstream of the IC insertion site in HLM23 (Fig. 6A), to ensure that the resulting truncated *NO06G03670* proteins would not contain any additional domains compared to the truncated *NO06G03670* of HLM23. The protospacer sequences sp1, sp2 and sp3 were incorporated into a Cas12a-encoding plasmid, generating pTKO-sp1, pTKO-sp2 and pTKO-sp3, respectively. The three plasmids were used to transform wild type *N. oceanica*, and ten mutant colonies per construct were screened by PCR and sequencing (Fig. 6A). 0/10, 2/10 and 0/10 colonies with nucleotide mutations were recovered for pTKO-sp1, pTKO-sp2 and pTKO-sp3, respectively. The pTKO-sp2 colonies three (pTKO-sp2-C3) and seven (pTKO-sp2-C7) showed a deletion of *NO06G03670* nucleotides 1781–1784 (CCAA). This deletion had occurred around the position expected for Cas12a-mediated cleavage, corresponding to nucleotides 18–21 of protospacer sp2 (Zetsche et al., 2015). The four nucleotide deletion causes a shift of the genetic reading frame, rendering *NO06G03670* non-functional in pTKO-sp2-C3 and pTKO-sp2-C7.

We examined the NL content of the confirmed *NO06G03670* knockout mutant strains pTKO-sp2-C3 and pTKO-sp2-C7. When grown at high light conditions, the strains had NL contents similar to that of HLM23, and significantly higher than that of wild type controls (Fig. 6B). This confirms that the high lipid phenotype of HLM23 is connected to the knockout of *NO06G03670*, and underpins the relevance of this gene as a control hub for growth-associated NL production

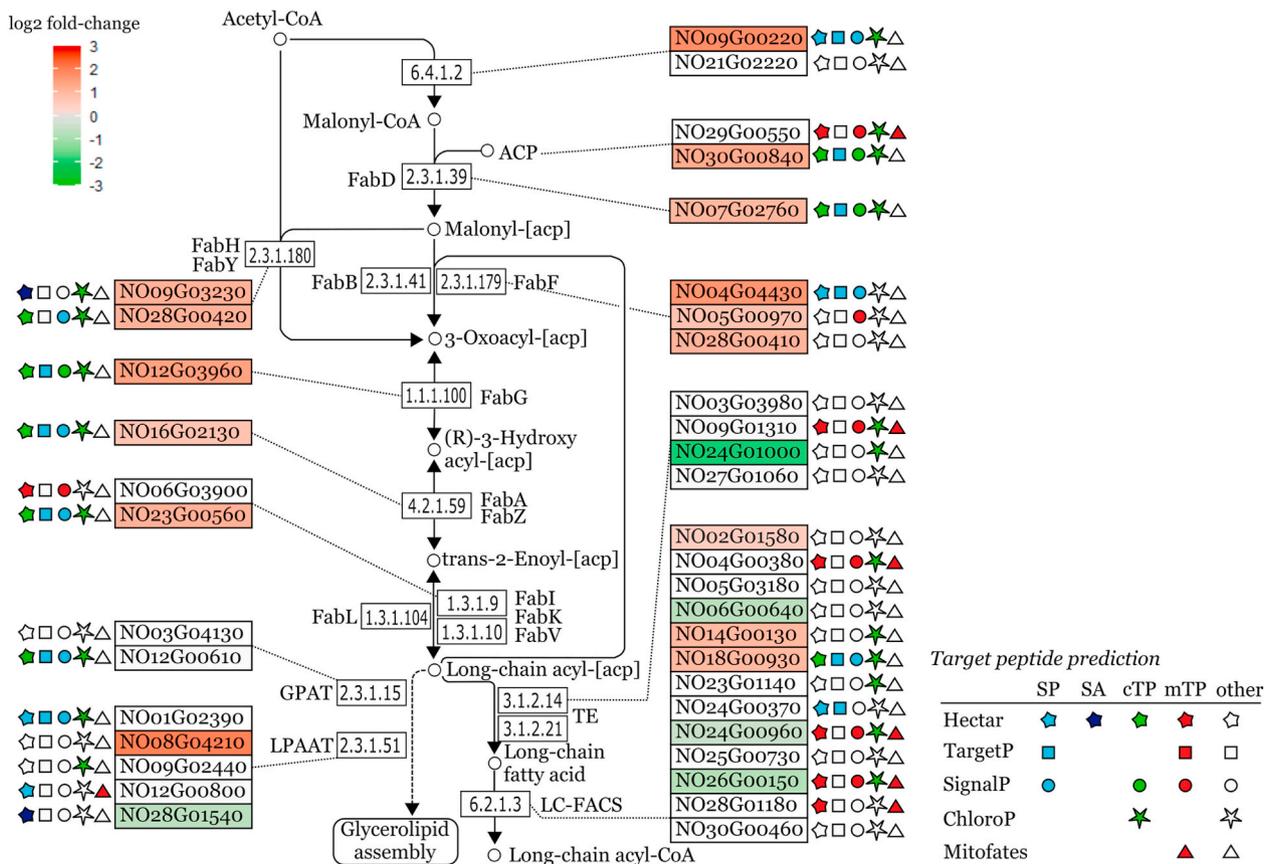
in *N. oceanica*. The connection between genotype and phenotype of other HLMs could be investigated accordingly during follow-up research.

### 2.11. Knockout of *NO06G03670* causes upregulation of metabolic pathways that are associated with fatty acid biosynthesis

In a computational study involving analysis of conserved transcription factor binding site motifs, Hu and colleagues predicted *NO06G03670* to attenuate the expression of genes related to fatty acid biosynthesis (FAB) and long chain fatty acyl-CoA synthetases (LC-FACS). LC-FACS activate fatty acids for different metabolic processes such as trafficking, lipid assembly or beta oxidation (Li-Beisson et al., 2010). In order to understand why the knockout of *NO06G03670* leads to an increase in lipid content, we analyzed gene expression in HLM23 by mRNA sequencing technology. Out of 9578 expressed genes, 1530 and 1495 genes were transcriptionally up- and downregulated, respectively, in HLM23, relative to the wild type (Fig. S4).

#### 2.11.1. Fatty acid biosynthesis genes are transcriptionally upregulated in HLM23

Plants and microalgae possess a prokaryotically-derived type II FAB pathway in their chloroplasts, which utilizes discrete, monofunctional enzymes to synthesize straight-chain fatty acids (Fig. 7). Out of 12 FAB genes that putatively encode these monofunctional enzymes, 10 were



**Fig. 7. Differential expression of genes encoding putative fatty acid biosynthesis enzymes in HLM23.** The pathway map illustrates the enzymatic conversion of acetyl-CoA to fatty acids, and transfer of fatty acyl chains to glycerol-3-phosphate or coenzyme A. The genes encoding the enzymes that putatively catalyze these reactions in *N. oceanica* are shown in solid boxes, with the exception of *NO29G00550* and *NO30G00840* which encode putative ACPs. Genes with a significant ( $P_{adjust} \leq 0.05$ ) differential expression in HLM23 relative to the wild type are colored according to the  $\log_2$  fold change in the mutant strain. Symbols next to boxes represent the subcellular localization, predicted by five *in silico* tools. Almost all genes putatively involved in fatty acid biosynthesis were transcriptionally upregulated in HLM23. Most of these genes encode proteins that may contain a chloroplast targeting peptide (cTP) or a signal peptide (SP), suggesting that plastidic fatty acid biosynthesis is enhanced at the transcriptional level in HLM23. TE: thioesterase; SP: signal peptide; SA: signal anchor; cTP: chloroplast targeting peptide; mTP: mitochondrion targeting peptide. The pathway map was adapted from the KEGG PATHWAY database (Kanehisa and Goto 2000).

transcriptionally upregulated ( $p_{\text{adjust}} < 0.05$ ) in HLM23 and two were not differentially expressed, compared to the wild type (Fig. 7). Most of the upregulated proteins were predicted to contain a chloroplast targeting peptide (cTP) or signal peptide (SP) by one or multiple *in silico* prediction tools, and may thus localize in the plastid. We also identified two acyl carrier protein (ACP) genes in *N. oceanica*, *NO29G00550* and *NO30G00840*, predicted to localize in the chloroplast and mitochondrion respectively. As carriers of the growing acyl chains, ACPs fulfill a central role in FAB and it has been shown that they are much stronger expressed than other FAB-related genes in plants (O'Hara et al., 2002). In HLM23, *NO30G00840* transcript abundance was >10-times higher than that of other FAB-associated genes and 1.22- $\log_2$  fold ( $p = 6.37e-12$ ) higher than in the wild type (Dataset S1), ranking the gene at position 28 of the strongest expressed genes of the entire transcriptome dataset. Expression of the putative mitochondrial ACP isoform *NO29G00550* was similar in HLM23 and the wild type and much lower than expression of the likely chloroplastic counterpart.

The first committed and rate-limiting step of FAB is the carboxylation of acetyl-CoA, carried out by acetyl-CoA carboxylase (ACCase, EC# 6.4.1.2) (Majerus and Kilburn 1969). We identified two isoforms for ACC: *NO21G02220* and *NO09G00220*. In particular, isoform *NO09G00220* shows high similarity with ACCases from other microalgae. Moreover, it had the highest transcript abundance among all FAB-associated enzymes in the wild type strain and it showed strong upregulation in HLM23, with transcript levels being 1.72- $\log_2$  fold ( $p = 5.21e-18$ ) increased relative to the wild type. Expression of putative chloroplastic candidates for the other FAB enzymes was moderately to strongly increased in HLM23, ranging from a 0.66- $\log_2$  fold ( $p = 2.16e-03$ ) increase for beta-hydroxyacyl-[acyl-carrier-protein] dehydratase (EC# 4.2.1.59) to a 1.59- $\log_2$  fold ( $p = 5.14e-17$ ) increase for beta-ketoacyl-[acyl-carrier-protein] synthase II (EC# 2.3.1.41/2.3.1.179).

The primary end products of plastidial FAB are C16:0 and C18:0 fatty acyl-ACPs (Li-Beisson et al., 2010), which can further be desaturated by stearoyl-ACP  $\Delta 9$  desaturases. In *Nannochloropsis*, only a single  $\Delta 9$  desaturase, *NO02G01510*, has been identified. *NO02G01510* expression was decreased by 173% ( $p = 1.43e-03$ ) in HLM23 compared to the wild type, which might explain the decreased fraction of MUFA in the neutral lipids of this mutant. In a previous study, Dong and co-workers found decreased *NO02G01510* protein levels during N-starvation (Dong et al., 2013), but other studies have shown that the gene is upregulated at the transcriptional level during N-stress (Li et al., 2014; Zienkiewicz et al., 2020; Janssen et al., 2020).

Fatty acyl-ACPs produced in FAB can directly be used by chloroplastic isoforms of glycerol-3-phosphate acyltransferase (GPAT, EC# 2.3.1.15) or lysophosphatidic acid acyltransferase (LPAAT, EC# 2.3.1.51) for glycerolipid assembly (Joyard et al., 2010). The putative GPAT enzymes of *N. oceanica* were not differentially expressed in HLM23, and expression of LPAAT isoforms *NO28G01540* was decreased by 170% compared to the wild type ( $p = 1.25e-02$ ). Expression of *NO08G04210*, which is the only *Nannochloropsis* LPAAT with no homologue in *A. thaliana* (Li et al., 2014), was increased by 1.93- $\log_2$  fold ( $p = 2.13e-09$ ).

In addition to GPAT and LPAAT, another enzyme that processes fatty acyl-ACPs is thioesterase, which hydrolyzes the thioester bond between ACP and the fatty acyl chain (EC# 3.1.2.21/3.1.2.14), releasing free fatty acids. Four *N. oceanica* enzymes (*NO03G03980*, *NO09G01310*, *NO24G01000* and *NO27G01060*) are predicted to contain a thioesterase domain, but only the putatively mitochondrial *NO09G01310* shows a high degree of similarity to known thioesterases. None of the four proteins were predicted to contain a SP or cTP by more than one prediction tool. Therefore, the plastidic thioesterase isoform of *N. oceanica* needs to be identified by biochemical analysis.

Free fatty acids are converted to acyl-CoA at the outer plastid envelope and in other subcellular compartments by long chain fatty acyl-CoA synthetase (LC-FACS) (Koo et al. 2004). Out of the 13 putative

LC-FACs of *N. oceanica*, three were up- and three were downregulated in HLM23. In particular, *NO18G00930* transcript was increased 0.94- $\log_2$  fold ( $p = 3.27e-04$ ) with a predicted chloroplastic localization. *NO14G00130* transcript levels were 0.95- $\log_2$  fold ( $p = 6.78e-04$ ) higher in HLM23 than in the wild type, however, subcellular localization and metabolic function of *NO14G00130* are unclear. Previously, Li and co-workers have shown that expression of *NO14G00130* gradually increases during N-starvation experiments, as well as for another LC-FACS isoform-encoding gene, *NO05G03180* (Li et al., 2014). However, in HLM23, *NO05G03180* was not differentially expressed compared to the wild type. *NO05G03180* is predicted to localize in the peroxisome, thus it may be responsible for the activation of fatty acids for beta oxidation.

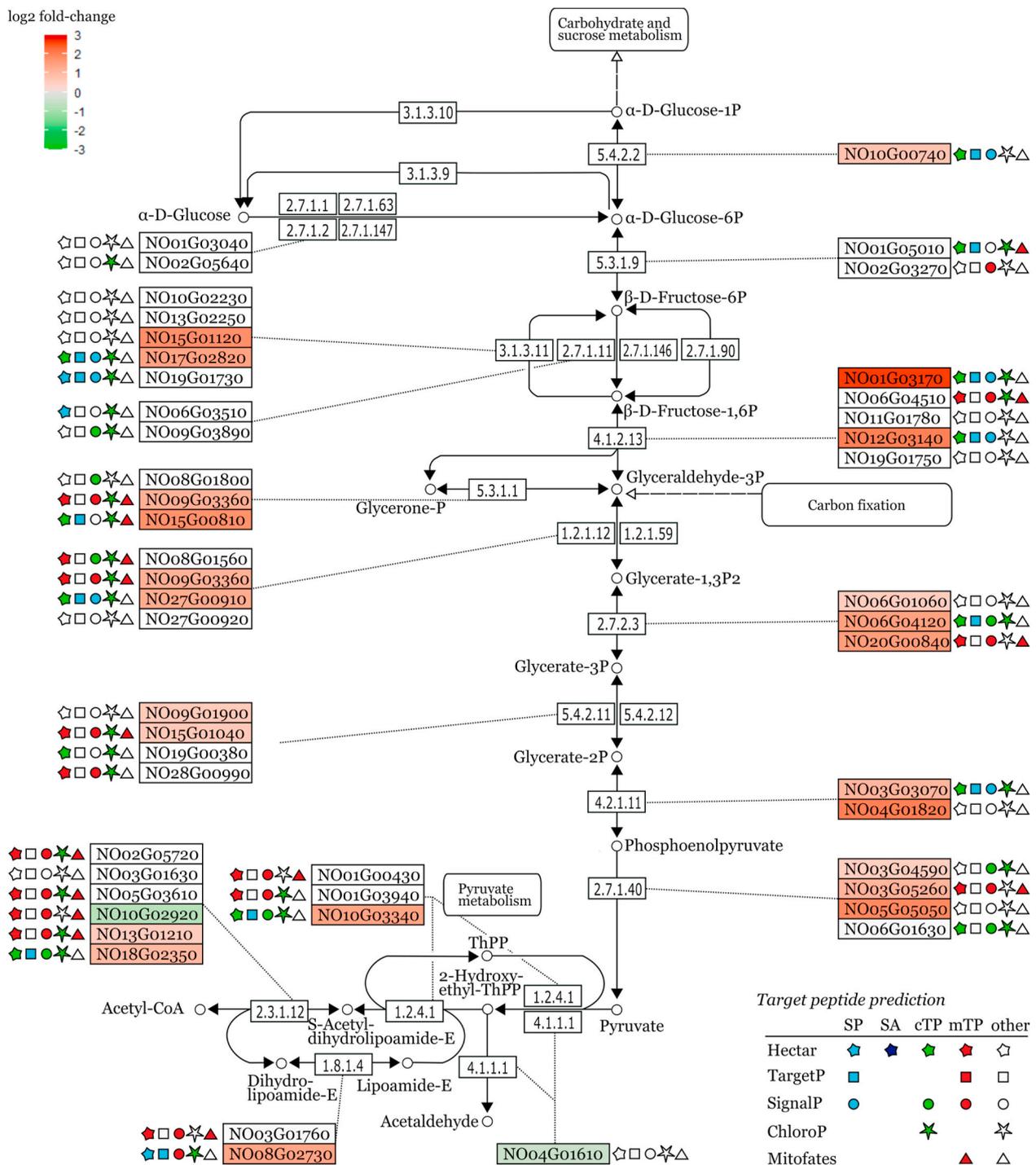
After activation by LC-FACS, fatty acyl-CoA is sequestered by acyl-CoA binding proteins (ACBPs), that i. a. Aid in transport to the ER for fatty acid modifications and lipid assembly (Knudsen et al., 1993). Out of the four putative ACBPs of *N. oceanica* (*NO16G02490*, *NO19G00260*, *NO23G00150*, *NO24G01960*), *NO19G00260* and *NO24G01960* transcript levels were similar in HLM23 and the wild type (Dataset S4), while *NO23G00150* expression was undetectable, which matches previous reports (Li et al., 2014). *NO16G02490* expression was increased 1.63- $\log_2$  fold ( $p = 3.55e-06$ ) in HLM23, putatively implicating this ACBP isoform in sequestration and transport of excess fatty acyl-CoA produced by the mutant.

#### 2.11.2. Supply of fatty acid building blocks may be enhanced in HLM23

The transcriptional upregulation of multiple chloroplastic FAB proteins including ACCase in HLM23 may well explain the increased lipid content observed for this mutant. To meet an increased demand of FAB substrates, the key fatty acid building block acetyl-CoA needs to be supplied at sufficient rates. The photosynthetically produced glyceraldehyde-3-phosphate (G3P), which is the primary product of the Calvin-Benson-Bassham (CBB) cycle, can either be used for production of hexoses through gluconeogenesis, or it can be converted to pyruvate and then acetyl-CoA through the “lower part” of glycolysis (i.e. the reactions from G3P to pyruvate). We mapped all the enzymes putatively involved in this part of central carbon metabolism (Fig. 8).

HLM23 showed transcriptional upregulation for multiple isoforms of all enzymes that participate in the conversion of G3P to pyruvate (Fig. 8; Dataset S2). In *N. oceanica*, the lower part of glycolysis is likely present in the cytosol, chloroplast and mitochondrion, whereas the upper part may only be fully functional in the chloroplast (Poliner et al., 2015). Enzymes of the lower part of glycolysis were previously shown to be co-regulated with fatty acid biosynthesis genes in *N. oceanica*, suggesting functional cooperation (Poliner et al., 2015). The rate-limiting step of the lower part of glycolysis is catalyzed by glyceraldehyde 3-phosphate dehydrogenase (GAPDH, EC# 1.2.1.12) (Shestov et al., 2014; Orlenko et al. 2016), which has four predicted isoforms in *N. oceanica*. Among them, the putative mitochondrial isoform *NO09G03360* and the putative chloroplastic isoform *NO27G00910* had a 1.1- $\log_2$  fold ( $p = 9.52e-08$ ) and 1.17- $\log_2$  fold ( $p = 3.22e-08$ ) increased transcript abundance, respectively. In agreement with this, several isoforms of all other enzymes involved in the lower part of glycolysis were also upregulated in HLM23. For each enzyme, at least one of the upregulated isoforms was predicted to localize in the chloroplast by one or multiple prediction tools, with the exception of phosphoglycerate mutase (PGM, EC# 5.4.2.11). At the same time, previous studies have shown that the reaction catalyzed by PGM does not pose a bottleneck for glycolysis among various organisms (Orlenko et al. 2016).

Glycolysis and FAB are bridged by a multi-enzyme complex called the pyruvate dehydrogenase complex (PDHC) that converts pyruvate into acetyl-CoA. PDHC consists of the subunits E1 (EC# 1.2.4.1), E2 (EC# 2.3.1.12) and E3 (EC# 1.8.4.1) which are present in the mitochondrion and chloroplast of *N. oceanica* (Fig. 8). The three putative chloroplastic E1, E2 and E3 subunits *NO10G03340*, *NO18G02350* and *NO08G02730* were upregulated 1.32- $\log_2$  fold ( $p = 3.11e-14$ ), 0.94- $\log_2$  fold ( $p = 5.92e-06$ ) and 1.3- $\log_2$  fold ( $p = 3.47e-09$ ), respectively, in



**Fig. 8.** Differential expression of genes encoding enzymes putatively involved in glycolysis, gluconeogenesis and pyruvate metabolism in HLM23. The schematic illustrates the main reactions of the Embden-Meyerhof-Parnas (EMP) pathway, and of pyruvate metabolism. Genes involved in both pathways were either transcriptionally upregulated or not differentially expressed in HLM23 relative to the wild type, with the exceptions of *NO04G01610* and *NO10G02920*. Among the upregulated genes, subcellular localization is predominantly predicted to be chloroplastic. Most EMP pathway enzymes can catalyze reactions in either the glycolytic or gluconeogenic direction. The pathway map was adapted from the KEGG PATHWAY database (Kanehisa and Goto 2000).

HLM23. Expression of putative mitochondrial PDHC subunits, on the other hand, was mainly unchanged. The mitochondrial PDHC provides acetyl-CoA not only as substrate for mitochondrial FAB, but more importantly for the citric acid cycle. Pyruvate can further be converted to acetyl-CoA via the PDHC-bypass. However, the highest expressed acetyl-CoA synthetase of this pathway lacks co-expression with FAB genes during diel cycles (Poliner et al., 2015), suggesting that the PDHC-bypass likely does not contribute to acetyl-CoA supply for lipid

biosynthesis in *N. oceanica*. The single copy pyruvate decarboxylase (PDC, EC# 4.1.1.1) *NO04G01610* that channels pyruvate into this bypass is transcriptionally downregulated in HLM23 ( $p=2.77e-02$ ), suggesting that this pathway is not responsible for the increased fatty acid content of the mutant. Expression of putative isoforms of the other PDHC-bypass enzymes aldehyde dehydrogenase and acetate-CoA ligase was also decreased in HLM23 (Dataset S2).

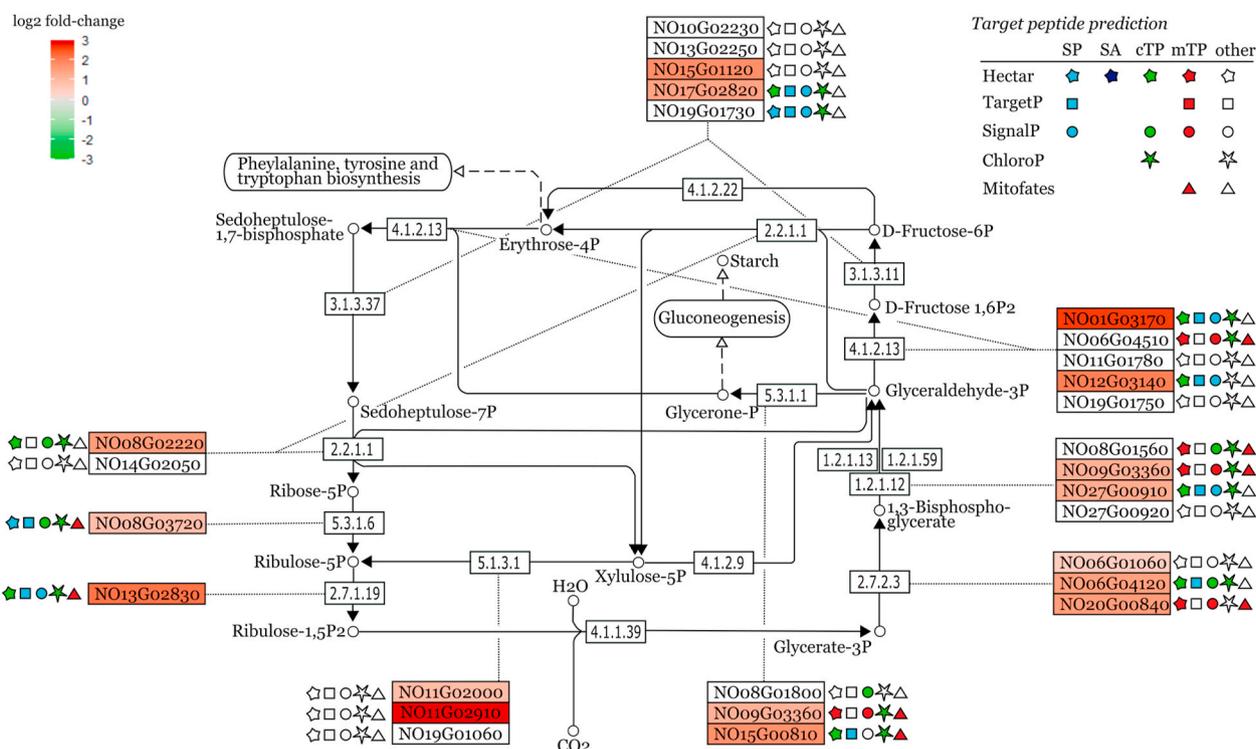
Not only the lower part but also the upper part of glycolysis was

transcriptionally upregulated in HLM23, and almost all of the upregulated genes are predicted as chloroplastic isoforms (Fig. 8; Dataset S2). Most of the glycolytic enzymes function not only in glycolysis but also in gluconeogenesis, with the exceptions of the glycolytic enzymes hexokinase, phosphofructokinase and pyruvate kinase. Phosphofructokinase (PFK, EC# 2.7.1.11) is substituted by fructose 1,6-bisphosphatase (FBPase, EC# 3.1.3.11) in gluconeogenesis. In HLM23, the putative chloroplastic FBPase *NO17G02820* is upregulated 1.39- $\log_2$  fold ( $p = 2.11\text{e-}09$ ), whereas expression of the glycolytic PFK, is not affected in the mutant. We also found overexpression of putative chloroplastic isoform *NO15G00810* of triose phosphate isomerase (TPI, EC# 5.3.1.1, increased 1.7- $\log_2$  fold,  $p = 1.36\text{e-}11$ ) and fructose-bisphosphate aldolase isoforms *NO12G03140* and *NO01G03170* (EC# 4.1.2.13, increased 1.94–2.55- $\log_2$  fold,  $p < 4.50\text{e-}18$ ), which catalyze the reversible steps of the upper part of glycolysis/gluconeogenesis. The overexpression of FBPase, in particular, may suggest an increased level of gluconeogenesis in HLM23. However, FBPase, along with TPI, fructose-bisphosphate aldolase, GAPDH and phosphoglycerate kinase (PGK, EC# 2.7.2.3), are also part of the CBB cycle which is responsible for carbon fixation (Fig. 9). Consequently, the upregulation of these enzymes in HLM23 could be related to an increased activity of glycolysis/gluconeogenesis, the CBB cycle, or both pathways. An increased flux capacity of the CBB cycle is further suggested by upregulation of CBB cycle-specific enzymes such as transketolase (EC# 2.2.1.1), phosphopentose isomerase (EC# 5.3.1.6), phosphopentose epimerase (EC# 5.1.3.1) and phosphoribulokinase (EC# 2.7.1.19) in HLM23 (Fig. 9, Dataset S3). Expression levels of ribulose 1,5-bisphosphate carboxylase (RuBisCO, EC# 4.1.1.39), a key enzyme of the CBB cycle, were not assessable in our experiments because in *Nannochloropsis* both enzyme subunits are encoded in the chloroplast genome. As chloroplast transcripts do not contain poly(A) tails, the RuBisCO transcripts were not analyzed during mRNA sequencing. Transcription of a CbbX protein homologue, which is an activator of red-type RuBisCO protein (Mueller-Cajar et al. 2011),

was increased 0.56- $\log_2$  fold ( $p = 2.24\text{e-}02$ ) in HLM23.

The upregulation of genes encoding enzymes involved in the CBB cycle, central carbon metabolism and chloroplastic FAB in HLM23 may be related to this mutant's altered ability to cope with higher light irradiation ( $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ ), which was reflected in an increased  $F_v/F_m$  under these conditions (Fig. 3C). Accordingly, when cultures were grown at lower irradiance conditions ( $150$  or  $350 \mu\text{mol m}^{-2} \text{s}^{-1}$ ),  $F_v/F_m$  of HLM23 and the wild type were similar and the differences in NL contents were less pronounced (Fig. S5). Previously, Yang and colleagues have shown that overexpression of chloroplastic FBPase in the green microalga *Chlorella vulgaris* caused an increase in photosynthetic quantum yield and oxygen evolution rates (Yang et al., 2017), showcasing a connection between expression levels of CBB cycle enzymes and photosynthetic performance. Future studies might shed light on the ability of HLM23 to cope with the adverse effects associated with high light conditions, and if there is a link to the transcriptional upregulation of metabolic pathways that directly utilize photosynthetic assimilates.

In line with the elevated  $F_v/F_m$ , HLM23 showed upregulation of several proteins that are directly or indirectly associated with the light reactions of photosynthesis (Dataset S4). Transcript levels of *NO24G02290*, encoding a chlorophyll *a-b* binding protein of the light harvesting complex, were increased 0.46- $\log_2$  fold ( $p = 4.29\text{e-}02$ ) in HLM23 compared to the wild type. Expression of four out of six putative ferredoxin proteins was increased 1.1–3.45- $\log_2$  fold ( $p < 8.78\text{e-}04$ ) and the only known ferredoxin NADP(+) oxidoreductase was upregulated 1.15- $\log_2$  fold ( $p = 3.07\text{e-}11$ ) in HLM23 (Dataset S4). Additionally, the only isoform of the highly expressed PSII manganese-stabilizing protein PsbO was upregulated 0.89- $\log_2$  fold ( $p = 2.09\text{e-}07$ ). PsbO aids in rapid turnovers of the oxygen evolving reactions (Popelkova and Yocum 2011), and is required for high photosynthetic quantum yields in *A. thaliana* (Yi et al., 2005). Expression of two magnesium chelatase (EC# 6.6.1.1) subunits ChlD and ChlH was increased 1.31- $\log_2$  fold ( $p = 3.54\text{e-}07$ ) and 2.02- $\log_2$  fold ( $p = 7.44\text{e-}25$ ) respectively in HLM23



**Fig. 9. Differential expression of genes encoding enzymes putatively involved in the Calvin-Benson-Bassham cycle in HLM23.** The schematic illustrates the reactions of the CBB cycle. Gene candidates for every reaction of the cycle were transcriptionally upregulated in HLM23 relative to the wild type. Most of the upregulated proteins are predicted to localize to the chloroplast. RuBisCO expression levels could not be quantified by the mRNA sequencing strategy of choice. The pathway map was adapted from the KEGG PATHWAY database (Kanehisa and Goto 2000).

(Dataset S4). Magnesium chelatase catalyzes the insertion of a magnesium ion into protoporphyrin IX, which is the first committed step of chlorophyll biosynthesis (Masuda 2008). The third subunit, ChlI is encoded in the chloroplast genome of *N. oceanica* and its expression was therefore not quantified.

Intriguingly, a putative glucose-6-phosphate/ $P_i$  translocator (GPT) was upregulated 3.54- $\log_2$  fold ( $p = 7.56\text{e-}06$ ) in HLM23 (Dataset S4). In *A. thaliana*, a GPT knockout is associated with decreased starch synthesis under high light conditions (Dyson et al., 2015). It was suggested that high cytosolic triose-phosphate levels can induce expression of GPT, which promotes the import of glucose-6-phosphate (G6P) from the cytosol to the chloroplast (Weise et al., 2019). Imported G6P can then fuel either carbohydrate synthesis or the chloroplastic G6P shunt, which corresponds to the oxidative phase of the pentose phosphate pathway and may have a stabilizing effect on photosynthesis (Sharkey and Weise 2016). The function of GPT in microalgae, however, remains obscure. Chloroplastic isoforms of the G6P shunt enzymes, such as 6-phosphogluconolactonase, were transcriptionally downregulated by up to  $-1.03\text{-}\log_2$  fold ( $p = 1.12\text{e-}03$ ) in HLM23, suggesting that the upregulation of GPT in this mutant has a different reason than contributing to activity of this pathway.

Lastly, it should be noted that multiple putative and confirmed transcriptional regulators were differentially expressed in HLM23. For instance, expression of the NobZIP1 transcription factor *NO19G01720* was increased 1.05- $\log_2$  fold ( $p = 3.16\text{e-}07$ ). Recently, Li and co-workers found that an *N. oceanica* mutant overexpressing a *NO19G01720* homologue had an increased lipid content, which the authors linked to a decreased expression of UDP-glucose dehydrogenase (UGDH), a key enzyme in cell wall carbohydrate polymer metabolism (Li et al., 2019). However, despite the upregulation of *NO19G01720*, expression of the UGDH gene *NO14G01750* was not decreased, but increased 0.7- $\log_2$  fold ( $p = 4.46\text{e-}03$ ) in HLM23, suggesting that *NO14G01750* expression could be modulated by multiple transcriptional regulators and not only by *NO19G01720*. This also indicates that lipid contents of HLM23 could potentially be further improved by silencing the gene expression of UGDH in this strain, for instance through RNA interference, as previously demonstrated for wild type *N. oceanica* (Li et al., 2019).

In conclusion, the knockout of the putative APETALA2-like transcription factor *NO06G03670* in HLM23 resulted in increased NL content and  $F_v/F_m$ , without affecting growth rate compared to the wild type. This phenotype was accompanied by a transcriptional upregulation of key components of several chloroplast-located metabolic pathways, including sugar metabolism, the conversion of pyruvate to acetyl-CoA, fatty acid biosynthesis and the CBB cycle. However, the exact cause of the alteration in the lipid content of *N. oceanica* due to the *NO06G03670* knockout remains to be elucidated. In this context, a loss-of function mutant of APETALA2 in *A. thaliana* showed changes in the ratio of hexose to sucrose during seed development, suggesting that APETALA2 may control seed mass by affecting sugar metabolism, in plants (Ohto et al., 2005). Because lipid biosynthesis from sugars and related substrates is a preferentially anabolic activity in the oleaginous *N. oceanica* (Alboresi et al., 2016), *NO06G03670* may be involved in directly or indirectly regulating sugar metabolism and trafficking, which in turn could affect photosynthetic carbon assimilation and partitioning, and consequently lipid content. In this regard, the increased  $F_v/F_m$  of HLM23 is intriguing and future studies are needed to elucidate the role of the APETALA2-like transcription factor in the regulatory network of *N. oceanica*. Engineering of microalgal strains with an increased photosynthetic efficiency is a difficult task that has thus far met with limited success. However, commercialization of microalgal-derived bulk commodities will heavily rely on an increased photosynthetic performance of microalgal strains (Benvenuti et al., 2017).

### 3. Conclusion

In this study, we developed a novel forward genetic screen for

*N. oceanica* based on insertional mutagenesis and FACS. We isolated multiple lipid-overproducing mutant strains by iteratively sorting an insertional mutant library five times. An optimized genotyping methodology based on the type IIS restriction endonuclease MmeI allowed us to identify the causative mutations in five mutants, which showed 12–59% increased NL contents  $DCW^{-1}$  relative to the wild type. The genes that were disrupted in these strains present potential targets for genetically engineering *Nannochloropsis* for an increased lipid productivity. Highest NL contents were found for strain HLM15, which is likely incapable of sustaining protein homeostasis due to a disruption of a DnaJ protein. The 59% increase in NL  $DCW^{-1}$  that we found for this mutant was accompanied by a 21% decrease in growth rate and susceptibility to high light. Two of the genes that were affected in the other mutant strains encode transcription factors, which highlights the importance of this type of protein as a master switch for metabolic regulation. One of these transcription factors, *NO09G01030*, is a homologue of a transcription factor of *N. gaditana* that was previously shown to be associated with lipid metabolism (Ajjawi et al., 2017). The other one, *NO06G03670*, was predicted to upregulate genes involved in protein homeostasis and downregulate chloroplastic genes associated with fatty acid biosynthesis and carbon flux to acetyl-CoA (Hu et al., 2014), which we confirmed by mRNA sequencing methodology. *NO06G03670* knockout strain HLM23 had a NL content of 30.2%  $DCW^{-1}$  and increased photosynthetic quantum yield under favorable biomass production conditions, and growth was not impaired. Accordingly, projected NL productivities of one-step batch cultivations with HLM23 are 39% higher compared to productivities when using wild type *N. oceanica*. The insights gained in this study increase the genetic knowledge about an industrially attractive microalga, and will be useful to formulate new hypotheses for the genetic engineering of *Nannochloropsis* strains with an increased NL productivity. While our findings significantly contribute to an improved understanding of *N. oceanica*, more studies are needed to elucidate the functions of the many genes that remain unknown. In this regard, our optimized MmeI-based genotyping methodology lays the foundation of future screenings for not only lipid accumulation but any desired phenotype.

## 4. Materials and methods

### 4.1. Strains and cultivation conditions

The microalgal strain used in this study, *N. oceanica* IMET1 was a kind gift by prof. Jian Xu (Qingdao Institute for Bioenergy and Bioprocess Technology, Chinese Academy of Sciences). Microalgae were cultured in artificial sea water (ASW, 419.23 mM NaCl, 22.53 mM  $Na_2SO_4$ , 5.42 mM  $CaCl_2$ , 4.88 mM  $K_2SO_4$ , 48.21 mM  $MgCl_2$  and 20 mM HEPES, pH 8), supplemented with 2 ml  $l^{-1}$  of Nutribloom plus (Necton, Portugal) growth media (ASW-NB) at 25 °C. For maintenance purposes and after fluorescence-activated cell sorting steps, cultures were incubated on an orbital shaker operated at 90 rpm in a climate-controlled chamber, aerated with ambient air and illuminated with warm-white fluorescent light bulbs at 50  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light intensity (“LL” conditions). An HT Multitron Pro (Infors Benelux, Netherlands) orbital shaker unit with 0.2%  $CO_2$ -enriched air was used for cultivation at 150  $\mu\text{mol m}^{-2} \text{s}^{-1}$  illumination intensity (“ML” conditions) at 120 rpm shaking frequency. Isolated high lipid mutant (HLM) strains were cultivated in 50 ml shake flasks at 600  $\mu\text{mol m}^{-2} \text{s}^{-1}$  (“HL” conditions) in an Algem HT24 photobioreactor (Algeniunity, UK) that was placed inside an HT Multitron Pro operated at conditions as described above. All experiments were carried out with a 16:8 h diurnal light cycle.

For cultivation on solid medium, ASW was supplemented with 1% (w/v) of agar (Merck) before autoclaving, cooled to 60 °C and mixed with 2 ml  $l^{-1}$  of nutribloom plus and 5  $\mu\text{g ml}^{-1}$  Zeocin before distribution into petri dishes for solidification. Algae-containing plates were incubated at 25 °C in ambient air under warm-white fluorescent light

bulbs at an intensity of 60–80  $\mu\text{mol m}^{-2} \text{s}^{-1}$  with a 16:8 h diurnal cycle.

#### 4.2. Construction of insertion cassette DNA

The insertion cassette (IC) carried the bleomycin resistance gene *shble* (GenBank accession number A31898.1) from *Streptoalloteichus hindustanus*, under control of the strong endogenous promoter of the violaxanthin chlorophyll *a* binding protein precursor (VCP) gene, including the Kozak sequence and first genetic intron. Two transcriptional terminators belonging to the  $\alpha$ -tubulin and *Clp* protease proteolytic subunit genes were added in head-to-head orientation downstream of *shble*, to safeguard transcriptional termination of genes at the insertion site regardless of IC orientation. Assembly of these elements into the pUC19 derivative plasmid pNIM14 was previously described (Naduthodi et al., 2019). Using pNIM14 as a template, we constructed pNIM21 by adding terminal MmeI restriction endonuclease recognition sequences to facilitate precise excision of the IC from genomic DNA during insertion site tracing. We further inserted ~100 nucleotides of artificial barcodes between these MmeI recognition sequences and the VCP promoter and *Clp* terminator elements respectively (Fig. 1A). This allowed us to design barcode-complementary cassette PCR primers for insertion site tracing, which reduced spurious primer annealing to endogenous genome sequences. The barcode sequences were inserted into pNIM14 as PCR primer overhangs and assembled using Gibson assembly technique (NEBuilder HiFi DNA Assembly Master Mix, NEB #E2621) with 22 nucleotides overlap. The first assembly fragment was PCR-amplified using oligonucleotides and with 1 ng of MlsI-linearized pNIM14 as template and an annealing temperature ( $T_a$ ) of 65 °C. The second assembly fragment was PCR-amplified using oligonucleotides and with 1 ng of ScaI-linearized pNIM14 as template and a  $T_a$  of 69 °C. Both PCRs were carried out using Q5 DNA polymerase (NEB #M0492), 0.25  $\mu\text{M}$  primers and the following cycling scheme: 30 s at 98 °C, 30 cycles (8 s at 98 °C, 10 s at  $T_a$ , 90 s at 72 °C), 3 min at 72 °C. Linearization of pNIM14 using MlsI (Thermo Fisher Scientific #FD1214) or ScaI (Thermo Fisher Scientific #FD0434) was carried out according to the manufacturer's instructions. PCR products were purified using spin columns (Zymo Research #D4004) and used in Gibson assembly according to the manufacturer's instructions. Correct assembly of pNIM21 was verified by PCR and sequencing. The IC was amplified from 1 ng ScaI-linearized (Thermo Fisher Scientific #FD1374) pNIM21 with Q5 DNA polymerase, using 0.5  $\mu\text{M}$  of 5'-GTTGGAATCTTTCAACACGCTAGG-3' and 5'-GTTGGATTGATACTGTTGTTGTTGG-3' with the following cycling scheme: 30 s at 98 °C, 35 cycles (5 s at 98 °C, 85 s at 72 °C), 5 min at 72 °C. The IC was purified (Thermo Fisher Scientific #K0832) according to the manufacturer's instructions, using the PCR cleanup version of the protocol, with the exception that nuclease-free (NF) H<sub>2</sub>O was used for elution of DNA from the column.

#### 4.3. Transformation of *N. oceanica*

*N. oceanica* IMET1 was transformed using electroporation similar to the protocol described by Vieler and colleagues (Vieler et al., 2012). Briefly, microalgal cells were cultivated at ML conditions and harvested during mid-exponential growth stage by centrifugation (2500 $\times$ g, 5 min, 4 °C). Cell pellets were washed thrice with 4 °C-cold 375 mM sorbitol and resuspended at 2.5  $\times 10^9$  cells ml<sup>-1</sup>. 200  $\mu\text{l}$  of cell suspension was mixed with 1  $\mu\text{g}$  of IC DNA in pre-cooled 2 mm electroporation cuvettes and rapidly pulsed using a Bio-Rad GenePulser II, set to exponential pulse decay with 11 kV cm<sup>-1</sup> electric field strength, 50  $\mu\text{F}$  capacitance and 600  $\Omega$  resistance. After the ~25 ms pulse, cells were immediately transferred to 5 ml of 22 °C ASW-NB and recovered for 24 h at dim light without agitation. Subsequently, cells were pelleted (2500 $\times$ g, 10 min), resuspended in a small amount of supernatant and plated onto ASW-NB plates containing 5  $\mu\text{g ml}^{-1}$  Zeocin and incubated as described previously for 4 weeks, before plates were flushed with ASW-NB containing 5  $\mu\text{g ml}^{-1}$  of

Zeocin, and pooled in a 250 ml shake flask, yielding the insertional mutant library.

#### 4.4. Flow cytometry analysis

BODIPY fluorescence emission was quantified by flow cytometry analysis using an SH800S (Sony Biotechnology, USA) cell sorter, equipped with a 488 nm laser and a 70  $\mu\text{m}$  nozzle microfluidic chip. Detector wavelengths for different channels were: 488 nm (forward scatter, gain 2); 488 nm (side scatter, gain 22%); 510  $\pm$  10 nm (BODIPY, gain 29%); 720  $\pm$  30 nm (chlorophyll *a* autofluorescence, gain 40%). For standard fluorescence analyses, a minimum of 50,000 events were screened per sample. Gating was applied as previously described (Südfeld et al., 2020) to remove background noise. For statistical data treatment, only events in the final gate ("living cells") were considered.

#### 4.5. Staining of intracellular neutral lipids with BODIPY

Neutral lipids (NLs) were visualized for flow cytometry and FACS using the fluorescent dye BODIPY 505/515 (Invitrogen, #D3921) as previously described (Südfeld et al., 2020). Microalgae were stained at an OD<sub>750</sub> of 0.3–0.5 with 6% (v/v) of DMSO and 1.2  $\mu\text{g ml}^{-1}$  BODIPY 505/515. After addition of the dye, cell suspensions were vortexed for 5 s and incubated in the dark for 15 min before being subjected to flow cytometry/FACS.

#### 4.6. FACS sorting for enriching high lipid mutants in the mutant library

Cell sorting was carried out using the same instrument and operational settings as during regular flow cytometry analysis. Cells were sorted directly into sterile 250 ml Erlenmeyer flasks, filled with 50 ml ASW-NB. Sorting was terminated at 300,000 sorted events in 'Ultra purity' sort mode of the SH800S software. Gating was applied as indicated previously (Südfeld et al., 2020) to filter out non-vital cells and background noise, with an additional final gate for high-NL cells (Fig. S1). Sorted cultures were recovered under LL conditions for 10 days (d). New cultures were inoculated to an OD<sub>750</sub> of 0.1 and acclimated to ML conditions for 3 d. Finally, cultures were diluted to an OD<sub>750</sub> of 0.1, incubated for 2 more days under ML conditions and subjected to the next round of sorting. ML conditions were employed during HLM enrichment and isolation by FACS, because they are sufficient to induce NL accumulation, but less stressful for cells than HL conditions. Therefore, using ML conditions prevented loss of high-light sensitive mutant strains during the library enrichment for HLMs.

#### 4.7. Strain isolation from enriched libraries using FACS

After 4 rounds of sorting, HLM-enriched libraries were recovered under LL conditions, set to an OD<sub>750</sub> of 0.1 and acclimated to ML conditions for 3 d. New cultures were inoculated to an OD<sub>750</sub> of 0.1 and grown under ML conditions for an additional 2 d before strains were isolated from the pooled libraries by depositing single cells onto agar plates using a high-NL gate as described above, and the 'Single cell' sort mode of the SH800S software.

#### 4.8. Characterization of isolated HLM strains

Four weeks after HLM strain isolation by FACS, all colonies showed identical morphology and comparable colony diameter on agar. Therefore, the first 17 viable colonies were picked from agar plates and expanded to shake flasks, which were incubated at LL conditions in duplicates. After ~1 week, fresh cultures were inoculated to an OD<sub>750</sub> of 0.07, and incubated in an Algem HT24 photobioreactor (Algenity, UK) operated at 350  $\mu\text{mol m}^{-2} \text{s}^{-1}$  for 3 d. Subsequently, cultures were diluted to an OD<sub>750</sub> of 0.07 and incubated at 600  $\mu\text{mol m}^{-2} \text{s}^{-1}$  for 14 d.

Growth curves were obtained by measuring the OD<sub>750</sub> daily, 1 h before onset of the dark phase of the diurnal light cycle. Maximum specific growth rate  $\mu_{max}$  was obtained as the slope of a linear model that was fitted on logarithmically-transformed OD<sub>750</sub> values, obtained within the first 72 h of the cultivation. For biochemical characterization, cultures were harvested 1 h before onset of the dark phase on the second day of cultivation at 600  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . HL conditions were employed to improve translatability of findings to outdoor production systems in future studies and to maximize the relevance of findings. Microalgal suspensions were centrifuged (4000×g, 10 min) and resuspended in 2 ml of ASW and 600  $\mu\text{l}$  of the suspension was subjected to lipid quantification as described below. The dry cell weight (DCW) concentration of the suspension was measured as previously described (Zhu and Lee 1997) with 0.5 M ammonium formate as wash buffer. Cell count of the suspension was determined using a Beckman Coulter Multisizer 3 (Beckman Coulter Inc., USA, 50  $\mu\text{m}$  orifice).

#### 4.9. Assessment of photosynthetic performance

The maximum efficiency of photosystem II photochemistry was assessed using *in vivo* chlorophyll fluorescence with an AquaPen AP 100-C (Photon System Instruments, Czech Republic) handheld fluorometer. Samples were dark adapted for 20 min prior to analysis according to the manufacturer's protocol.

#### 4.10. Identification of cassette insertion sites

Cassette PCR was carried out employing a modified version of procedures reported by Goodman and co-workers (Goodman et al., 2009) and Zhang and co-workers (Zhang et al., 2014). As described above, the IC carried terminal recognition sites for the type IIS restriction enzyme MmeI. MmeI induces a 2 nucleotide staggered-end cut 20–21 nucleotides outside of its recognition site. The recognition sites were oriented in a way that the MmeI would cut in the flanking genomic DNA of transformant strains, facilitating 'excision' of the ends of the IC together with 20–21 nucleotides of flanking sequences. For this, genomic DNA was extracted from exponentially growing cultures. Briefly,  $\sim 6 \times 10^7$  cells were pelleted (5000 × g, 3 min) and resuspended in 150  $\mu\text{l}$  of lysis buffer as described by Daboussi and colleagues (Daboussi et al., 2014). The suspension was vortexed for 60 s at high speed, incubated overnight at 22 °C, again vortexed for 60 s and pelleted (5000 × g, 5 min). The supernatant was purified using the FavorPrep column purification kit (Bio-Connect B.V. #FAGDC001) according to the manufacturer's instructions, except that NF H<sub>2</sub>O was used for eluting DNA from the column. DNA concentration was determined using a NanoDrop device, and 350 ng was digested with 0.16  $\mu\text{l}$  MmeI (NEB #R0637) at 37 °C for 10 min according to the manufacturer's instructions. Integrity of genomic DNA and successful digestion were verified by agarose gel electrophoresis. Double stranded DNA adapters were assembled by mixing 100  $\mu\text{M}$  stock solutions of oligonucleotides and 5'-P-CTCCACGTTACCC-NH<sub>3</sub>-3' (5'-phosphate-modified, 3'-amine modified), heating the mixture for 5 min at 93 °C followed by slow cooling to 22 °C (~1 h). The bottom strand of the adapter was designed with a 5'-phosphate and 3'-NH<sub>3</sub> modification to increase ligation efficiency and to prevent unspecific amplification during the PCR steps. The top strand was designed to have a NN-3' overhang after adapter formation to facilitate annealing to the NN-3' overhang of MmeI-digested DNA fragments. Roughly 13  $\mu\text{g}$  of MmeI-digested DNA was mixed with 4.8  $\mu\text{l}$  of 50  $\mu\text{M}$  adapters and subjected to ligation using 2.5 U of T4 DNA ligase (Thermo Fisher Scientific #EL0014) in an 8  $\mu\text{l}$  reaction for 60 min at 22 °C. The reaction was stopped by heat inactivation of the enzyme and the solution was diluted 1:5 with NF H<sub>2</sub>O. To amplify the target sequence over the genomic background, 1  $\mu\text{l}$  of the mixture was used as template for nested PCR (2 rounds total) with Taq polymerase (Thermo Fisher Scientific #K1081) using 0.25  $\mu\text{M}$  of primers specific for the adapter and for either the

5'-end or the 3'-end of the IC respectively (two separate nested PCRs per strain). The first PCR iteration was carried out using oligonucleotide 5'-TAGGCAGGCGGGCACCTCGCGTTAGTG-3' together with either 5'-CGGAGATGCGTACCGTAAGGTGGCATTAGC-3' or 5'-CTCGGCTGGTATCTGAGGAGTAGCCACAC-3' in the reactions specific for the 5'- and 3'-end of the IC respectively. For increased specificity, PCRs were run with touchdown settings, as follows: 5 min at 95 °C, 7 cycles (30 s at 95 °C, 45 s at 72 °C), 32 cycles (30 s at 95 °C, 15 s at  $T_a$ , 30 s at 72 °C), 10 min at 72 °C.  $T_a$  was 69.2 and 70.6 °C for the 5'- and 3'-end specific reactions respectively. PCR products were diluted 1:50 with NF H<sub>2</sub>O and subjected to a second round of PCR with nested primers, namely 5'-GTTAGTGCTGGGTCTAGGCGCTCTGG-3' together with either 5'-AGCCAGACATCATCCATCGCCTCTGATCG-3' or 5'-CGAGGAAGTAGACTGTGACGTTGGCGAT-3' for the 5'- and 3'-end specific reactions respectively. Cycling schemes were similar to the first round of PCR, but cycle number was 28 instead of 32.  $T_a$  during these 28 cycles was 70.3 and 68.9 °C for the 5'- and 3'-end specific reactions respectively. PCR products were purified and sequenced in order to reveal the nucleotide sequence surrounding the IC. The genomic position of IC integration was then determined by using the flanking genomic nucleotide sequences in BLASTn queries against the *N. oceanica* IMET1 genome version 2 (Gong et al., 2020), using standard settings, except changing the E value to 0.01. The overall procedure is summarized in Figure S3.

#### 4.11. Quantification of neutral and polar lipids via GC-FID

Neutral and polar lipid contents were determined using a modified version of the protocol described by Remmers et al., (2017). 600  $\mu\text{l}$  of an algal suspension with known cell concentration and DCW concentration was subjected to fatty acid extraction, separation into neutral and polar acyl lipids, methylation and quantification. The suspension was freeze-dried in a lysing matrix (#116914050-CF, MP Biomedicals) and subjected to mechanical cell disruption and lipid extraction with a chloroform:methanol (1:1.25) solution containing the 2 internal standards triptadecanoin (T4257; Sigma-Aldrich) and 1,2-didecanoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (840,434, Avanti Polar Lipids Inc.) at a known concentration. Polar and apolar lipids were separated using Sep-Pak Vac silica cartridges (6 cc, 1000 mg; Waters). Neutral lipids were eluted with a hexane:diethylether (7:1 v/v) solution, and polar lipids using methanol:acetone:hexane (2:2:1 v/v). Solvents were evaporated under N<sub>2</sub> gas and lipid fractions were methylated in 5% (v/v) H<sub>2</sub>SO<sub>4</sub>-containing MeOH (1 h, 100 °C), extracted with hexane and analyzed by gas chromatography (GC-FID). Fatty acids were quantified based on the relative responses of individual fatty acids compared to the signal of the internal standard and normalized to the DCW or cell concentration.

#### 4.12. Targeted knockout of NO06G03670

Targeted gene knockouts were produced using CRISPR-Cas technology as previously described (Poliner et al. 2018) [Mihris Naduthodi, personal communication]. Three plasmids pTKO-sp1, pTKO-sp2 and pTKO-sp3 were generated that encode an FnCas12a variant (Zetsche et al., 2015) and three different CRISPR RNAs (crRNAs) crRNA1, crRNA2 and crRNA3, containing three different protospacer sequences (sp1, sp2 and sp3, respectively), targeting the first exon of the gene NO06G03670 (Fig. 6A). The FnCas12a was obtained from plasmid pY004 (pcDNA3.1-hFnCpf1) which was a gift from Feng Zhang (Addgene plasmid # 69,976; <http://n2t.net/addgene:69976>; RRID:Addgene\_69,976). The protospacers were designed using CHOPCHOP v3 (Labun et al., 2019), selecting the software options CRISPR/Cpf1 or CasX, and sgRNA length without PAM: 25, and the 5'-PAM sequence TTTN. Targets that were predicted as highly efficient were curated to select an appropriately positioned 5'-PAM sequence TTTV (V = A/C/G), and restriction-cloned into the vector pNAD\_BaeI as follows. The vector

(2 µg) was digested for 60 min using 10 U BaeI (NEB #R0613) according to the manufacturer's instructions, and the linear product was purified from a 0.5% (w/v) agarose gel after 90 min of electrophoresis at 100 V (Thermo Fisher Scientific #K0832). Inserts containing sp1, sp2 and sp3 sequences were synthesized as single-stranded oligonucleotides and annealed by mixing 100 µM solutions in equal volumes, heating at 95 °C for 5 min, and slowly cooling the suspension to 22 °C (~1 h). The pTKO-sp1 insert was created by annealing 5'-ACCACCTCCCCTTGA CGTTACTGGCCG-3' and 5'-AAGTAACGTCAAGGGGAGGGTGGT ATCTA-3', the pTKO-sp2 insert was created by annealing 5'-CCCTTCCATTGACCAACCAAAAATGGCCG-3' and 5'-ATTTTGGT TGGTCAAATGGAAGGGATCTA-3', and the pTKO-sp3 insert was created by annealing 5'-GTTGAGGCGCATCGGGTATTGCTGGCCG-3' and 5'-AGGCAATCACCCGATGCGCTCAACATCTA-3'. Then, 40 ng of linearized vector was ligated with 50 pmol of double-stranded insert using 1.25 Weiss U of T4 DNA ligase (Thermo Fisher Scientific #EL0011) at 22 °C for 60 min. Correct assembly of recombinant plasmids was verified by sequencing.

The derived plasmids pTKO-sp1, pTKO-sp2 and pTKO-sp3 were transformed into wild type *N. oceanica* as described above, but using 3 µg of circular plasmid instead of linear DNA, and in the presence of 30 µg denatured salmon sperm DNA (Thermo Fisher Scientific #AM9680) per transformation reaction before applying the electroporation pulse. The electroporation mixture was plated on Zeocin-containing agar for 4 weeks, after which transformants were re-streaked on fresh Zeocin-containing agar. After 1 week, colonies were transferred to 48 well microtiter plates with Zeocin-containing ASW-NB medium. After 1 week at ML conditions, 1.5 µl of culture was used as template in 30 µl PCR reactions, using Phire polymerase (Thermo Fisher Scientific #F160) with the following thermocycling conditions: 5 min at 98 °C, 35 cycles (5 s at 98 °C, 5 s at 61.2 °C, 60 s at 72 °C), and 3 min at 72 °C. Oligonucleotides 5'-ACTTCACAAGGTATCTCACCTCATC-3' and 5'-CTAAGGGGAGTGGAAAGAAAAG-3' were used as PCR primers at 0.25 µM. PCR products were purified and analyzed using Sanger sequencing at MacroGen Europe B.V. (Amsterdam, Netherlands).

#### 4.13. mRNA sequencing

For transcriptomic analysis, microalgal cultures were inoculated in ASW-NB at an OD<sub>750</sub> of 0.1 and cultivated in 50 ml shake flasks at 600 µmol m<sup>-2</sup> s<sup>-1</sup> in an Algem HT24 photobioreactor (Algenuity, UK) operated as described in section 3.1. After 3 d, 22 ml of culture was harvested (4255×g, 10 min, 4 °C) and cell pellets were snap-frozen. RNA extraction, mRNA library preparation and sequencing, transcript mapping and quantification were performed by Novogene (UK) Company Limited (N = 3). RNA was extracted from cell pellets using the RNeasy Plus Universal Mini Kit (Qiagen #73404) according to the manufacturer's instructions. RNA purity was verified on 1% agarose gels and using a NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA integrity and quantity was assessed using 1% agarose gels and the Bioanalyzer 2100 RNA 6000 Nano Kit (Agilent Technologies, CA, USA). A total of 1 µg RNA was used for RNA-seq library preparation with NEBNext Ultra<sup>TM</sup> RNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's instructions. Briefly, poly-T oligonucleotides attached to magnetic beads were used to purify mRNA from total RNA samples. mRNA was fragmented at elevated temperatures in NEBNext First Strand Synthesis Reaction Buffer (5X) and first strand cDNA was reverse transcribed using random hexamers and M-MuLV reverse transcriptase (RNase H-). Second strand cDNA was synthesized using DNA Polymerase I and RNase H. Blunt ends were created by exonuclease/polymerase treatment. 3'-DNA ends were adenylated to allow for ligation of NEBNext adaptors, containing a hairpin loop structure. Fragments of ~150–200 nucleotides length were purified using the AMPure XP system (Beckman Coulter, Beverly, USA). Hairpin loops were opened by treatment of size-selected cDNA fragments with 3 µl USER enzyme (NEB,

USA) at 37 °C for 15 min, followed by heat inactivation of the enzyme at 95 °C for 5 min. Linear double-stranded cDNA fragments were subjected to PCR using Phusion High-Fidelity DNA polymerase with Index primers or Universal PCR primers provided for the NEBNext Ultra<sup>TM</sup> RNA Library Prep system. Barcoded PCR products were purified using the AMPure XP system, and the library quality was checked using an Agilent Bioanalyzer 2100. Index-coded samples were clustered using a cBot Cluster Generation System with the PE Cluster Kit cBot-HS (Illumina) according to the manufacturer's recommendations. After this, libraries were sequenced on an Illumina Novaseq 6000 platform and paired-end reads were generated. FASTQ raw reads were preprocessed using fastp (Chen et al., 2018) to remove reads of low quality and reads containing poly-N or adapter sequences.

The *N. oceanica* IMET1v2 reference genome and gene model annotation were retrieved from <ftp://nandesyn.single-cell.cn/pub/> (Gong et al., 2020). Cleaned paired-end reads were mapped to the reference genome using HISAT2 (Kim et al., 2019). Mapped reads were quantified using the featureCounts package (Liao et al. 2014) of R statistical software (R Development Core Team, 2018). FPKM values were calculated for gene count comparisons within a sample. For differential expression analysis, gene counts were normalized and compared for log<sub>2</sub> fold change and statistical significant difference between sample groups using the DESeq2 package (Love et al. 2014) of the R Bioconductor project (Gentleman et al., 2004). The false discovery rate was controlled by adjusting the p-values using the Benjamini and Hochberg correction (Benjamini and Hochberg, 1995). Genes were considered differentially expressed between sample groups for  $p_{\text{adjust}} \leq 0.05$ .

Gene candidates associated with specific metabolic pathways were retrieved from the *Nannochloropsis* biochemical pathway database NannoCyc (Gong et al., 2020). Functional gene affiliations in this database are predicted using PathwayTools (Karp et al. 2002). Gene candidates for different biochemical pathways were then pooled with non-redundant genes of the same functions identified in previous studies (Vieler et al., 2012; Dong et al., 2013; Li et al., 2014; Hu et al., 2014; Poliner et al., 2015). Gene sets were manually curated by analyzing the encoded amino acid sequences for similarities with functionally annotated proteins in the *Non-redundant protein sequences* database of the NCBI (Agarwala et al., 2018), using the protein-protein BLAST feature operated with standard algorithm parameters. Proteins were further analyzed for conserved domains and domain architecture using the NCBI's conserved domain database (Marchler-Bauer et al., 2017). Sub-cellular localization of encoded proteins was analyzed using the *in silico* prediction tools HECTAR 1.3 (Gschoeßl et al. 2008), SignalP 5.0 (Armenteros et al., 2019), TargetP 2.0 (Armenteros et al., 2019), ChloroP 1.1 (Emanuelsson et al. 1999), MitoFates 1.2 (Fukasawa et al., 2015) and for certain genes using PredPlantPTS1 (Lingner et al., 2011). Log<sub>2</sub> fold changes,  $p_{\text{adjust}}$  values and localization predictions for genes from different biochemical pathways are listed in supplementary datasets S1-4.

#### 4.14. Statistical data treatment

R statistical software (R Development Core Team, 2018) was used for all data processing and statistical tests. Two-way ANOVA was employed to test for significant main effects. In case of a significant ANOVA outcome, Tukey's HSD test was used to compare the means of multiple groups with each other. Differences were considered significant in case of  $p < 0.05$ . When only two groups were compared, Welch's two-tailed *t*-test was employed instead of Tukey's HSD test. For quantitative flow cytometry analyses, we considered the median of the fluorescence distribution of gated events in a specific channel (FL1-A for measuring BODIPY fluorescence, FL5-A for measuring chlorophyll *a* auto-fluorescence) as representative value for one sample.

## Data availability

The RNA-Seq data produced in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus GSE167058.

## CRedit authorship contribution statement

**Christian Südfeld:** Conceptualization, Methodology, Investigation, Validation, Software, Formal analysis, Writing – original draft, Writing – review & editing. **Míchal Hubáček:** Conceptualization, Methodology, Investigation. **Daniel Figueiredo:** Conceptualization, Methodology, Investigation. **Mihris I.S. Naduthodi:** Conceptualization, Methodology. **John van der Oost:** Writing – review & editing. **René H. Wijffels:** Conceptualization, Project administration, Supervision, Writing – review & editing. **Maria J. Barbosa:** Project administration, Supervision, Writing – review & editing, Funding acquisition. **Sarah D'Adamo:** Project administration, Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was part of the Netherlands Organization of Scientific Research (NWO) Building Blocks of Life programme (grant number 737.016.007), and of the MAGNIFICENT project, funded by the Bio-Based Industries Joint Undertaking under the European Union's Horizon 2020 research and innovation program (grant number 745754).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jymben.2021.04.012>.

## Statement of informed consent, human/animal rights

No conflicts, informed consent, or human or animal rights are applicable to this study.

## References

- Agarwala, Richa, Barrett, Tanya, Beck, Jeff, Benson, Dennis A., Bollin, Colleen, Bolton, Evan, Bourexis, Devon, et al., 2018. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 46 (D1), D8–D13. <https://doi.org/10.1093/nar/gkx1095>.
- Ahmad, Atta, Bhattacharya, Akash, McDonald, Ramsay A., Cordes, Melissa, Ellington, Benjamin, Bertelsen, Eric B., Zunderweg, Erik R.P., 2011. Heat shock protein 70 kDa chaperone/DnaJ co-chaperone complex employs an unusual dynamic interface. *Proc. Natl. Acad. Sci. U.S.A.* 108 (47), 18966–18971. <https://doi.org/10.1073/pnas.1111220108>.
- Ajjawi, Imad, Verruto, John, Aquí, Moena, Soriaga, Leah B., Coppersmith, Jennifer, Kwok, Kathleen, Peach, Luke, et al., 2017. Lipid production in *Nannochloropsis gaditana* by decreasing expression of a single transcriptional regulator. *Nat. Biotechnol.* 35 (7), 647–652. <https://doi.org/10.1038/nbt.3865>.
- Alboresi, Alessandro, Perin, Giorgio, Vitulo, Nicola, Diretto, Gianfranco, Block, Maryse, Jouhet, Juliette, Meneghesso, Andrea, et al., 2016. Light remodels lipid biosynthesis in *Nannochloropsis gaditana* by modulating carbon partitioning between organelles. *Plant Physiol.* 171 (4), 2468–2482. <https://doi.org/10.1104/pp.16.00599>.
- Alonso, José M., Stepanova, Anna N., Leisse, Thomas J., Kim, Christopher J., Chen, Huaming, Shinn, Paul, Stevenson, Denise K., et al., 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science* 301 (5633), 653–657. <https://doi.org/10.1126/science.1086391>.
- Archambault, Steven, Starbuck Downes, Cara Meghan, Van Voorhies, Wayne, Erickson, Christopher A., Lammers, Peter, 2014. *Nannochloropsis* sp. algae for use as biofuel: analyzing a translog production function using data from multiple sites in the southwestern United States. *Algal Research* 6 (PB), 124–131. <https://doi.org/10.1016/j.algal.2014.10.005>.

- Armenteros, Almagro, Juan, José, Tsigiris, Konstantinos D., Sønderby, Casper Kaae, Petersen, Thomas Nordahl, Winther, Ole, Brunak, Søren, Gunnar von Heijne, Nielsen, Henrik, 2019a. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* 37 (4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>.
- Armenteros, Jose Juan Almagro, Salvatore, Marco, Emanuelsson, Olof, Winther, Ole, Von Heijne, Gunnar, Elofsson, Arne, Nielsen, Henrik, 2019b. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance* 2 (5). <https://doi.org/10.26508/lsa.201900429>.
- Arnold, C., Hodgson, L.J., 1991. Vectorette PCR: a novel approach to genomic walking. *Genome Res.* 1 (1), 39–42. <https://doi.org/10.1101/gr.1.1.39>.
- Auerbach, C., Robson, J.M., 1944. Production of Mutations by Allyl Isothiocyanate. "Nature Publishing Group. <https://doi.org/10.1016/154081a0>.
- Austin, Christopher P., Battey, James F., Bradley, Allan, Bucan, Maja, Capecchi, Mario, Collins, Francis S., Dove, William F., et al., 2004. The Knockout Mouse Project. "Nature Publishing Group.
- Baker, Michael J., Frazier, Ann E., Gulbis, Jacqueline M., Ryan, Michael T., 2007. Mitochondrial Protein-Import Machinery: Correlating Structure with Function. *Elsevier Current Trends*. <https://doi.org/10.1016/j.tcb.2007.07.010>.
- Basseri, Sana, Austin, Richard C., 2012. Endoplasmic reticulum stress and lipid metabolism: mechanisms and therapeutic potential. <https://doi.org/10.1155/2012/841362>.
- Benjamini, Yoav, Hochberg, Yosef, 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B* 57 (1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Benvenuti, Giulia, Ruiz, Jesús, Lamers, Packo P., Bosma, Rouke, Wijffels, René H., Barbosa, Maria J., 2017. Towards microalgal triglycerides in the commodity markets. *Biotechnol. Biofuels* 10 (1), 188. <https://doi.org/10.1186/s13068-017-0873-2>.
- Beutler, Bruce, Jiang, Zhengfan, Georgel, Philippe, Crozat, Karine, Croker, Ben, Rutschmann, Sophie, Du, Xin, Hoebe, Kasper, 2006. Genetic analysis of host resistance: toll-like receptor signaling and immunity at large. *Annu. Rev. Immunol.* 24 (1), 353–389. <https://doi.org/10.1146/annurev.immunol.24.021605.090552>.
- Blaby-Haas, Crysten E., Merchant, Sabeeha S., 2019. Comparative and Functional Algal Genomics. *Annu. Rev. Inc.* <https://doi.org/10.1146/annurev-arplant-050718-095841>.
- Bragantini, Benoit, Tiotiu, Decebal, Rothé, Benjamin, Saliou, Jean Michel, Marty, Héléne, Cianféran, Sarah, Charpentier, Bruno, Quinternet, Marc, Manival, Xavier, 2016. Functional and structural insights of the zinc-finger HIT protein family members involved in box C/D snoRNP biogenesis. *J. Mol. Biol.* 428 (11), 2488–2506. <https://doi.org/10.1016/j.jmb.2016.04.028>.
- Brocard, Cécile, Hartig, Andreas, 2006. "Peroxisome Targeting Signal 1: Is it Really a Simple Tripeptide?" *Elsevier*. <https://doi.org/10.1016/j.bbamcr.2006.08.022>.
- Calder, Philip C., 2006. n-3 Polyunsaturated fatty acids, inflammation, and inflammatory diseases. In: *American Journal of Clinical Nutrition*, 83:1505S–1519S. 6. American Society for Nutrition. <https://doi.org/10.1093/ajcn/83.6.1505s>.
- Cavalier-Smith, T., 2018. Kingdom Chromista and its eight phyla: a new synthesis emphasising periplastid protein targeting, cytoskeletal and periplastid evolution, and ancient divergences. *Protoplasma* 255 (1), 297–357. <https://doi.org/10.1007/s00709-017-1147-3>.
- Cecchin, Michela, Berteotti, Silvia, Paltrinieri, Stefania, Vigliante, Ivano, Iadarola, Barbara, Giovannone, Barbara, Maffei, Massimo E., Delledonne, Massimo, Ballottari, Matteo, 2020. Improved lipid productivity in *Nannochloropsis gaditana* in nitrogen-replete conditions by selection of pale green mutants. *Biotechnol. Biofuels* 13 (1), 78. <https://doi.org/10.1186/s13068-020-01718-8>.
- Chen, Shifu, Zhou, Yanqing, Chen, Yaru, Gu, Jia, 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
- Chini Zittelli, G., Lavista, F., Bastianini, A., Rodolfi, L., Vincenzini, M., Tredici, M.R., 1999. Production of eicosapentaenoic acid by *Nannochloropsis* sp. cultures in outdoor tubular photobioreactors. *J. Biotechnol.* 70, 299–312. [https://doi.org/10.1016/S0168-1656\(99\)00082-6](https://doi.org/10.1016/S0168-1656(99)00082-6), 1–3. Elsevier Sci B.V.
- Naduthodi, Mihris Ibnu Saleem, Claessens, Nico J., D'Adamo, Sarah, van der Oost, John, Barbosa, Maria J., 2021. Synthetic biology approaches to enhance microalgal productivity. *Trends Biotechnol.* <https://doi.org/10.1016/j.tibtech.2020.12.010>. February.
- Cowell, I.G., 1997. "Yeast two-hybrid library screening. *Methods in Molecular Biology (Clifton, N.J.)* 69, 185–202. <https://doi.org/10.1385/0-89603-383-x:185>.
- Daboussi, Fayza, Leduc, Sophie, Maréchal, Alan, Dubois, Gwendoline, Guyot, Valérie, Perez-Michaut, Christophe, Amato, Alberto, et al., 2014. Genome engineering empowers the diatom *Phaeodactylum tricornutum* for biotechnology. *Nat. Commun.* 5 (May), 98–106. <https://doi.org/10.1038/ncomms4831>.
- Derelle, R., López-García, P., Timpano, H., Moreira, D., 2016. A phylogenomic framework to study the diversity and evolution of stramenopiles (=Heterokonts). *Mol. Biol. Evol.* 33 (11), 2890–2898. <https://doi.org/10.1093/molbev/msw168>.
- Doan, Thi Thai Yen, Obbard, Jeffrey Philip, 2012. Enhanced intracellular lipid in *Nannochloropsis* sp. via random mutagenesis and flow cytometric cell sorting. *Algal Research* 1 (1), 17–21. <https://doi.org/10.1016/j.algal.2012.03.001>.
- Dolch, Lina Juana, Rak, Camille, Perin, Giorgio, Tourcier, Guillaume, Broughton, Richard, Leterrier, Marina, Morosinotto, Tomas, et al., 2017. A palmitic acid elongase affects eicosapentaenoic acid and plastidial monogalactosyldiacylglycerol levels in *nannochloropsis*. *Plant Physiol.* 173 (1), 742–759. <https://doi.org/10.1104/pp.16.01420>.
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. "Touchdown" PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* 19 (14), 4008. <https://doi.org/10.1093/nar/19.14.4008>.

- Dong, Hong Po, Williams, Ernest, Wang, Da Zhi, Xie, Zhang Xian, Hsia, Ru Ching, Jenck, Alizée, Halden, Rolf, Li, Jing, Chen, Feng, Place, Allen R., 2013. Responses of *Nannochloropsis oceanica* IMET1 to long-term nitrogen starvation and recovery. *Plant Physiol.* 162 (2), 1110–1126. <https://doi.org/10.1104/pp.113.214320>.
- Cabanelas, Iago Teles Dominguez, Van Der Zwart, Mathijs, Kleinegris, Dorinde M.M., Wijffels, René H., Barbosa, Maria J., 2016. Sorting cells of the microalga *Chlorococcum littorale* with increased triacylglycerol productivity. *Biotechnol. Biofuels* 9 (1), 183. <https://doi.org/10.1186/s13068-016-0595-x>.
- Dyson, Beth C., Allwood, J. William, Feil, Regina, Xu, Yun, Miller, Matthew, Bowsher, Caroline G., Goodacre, Royston, Lunn, John E., Johnson, Giles N., 2015. Acclimation of metabolism to light in *Arabidopsis thaliana*: the glucose 6-phosphate/phosphate translocator GPT2 directs metabolic acclimation. *Plant Cell Environ.* 38 (7), 1404–1417. <https://doi.org/10.1111/pce.12495>.
- Emanuelsson, Olof, Nielsen, Henrik, Von Heijne, Gunnar, 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8 (5), 978–984. <https://doi.org/10.1110/ps.8.5.978>.
- Erfellinck, Marie Laure, Ribeiro, Bianca, Perassolo, Maria, Pauwels, Laurens, Pollier, Jacob, Storme, Veronique, Goossens, Alain, 2018. “A user-friendly platform for yeast two-hybrid library screening using next generation sequencing.” Edited by Alan M. Moses. *PLoS One* 13 (12), e0201270. <https://doi.org/10.1371/journal.pone.0201270>.
- Fei, Weihua, Wang, Han, Bielby, Christopher, Yang, Hongyuan, 2009. Conditions of endoplasmic reticulum stress stimulate lipid droplet formation in *Saccharomyces cerevisiae*. *Biochem. J.* 424 (1), 61–67. <https://doi.org/10.1042/BJ20090785>.
- Fransen, Marc, Amery, Leen, Hartig, Andreas, Brees, Chantal, Rabijns, Anja, Mannaerts, Guy P., Van Veldhoven, Paul P., 2008. Comparison of the PTS1- and rab8b-binding properties of Pex5p and Pex5Rp/TRIP8b. *Biochim. Biophys. Acta Mol. Cell Res.* 1783 (5), 864–873. <https://doi.org/10.1016/j.bbamcr.2008.02.013>.
- Fukasawa, Yoshinori, Tsuji, Junko, Fu, Suzu Chin, Tomii, Kentaro, Horton, Paul, Imai, Kenichiro, 2015. MitoFates: improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol. Cell. Proteomics* 14 (4), 1113–1126. <https://doi.org/10.1074/mcp.M114.043083>.
- Gatsos, Xenia, Perry, Andrew J., Anwari, Khatira, Dolezal, Pavel, Wolyneec, P. Peter, Likić, Vladimir A., Purcell, Anthony W., Buchanan, Susan K., Lithgow, Trevor, 2008. *Protein Secretion and Outer Membrane Assembly in Alphaproteobacteria*. Oxford Academic.
- Genevaux, Pierre, Schwager, Françoise, Georgopoulos, Costa, Kelley, William L., 2002. Scanning mutagenesis identifies amino acid residues essential for the in vivo activity of the *Escherichia coli* DnaJ (Hsp40) J-domain. *Genetics* 162 (3), 1045–1053. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1462316/>.
- Gentleman, Robert C., Carey, Vincent J., Bates, Douglas M., Bolstad, Ben, Dettling, Marcel, Dudoit, Sandrine, Ellis, Byron, et al., 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 5 (10) <https://doi.org/10.1186/gb-2004-5-10-r80>.
- Ghazaei, Ciamak, 2017. Role and mechanism of the Hsp70 molecular chaperone machines in bacterial pathogens. *J. Med. Microbiol.* 66 (3), 259–265. <https://doi.org/10.1099/jmm.0.000429>.
- Gong, Yanhai, Kang, Nam Kyu, Kim, Young Uk, Wang, Zengbin, Wei, Li, Xin, Yi, Shen, Chen, et al., 2020. The NanDeSyn Database for *Nannochloropsis* systems and synthetic biology. *Plant J.* <https://doi.org/10.1111/tpl.15025>, October, tpl.15025.
- González-Ballester, David, De Montaigne, Amaury, Galván, Aurora, Fernández, Emilio, 2005. Restriction enzyme site-directed amplification PCR: a tool to identify regions flanking a marker DNA. *Anal. Biochem.* 340 (2), 330–335. <https://doi.org/10.1016/j.ab.2005.01.031>.
- Goodman, Andrew L., McNulty, Nathan P., Zhao, Yue, Leip, Douglas, Mitra, Robi D., Lozupone, Catherine A., Knight, Rob, Gordon, Jeffrey I., 2009. Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* 6 (3), 279–289. <https://doi.org/10.1016/j.chom.2009.08.003>.
- Greco, R., Ouwerkerk, P.B.F., Sallaud, C., Kohli, A., Colombo, L., Puigdomènech, P., Guiderdoni, E., Christou, P., Hoge, J.H.C., Pereira, A., 2001. *Transposon Insertional Mutagenesis in Rice*. *American Society of Plant Biologists*, 10.1104/125.3.1175.
- Gschloessl, Bernhard, Guermeur, Yann, Cock, J. Mark, 2008. HECTAR: a method to predict subcellular targeting in heterokonts. *BMC Bioinf.* 9, 393. <https://doi.org/10.1186/1471-2105-9-393>, September.
- Hanschen, Erik R., Starkenburg, Shawn R., 2020. The State of Algal Genome Quality and Diversity. *Elsevier B.V.* <https://doi.org/10.1016/j.algal.2020.101968>.
- Harewood, Louise, Schütz, Frédéric, Boyle, Shelagh, Perry, Paul, Delorenzi, Mauro, Bickmore, Wendy A., Reymond, Alexandre, 2010. The effect of translocation-induced nuclear reorganization on gene expression. *Genome Res.* 20 (5), 554–564. <https://doi.org/10.1101/gr.103622.109>.
- Hirano, Tatsuya, Kinoshita, Noriyuki, Morikawa, Kosuke, Yanagida, Mitsuhiro, 1990. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc 2+. *Cell* 60 (2), 319–328. [https://doi.org/10.1016/0092-8674\(90\)90746-2](https://doi.org/10.1016/0092-8674(90)90746-2).
- Hu, Jianqiang, Wang, Dongmei, Li, Jing, Gongchao, Ning, Kang, Xu, Jian, 2014. Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae *Nannochloropsis*. *Sci. Rep.* 4 (1), 5454. <https://doi.org/10.1038/srep05454>.
- Huang, Zixia, Teeling, Emma C., 2017. “ExUTR: a novel pipeline for large-scale prediction of 3'-UTR sequences from NGS data. *BMC Genom.* 18 (1), 847. <https://doi.org/10.1186/s12864-017-4241-1>.
- Hyka, P., Lickova, S., Příbyl, P., Melzoch, K., Kovar, K., 2013. Flow cytometry for the development of biotechnological processes with microalgae. <https://doi.org/10.1016/j.biotechadv.2012.04.007>.
- Janssen, Jorijn H., Lamers, Packo P., de Vos, Ric C.H., Wijffels, René H., Barbosa, Maria J., 2019. Translocation and de novo synthesis of eicosapentaenoic acid (EPA) during nitrogen starvation in *Nannochloropsis gaditana*. *Algal Research* 37 (January), 138–144. <https://doi.org/10.1016/j.algal.2018.11.025>.
- Janssen, Jorijn H., Spoelder, Jacco, Koehorst, Jasper J., Schaap, Peter J., Wijffels, René H., Barbosa, Maria J., 2020. Time-dependent transcriptome profile of genes involved in triacylglycerol (TAG) and polyunsaturated fatty acid synthesis in *Nannochloropsis gaditana* during nitrogen starvation. *J. Appl. Phycol.* 32 (2), 1153–1164. <https://doi.org/10.1007/s10811-019-02021-2>.
- Joyard, Jacques, Ferro, Myriam, Masselon, Christophe, Seigneurin-Berny, Daphné, Salvi, Daniel, Garin, Jérôme, Rolland, Norbert, 2010. Chloroplast Proteomics Highlights the Subcellular Compartmentation of Lipid Metabolism. *Pergamon*. <https://doi.org/10.1016/j.plipres.2009.10.003>.
- Kanehisa, Minoru, Goto, Susumu, 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Oxford University Press. <https://doi.org/10.1093/nar/28.1.27>.
- Karp, Peter D., Paley, Suzanne, Romero, Pedro, 2002. The pathway tools software. *Bioinformatics* 18 (1), S225–S232. [https://doi.org/10.1093/bioinformatics/18.suppl\\_1.S225](https://doi.org/10.1093/bioinformatics/18.suppl_1.S225). Oxford University Press.
- Kim, Daehwan, Paggi, Joseph M., Park, Chanhee, Bennett, Christopher, Salzberg, Steven L., 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Knothe, Gerhard, 2008. “Designer” biodiesel: optimizing fatty ester composition to improve fuel properties. *Energy Fuel.* 22 (2), 1358–1364. <https://doi.org/10.1021/ef700639e>.
- Knudsen, Jens, Mandrup, Susanne, Rasmussen, Jan Trige, Andreassen, Per Hove, Poulsen, Flemming, Kristiansen, Karsten, 1993. The function of acyl-CoA-binding protein (ACBP)/Diazepam binding inhibitor (DBI). *Mol. Cell. Biochem.* 123 (1–2), 129–138. <https://doi.org/10.1007/BF01076484>.
- Kong, Que, Yuan, Ling, Ma, Wei, 2019. Wrinkled 1, a ‘master regulator’ in transcriptional control of plant oil biosynthesis. *MDPI AG*. <https://doi.org/10.3390/plants8070238>.
- Koo, Abraham J.K., Ohlrogge, John B., Pollard, Mike, 2004. On the export of fatty acids from the chloroplast. *J. Biol. Chem.* 279 (16), 16101–16110. <https://doi.org/10.1074/jbc.M311305200>.
- Labun, Kornel, Montague, Tessa G., Krause, Maximilian, Torres Cleuren, Yamila N., Tjeldnes, Håkon, Valen, Eivind, 2019. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res.* 47 (W1), W171–W174. <https://doi.org/10.1093/nar/gkz365>.
- Lamb, John R., Tugendreich, Stuart, Hieter, Phil, 1995. Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem. Sci.* 20 (7), 257–259. [https://doi.org/10.1016/S0968-0004\(00\)89037-4](https://doi.org/10.1016/S0968-0004(00)89037-4).
- Li, Jing, Han, Danxiang, Wang, Dongmei, Ning, Kang, Jia, Jing, Wei, Li, Jing, Xiaoyan, et al., 2014. Choreography of transcriptomes and lipidomes of *Nannochloropsis* reveals the mechanisms of oil synthesis in microalgae. *Plant Cell* 26 (4), 1645–1665. <https://doi.org/10.1105/tpc.113.121418>.
- Li, Xiaobo, Zhang, Ru, Patena, Weronika, Gang, Spencer S., Blum, Sean R., Ivanova, Nina, Yue, Rebecca, et al., 2015. An indexed, mapped mutant library enables reverse genetics studies of biological processes in *Chlamydomonas reinhardtii*. *Plant Cell* 28 (2), 367–387. <https://doi.org/10.1105/tpc.15.00465>.
- Li, Da Wei, Balamurugan, Srinivasan, Yang, Yu Feng, Zheng, Jian Wei, Huang, Dan, Zou, Li Gong, Yang, Wei Dong, Liu, Jie Sheng, Guan, Yuanfang, Li, Hong Ye, 2019. Transcriptional regulation of microalgae for concurrent lipid overproduction and secretion. *Science Advances* 5 (1), eaau3795. <https://doi.org/10.1126/sciadv.aau3795>.
- Li-Beisson, Yonghua, Shorrosh, Basil, Beisson, Fred, Andersson, Mats X., Arondel, Vincent, Bates, Philip D., Baud, Sébastien, et al., 2010. Acyl-lipid metabolism. *Arabidopsis Book* 8, e0133. <https://doi.org/10.1199/tab.0133>, January.
- Liao, Yang, Smyth, Gordon K., Shi, Wei, 2014. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
- Lingner, Thomas, Kataya, Amr R., Antonicelli, Gerardo E., Benichou, Aline, Nilssen, Kjersti, Chen, Xiong Yan, Siemsen, Tanja, Morgenstern, Burkhard, Meinicke, Peter, Reumann, Sigrun, 2011. Identification of novel plant peroxisomal targeting signals by a combination of machine learning methods and in vivo subcellular targeting analyses. *Plant Cell* 23 (4), 1556–1572. <https://doi.org/10.1105/tpc.111.084095>.
- Liu, Junying, Song, Yunneng, Qiu, Wen, 2017. Oleaginous microalgae *Nannochloropsis* as a new model for biofuel production: review & analysis. *Renew. Sustain. Energy Rev.* 72, 154–162. <https://doi.org/10.1016/j.rser.2016.12.120>.
- Love, Michael I., Huber, Wolfgang, Anders, Simon, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Lu, Rwei Min, Hwang, Yu Chyi, Liu, I. Ju, Lee, Chi Chiu, Tsai, Han Zen, Li, Hsin Jung, Wu, Han Chung, 2020. *Development of Therapeutic Antibodies for the Treatment of Diseases*. *BioMed Central Ltd.*
- Lu, Yandu, Gan, Qinhua, Iwai, Masakazu, Alboresi, Alessandro, Burlacot, Adrien, Dautermann, Oliver, Takahashi, Hiroko, et al., 2021. Role of an ancient light-harvesting protein of PSI in light absorption and photoprotection. *Nat. Commun.* 12 (1), 679. <https://doi.org/10.1038/s41467-021-20967-1>.
- Ma, Yubin, Wang, Zhiyao, Yu, Changjiang, Yin, Yehu, Zhou, Gongke, 2014. Evaluation of the potential of 9 *Nannochloropsis* strains for biodiesel production. *Bioresour. Technol.* 167, 503–509. <https://doi.org/10.1016/j.biortech.2014.06.047>, September.
- Ma, Xiao Nian, Chen, Tian Peng, Yang, Bo, Liu, Jin, Chen, Feng, 2016. Lipid production from *nannochloropsis*. *Mar. Drugs* 14 (4), 61. <https://doi.org/10.3390/md14040061>.

- Majerus, P.W., Kilburn, E., 1969. Acetyl coenzyme A carboxylase. The roles of synthesis and degradation in regulation of enzyme levels in rat liver. *J. Biol. Chem.* 244 (22), 6254–6262. [https://doi.org/10.1016/S0021-9258\(18\)63531-8](https://doi.org/10.1016/S0021-9258(18)63531-8).
- Marchler-Bauer, Aron, Bo, Yu, Han, Lianyi, He, Jane, Lanczycki, Christopher J., Lu, Shennan, Chitsaz, Farideh, et al., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45 (D1), D200–D203. <https://doi.org/10.1093/nar/gkw1129>.
- Masuda, Tatsuru, 2008. Recent Overview of the Mg Branch of the Tetrapyrrole Biosynthesis Leading to Chlorophylls. Springer. <https://doi.org/10.1007/s11200-008-9291-4>.
- Meng, Yingying, Jiang, Junpeng, Wang, Haitao, Cao, Xupeng, Xue, Song, Yang, Qing, Wang, Weiliang, 2015. The characteristics of TAG and EPA accumulation in *Nannochloropsis oceanica* IMET1 under different nitrogen supply regimes. *Bioresour. Technol.* 179 (March), 483–489. <https://doi.org/10.1016/j.biortech.2014.12.012>.
- Mittler, Ron, 2002. Oxidative Stress, Antioxidants and Stress Tolerance. Elsevier Current Trends. [https://doi.org/10.1016/S1360-1385\(02\)02312-9](https://doi.org/10.1016/S1360-1385(02)02312-9).
- Naduthodi, Mihris Ibnu Saleem, Mohanraj, Prarthana, Südfeld, Christian, D'Adamo, Sarah, Barbosa, Maria J., Van Der Oost, John, 2019. CRISPR-Cas ribonucleoprotein mediated homology-directed repair for efficient targeted genome editing in microalgae *Nannochloropsis oceanica* IMET1. *Biotechnol. Biofuels* 12 (1), 66. <https://doi.org/10.1186/s13068-019-1401-3>.
- Moresco, Eva Marie Y., Li, Xiaohong, Beutler, Bruce, 2013. Going forward with genetics: recent technological advances and forward genetics in mice. *Am. J. Pathol.* 182 (5), 1462–1473. <https://doi.org/10.1016/j.ajpath.2013.02.002>.
- Mozaffarian, Dariush, Wu, Jason H.Y., 2011. Omega-3 Fatty Acids and Cardiovascular Disease: Effects on Risk Factors, Molecular Pathways, and Clinical Events. Elsevier USA. <https://doi.org/10.1016/j.jacc.2011.06.063>.
- Mueller-Cajar, Oliver, Stotz, Mathias, Wendler, Petra, Hartl, F. Ulrich, Bracher, Andreas, Hayer-Hartl, Manajit, 2011. Structure and function of the AAA + protein CbbX, a red-type rubisco activase. Ulrich Hartl, andreas bracher, and manajit hayer-hartl *Nature* 479 (7372), 194–199. <https://doi.org/10.1038/nature10568>.
- Ohto, Masa Aki, Fischer, Robert L., Goldberg, Robert B., Nakamura, Kenzo, Harada, John J., 2005. Control of seed mass by APETALA2. *Proc. Natl. Acad. Sci. U.S.A.* 102 (8), 3123–3128. <https://doi.org/10.1073/pnas.0409858102>.
- Orlenko, Alena, Hermansen, Russell A., Liberles, David A., 2016. Flux control in glycolysis varies across the tree of life. *J. Mol. Evol.* 82 (2–3), 146–161. <https://doi.org/10.1007/s00239-016-9731-2>.
- Osoerio, Hector, Jara, Carol, Fuenzalida, Karen, Rey-Jurado, Emma, Vázquez, Mónica, 2019. High-efficiency nuclear transformation of the microalgae *Nannochloropsis oceanica* using Tn5 Transposome for the generation of altered lipid accumulation phenotypes. *Biotechnol. Biofuels* 12 (1), 134. <https://doi.org/10.1186/s13068-019-1475-y>.
- O'Hara, Paul, Slabas, Antoni R., Fawcett, Tony, 2002. Fatty acid and lipid biosynthetic genes are expressed at constant molar ratios but different absolute levels during embryogenesis. *Plant Physiol.* 129 (1), 310–320. <https://doi.org/10.1104/pp.010956>.
- Pereira, Hugo, Schulze, Peter S.C., Schüler, Lisa Maylin, Santos, Tamára, Barreira, Luísa, Varella, João, 2018. Fluorescence Activated Cell-Sorting Principles and Applications in Microalgal Biotechnology. Elsevier B.V. <https://doi.org/10.1016/j.algal.2017.12.013>.
- Poliner, Eric, Panchy, Nicholas, Newton, Linsey, Wu, Guangxi, Lapinsky, Andrew, Bullard, Blair, Zienkiewicz, Agnieszka, Benning, Christoph, Shiu, Shin Han, Farré, Eva M., 2015. Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles. *Plant J.* 83 (6), 1097–1113. <https://doi.org/10.1111/tpj.12944>.
- Poliner, Eric, Farré, Eva M., Benning, Christoph, 2018. “Advanced Genetic Tools Enable Synthetic Biology in the Oleaginous Microalgae *Nannochloropsis* sp.” *Plant Cell Reports*. <https://doi.org/10.1007/s00299-018-2270-0>.
- Popelkova, Hana, Yocum, Charles F., 2011. PsbO, the Manganese-Stabilizing Protein: Analysis of the Structure-Function Relations that Provide Insights into its Role in Photosystem II. Elsevier. <https://doi.org/10.1016/j.jphotobiol.2011.01.015>.
- Qiu, X.B., Shao, Y.M., Miao, S., Wang, L., 2006. The Diversity of the DnaJ/Hsp40 Family, the Crucial Partners for Hsp70 Chaperones. Springer. <https://doi.org/10.1007/s00018-006-6192-6>.
- R Development Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- Ram, Hasthi, Soni, Praveen, Salvi, Prafull, Gandass, Nishu, Sharma, Ankita, Kaur, Amandeep, Sharma, Tilak Raj, 2019. Insertional mutagenesis approaches and their use in rice for functional genomics. MDPI AG. <https://doi.org/10.3390/plants8090310>.
- Remmers, I.M., Hidalgo-Ulloa, A., Brandt, B.P., Evers, W.A.C., Wijffels, R.H., Lamers, P. P., 2017. Continuous versus Batch Production of Lipids in the Microalgae *Acetodesmus Obliquus*. *Bioresour. Technol.* <https://doi.org/10.1016/j.biortech.2017.04.093>.
- Riechmann, José Luis, Meyerowitz, Elliot M., 1998. The AP2/EREBP Family of Plant Transcription Factors. *De Gruyter*. <https://doi.org/10.1515/bchm.1998.379.6.633>.
- Rodolfi, Liliana, Zittelli, Graziella Chini, Bassi, Niccolo, Padovani, Giulia, Biondi, Natascia, Bonini, Gimena, Tredecchi, Mario R., 2009. Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnol. Bioeng.* 102 (1), 100–112. <https://doi.org/10.1002/bit.22033>.
- Rumin, Judith, Nicolau, Elodie, de Oliveira, Raimundo Gonçalves, Fuentes-Grünwald, Claudio, Flynn, Kevin J., Picot, Laurent, 2020. A bibliometric analysis of microalgae research in the world, Europe, and the European Atlantic area. MDPI AG. <https://doi.org/10.3390/md18020079>.
- Ryu, Ae Jin, Kang, Nam Kyu, Jeon, Seungjib, Hur, Dong Hoon, Lee, Eun Mi, Lee, Do Yup, Jeong, Byeong Ryool, Chang, Yong Keun, Jeong, Ki Jun, 2020. Development and characterization of a *Nannochloropsis* mutant with simultaneously enhanced growth and lipid production. *Biotechnol. Biofuels* 13 (1). <https://doi.org/10.1186/s13068-020-01681-4>.
- Sahi, Chandan, Craig, Elizabeth Anne, 2007. Network of general and specialty J protein chaperones of the yeast cytosol. *Proc. Natl. Acad. Sci. U.S.A.* 104 (17), 7163–7168. <https://doi.org/10.1073/pnas.0702357104>.
- San Pedro, A., González-López, C.V., Ación, F.G., Molina-Grima, E., 2015. Outdoor pilot production of *nannochloropsis gaditana*: influence of culture parameters and lipid production rates in raceway ponds. *Algal Research* 8, 205–213. <https://doi.org/10.1016/j.algal.2015.02.013>.
- Seol, Ja Hwan, Shim, Eun Yong, Lee, Sang Eun, 2018. Microhomology-mediated End Joining: Good, Bad and Ugly. Elsevier B.V.
- Shank, K.J., Su, P., Brglez, I., Boss, W.F., Dewey, R.E., Boston, R.S., 2001. Induction of lipid metabolic enzymes during the endoplasmic reticulum stress response in plants. *Plant Physiol.* 126 (1), 267–277. <https://doi.org/10.1104/pp.126.1.267>.
- Sharkey, Thomas D., Weise, Sean E., 2016. The glucose 6-phosphate shunt around the Calvin-Benson cycle. *J. Exp. Bot.* 67 (14), 4067–4077. <https://doi.org/10.1093/jxb/erv484>.
- Shelake, Rahul Mahadev, Pramanik, Dibyajyoti, Kim, Jae-Yean, 2019. Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. *Plant Biotechnology Reports* 13 (5), 423–445. <https://doi.org/10.1007/s11816-019-00562-z>.
- Shestov, Alexander A., Liu, Xiaojing, Ser, Zheng, Cluntun, Ahmad A., Hung, Yin P., Huang, Lei, Kim, Dongsung, et al., 2014. Quantitative determinants of aerobic glycolysis identify flux through the enzyme GAPDH as a limiting step. *eLife* 3 (July2014), 1–18. <https://doi.org/10.7554/eLife.03342>.
- Shi, Y., Liu, M., Ding, W., Liu, J., 2020. Novel insights into phosphorus deprivation boosted lipid synthesis in the marine alga *nannochloropsis oceanica* without compromising biomass production. *J. Agric. Food Chem.* 68 (41), 11488–11502. <https://doi.org/10.1021/acs.jafc.0c04899>.
- Siebert, Paul D., Chenchik, Alex, Kellogg, David E., Lukyanov, Konstantin A., Lukyanov, Sergey A., 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res.* 23 (6), 1087–1088. <https://doi.org/10.1093/nar/23.6.1087>.
- Stadler, L.J., 1928. Genetic effects of X-rays in maize. *Proc. Natl. Acad. Sci. Unit. States Am.* 14 (1), 69–75. <https://doi.org/10.1073/pnas.14.1.69>.
- Südfeld, Christian, Hubáček, Michal, D'Adamo, Sarah, Wijffels, René H., Barbosa, Maria J., 2020. Optimization of high-throughput lipid screening of the microalgae *Nannochloropsis oceanica* using BODIPY 505/515. *Algal Research* 53 (March), 102138. <https://doi.org/10.1016/j.algal.2020.102138>.
- Terashima, Mia, Freeman, Elizabeth S., Jinkerson, Robert E., Jonikas, Martin C., 2015. A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid *Chlamydomonas* mutants. *Plant J.* 81 (1), 147–159. <https://doi.org/10.1111/tpj.12682>.
- Van Campenhout, Claude, Cabochette, Pauline, Veillard, Anne Clémence, Laczik, Miklos, Zelisko-Schmidt, Agnieszka, Sabatel, Céline, Dhainaut, Maxime, Vanhollebeke, Benoit, Gueydan, Cyril, Kruijs, Véronique, 2019. Guidelines for optimized gene knockout using CRISPR/Cas9. *Biotechniques* 66 (6), 295–302. <https://doi.org/10.2144/btn-2018-0187>.
- Vieler, Astrid, Wu, Guangxi, Tsai, Chia Hong, Bullard, Blair, Cornish, Adam J., Harvey, Christopher, Reza, Ida Barbara, et al., 2012. Genome, functional gene annotation, and nuclear transformation of the heterokont oleaginous alga *nannochloropsis oceanica* CCMP1779. *PLoS Genet.* 8 (11), e1003064. <https://doi.org/10.1371/journal.pgen.1003064>.
- Vos, Michel J., Hageman, Jurre, Carra, Serena, Kampinga, Harm H., 2008. Structural and Functional Diversities between Members of the Human HSPB, HSPH, HSPA, and DNAJ Chaperone Families. American Chemical Society. <https://doi.org/10.1021/bi800639z>.
- Wang, Dongmei, Ning, Kang, Li, Jing, Hu, Jianqiang, Han, Danxiang, Wang, Hui, Zeng, Xiaowei, et al., 2014. *Nannochloropsis* genomes reveal evolution of microalgal oleaginous traits. *PLoS Genet.* 10 (1) <https://doi.org/10.1371/journal.pgen.1004094>.
- Wang, Qintao, Lu, Yandu, Xin, Yi, Li, Wei, Huang, Shi, Xu, Jian, 2016a. Genome editing of model oleaginous microalgae *Nannochloropsis* spp. by CRISPR/Cas9. *Plant J.* 88 (6), 1071–1081. <https://doi.org/10.1111/tpj.13307>.
- Wang, Song, Zhang, Lin, Yang, Guanpin, Han, Jichang, Thomsen, Laurenz, Pan, Kehou, 2016b. Breeding 3 elite strains of *Nannochloropsis oceanica* by nitrosoguanidine mutagenesis and robust screening. *Algal Research* 19, 104–108. <https://doi.org/10.1016/j.algal.2016.07.021>.
- Wei, L., El Hajjami, M., Shen, C., You, W., Lu, Y., Li, J., et al., 2019. Transcriptomic and proteomic responses to very low CO2 suggest multiple carbon concentrating mechanisms in *Nannochloropsis oceanica*. *Biotechnol. Biofuels* 12 (1). <https://doi.org/10.1186/s13068-019-1506-8>.
- Weise, Sean E., Liu, Tiffany, Childs, Kevin L., Preiser, Alyssa L., Katulski, Hailey M., Perrin-Porzondek, Christopher, Sharkey, Thomas D., 2019. Transcriptional regulation of the glucose-6-phosphate/phosphate translocator 2 is related to carbon exchange across the chloroplast envelope. *Front. Plant Sci.* 10 (June), 827. <https://doi.org/10.3389/fpls.2019.00827>.
- Xie, Zhouli, Nolan, Trevor M., Jiang, Hao, Yin, Yanhai, 2019. AP2/ERF Transcription Factor Regulatory Networks in Hormone and Abiotic Stress Responses in Arabidopsis. *Frontiers Media S.A.* <https://doi.org/10.3389/fpls.2019.00228>.

- Xiong, Yan, Contento, Anthony L., Nguyen, Phan Quang, Bassham, Diane C., 2007. Degradation of oxidized proteins by autophagy during oxidative stress in arabidopsis. *Plant Physiol.* 143 (1), 291–299. <https://doi.org/10.1104/pp.106.092106>.
- Yamaoka, Yasuyo, Shin, Seungjun, Choi, Bae Young, Kim, Hanul, Jang, Sunghoon, Kajikawa, Masataka, Yamano, Takashi, et al., 2019. The bZIP1 transcription factor regulates lipid remodeling and contributes to ER stress management in *chlamydomonas reinhardtii*. *Plant Cell* 31 (5), 1127–1140. <https://doi.org/10.1105/tpc.18.00723>.
- Yang, Bo, Liu, Jin, Ma, Xiaonian, Guo, Bingbing, Liu, Bin, Wu, Tao, Jiang, Yue, Chen, Feng, 2017. Genetic engineering of the Calvin cycle toward enhanced photosynthetic CO<sub>2</sub> fixation in microalgae. *Biotechnol. Biofuels* 10 (1), 229. <https://doi.org/10.1186/s13068-017-0916-8>.
- Yao, Lina, Shen, Hui, Wang, Nan, Tatlay, Jaspaul, Li, Liang, Tan, Tin Wee, Lee, Yuan Kun, 2017. Elevated acetyl-CoA by amino acid recycling fuels microalgal neutral lipid accumulation in exponential growth phase for biofuel production. *Plant Biotechnology Journal* 15 (4), 497–509. <https://doi.org/10.1111/pbi.12648>.
- Yi, Xiaoping, McChargue, Myriam, Laborde, Susan, Frankel, Laurie K., Bricker, Terry M., 2005. The manganese-stabilizing protein is required for photosystem II assembly/stability and photoautotrophy in higher plants. *J. Biol. Chem.* 280 (16), 16170–16174. <https://doi.org/10.1074/jbc.M501550200>.
- Youker, Robert T., Brodsky, Jeffrey L., 2007. Regulation of Hsp70 Function: Hsp40 Co-chaperones and Nucleotide Exchange Factors." in *Cell Stress Proteins*, vols. 209–27. Springer, New York. [https://doi.org/10.1007/978-0-387-39717-7\\_9](https://doi.org/10.1007/978-0-387-39717-7_9).
- Young, Jason C., 2010. Mechanisms of the Hsp70 chaperone system. *Biochem. Cell. Biol.* 88, 291–300. 2. <https://doi.org/10.1139/O09-175>. National Research Council of Canada.
- Young, Jason C., Barral, José M., Hartl, F. Ulrich, 2003. *More than Folding: Localized Functions of Cytosolic Chaperones*. Elsevier Current Trends.
- Zetsche, Bernd, Gootenberg, Jonathan S., Abudayyeh, Omar O., Slaymaker, Ian M., Makarova, Kira S., Essletzbichler, Patrick, Volz, Sara E., et al., 2015. Cpf1 is a single RNA-guided endonuclease of a Class 2 CRISPR-Cas system. *Cell* 163 (3), 759–771. <https://doi.org/10.1016/j.cell.2015.09.038>.
- Zeytuni, Natalie, Zarivach, Raz, 2012. Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Modul. *Cell Press*. <https://doi.org/10.1016/j.str.2012.01.006>.
- Zeytuni, Natalie, Ozyamak, Ertan, Ben-Harush, Kfir, Davidov, Geula, Levin, Maxim, Gat, Yair, Moyal, Tal, Brik, Ashraf, Komeili, Arash, Zarivach, Raz, 2011. Self-recognition mechanism of mamA, a magnetosome-associated TPR-containing protein, promotes complex assembly. *Proc. Natl. Acad. Sci. U.S.A.* 108 (33), E480–E487. <https://doi.org/10.1073/pnas.1103367108>.
- Zhang, Zigu, Gurr, Sarah Jane, 2000. Walking into the unknown: a 'step down' PCR-based technique leading to the direct sequence analysis of flanking genomic DNA. *Gene* 253 (2), 145–150. [https://doi.org/10.1016/S0378-1119\(00\)00289-4](https://doi.org/10.1016/S0378-1119(00)00289-4).
- Zhang, R., Patena, W., Armbruster, U., Gang, S.S., Blum, S.R., Jonikas, M.C., 2014. High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell* 26 (4), 1398–1409. <https://doi.org/10.1105/tpc.114.124099>.
- Zhu, C.J., Lee, Y.K., 1997. Determination of biomass dry weight of marine microalgae. *J. Appl. Phycol.* 9 (2), 189–194. <https://doi.org/10.1023/A:1007914806640>.
- Zienkiewicz, Agnieszka, Zienkiewicz, Krzysztof, Poliner, Eric, Pulman, Jane A., Du, Zhi Yan, Stefano, Giovanni, Tsai, Chia Hong, et al., 2020. The microalga nanochloropsis during transition from quiescence to autotrophy in response to nitrogen availability. *Plant Physiol.* 182 (2), 819–839. <https://doi.org/10.1104/pp.19.00854>.