Identification of transcription factors involved in the phenotype of a domesticated oleaginous microalgae strain of *Tisochrysis lutea*

S. Thiriet-Rupert[a],[*],[1], G. Carrier[a], C. Trottier[a],[2], D. Eveillard[b], B. Schoefs[c], G. Bougaran[a], J.-P. Cadoret[a],[d], B. Chénais[c], B. Saint-Jean[a]

[a] *IFREMER, Physiology and Biotechnology of Algae Laboratories, Rue de l'Ile d'Yeu, 44311, Nantes, France*
[b] *LS2N, UMR6004, CNRS, Université de Nantes, Ecole Centrale de Nantes, IMTA, Nantes, France*
[c] *MIMMA, Metabolism, Engineering of Microalgal Biomolecules and Applications (MIMMA), Mer Molécules Santé, IUML, FR3473 CNRS, UBL, University of Le Mans, Le Mans, France*
[d] *Greensea SA, Meze, France*

ARTICLE INFO

ABSTRACT

Microalgae are an outstanding source of high value products with applications in food, feed or biofuel production. Among these promising organisms, the haptophyte *Tisochrysis lutea* is commonly used as a feed for shellfish and shrimps in aquaculture. Its capacity to produce high amounts of docosahexanoic acid and storage lipids is also of economic interest for nutrition and energy production. Consequently, understanding its lipid accumulation under nitrogen deprivation is of major interest.

Here, we aimed to identify Transcription Factors (TFs) involved in the establishment of the mutant phenotype of the 2Xc1 strain of *T. lutea*, which accumulates double the quantity of storage lipids under nitrogen deprivation than the wild type strain (WTc1). Strains were grown in chemostats and subjected to different nitrogen availability (limitation, repletion and depletion). Using RNA-seq data, the differentially expressed genes (DEGs) between strains were identified and summarized as a co-expression network to pinpoint putative major TFs in mutant phenotypes. This analysis was followed by a complementary Weighted Gene Correlation Network Analysis in order to classify genes based on their relative importance to mutant phenotype features, regardless of annotation biases due to the lack of functional annotation of the *Tisochrysis lutea* draft genome. This network-like strategy allowed the identification of seven TF candidates related to key functions in the mutant strain compared with WTc1. In particular, MYB-2R_14 and NF-YB_2 TFs are related to photosynthesis, oxidative stress response and triacylglycerol synthesis. GATA_2, MYB-rel_11 and MYB-2R_20 TFs are likely to be related to nitrogen uptake or carbon and nitrogen recycling, feeding carbohydrate synthesis in the form of chrysolaminarin. Finally, a q-RT-PCR approach further characterized the role of MYB-rel_11 and MYB-2R_20, revealing an expression pattern dependent on nitrogen availability.

1. Introduction

Although they account for < 1% of the photosynthetic biomass of our planet, the phytoplankton living in the photic zone of the ocean are responsible for around 45% of annual net primary productivity [1]. Among them, microalgae remain corner-stone species for aquatic ecosystem behaviours. Not only are they at the base of trophic chains, but are also key players in nitrogen, phosphorus or carbon biogeochemical cycles [2]. The number of microalgae species is thought to be as many as 70,000 [3] and evenly distributed in the eukaryotic tree [4]. This broad diversity is a source of a variety of high value compounds that, together with the capacity of microalgae to achieve a greater biomass production from light than terrestrial plants, makes microalgae of outstanding biotechnological interest [5–9] and has led to the design of many studies to enhance their production of high value products [10–13]. Among microalgae, haptophytes are distributed worldwide and participate greatly in global climate regulation as a carbon sink [14]. They produce alkenones, long-chain unsaturated methyl and ethyl

n-ketones (C37 to C40), which are exclusive to five haptophytes genera [15,16] and widely used as biomarkers for the reconstruction of marine paleoclimates [16]. The non-calcifying haptophyte *Tisochrysis lutea* is used in aquaculture as a feed for shellfish and shrimps because of its attractive fatty acid content [17]. Its production of docosahexanoic acid (DHA) and storage lipids is also of interest from nutritional and energy production perspectives [18–20]. In such a biotechnological context, studying microalgae mutant strains favourable for accumulation of industrial compounds is of primary interest. In particular, studying these strains can help us to better understand the mechanisms underlying interesting phenotypes. Because these key mechanisms are the consequence of genome expression modulation in which Transcription Factors (TFs) are critical players, we sought to identify TFs involved in a given phenotype, as well as their related genes, using network analysis.

This work focuses on the haptophyte microalga *T. lutea*, for which a mutant strain over-accumulating storage lipids (*T. lutea* 2Xc1) was selected by UVc random mutagenesis followed by flow cytometry selection for cell storage lipid content [21]. Previous proteomic and transcriptomic analyses revealed that wild type (WTc1) and mutant strains behave differently during early phases of nitrogen starvation, and suggested that proteins involved in carbon homeostasis, lipid metabolism and carbohydrate catabolism are likely involved in lipid accumulation [22,23]. A previous study investigated the metabolic strain specificities following fine-tuned changes of nitrogen availability, confirming the neutral lipid over-accumulation of 2Xc1, as well as an increase of its carbon accumulation capacity due to an increase in cell carbohydrate content [24] (Fig. 1).

To get further insight into the regulation of these metabolic changes, we analyzed the transcriptomic data generated during the experiment in [24]. Given the crucial role of TFs to monitor metabolism and their recent identification in *T. lutea* [25], we aimed to identify those TFs involved in the mutant phenotype of the 2Xc1 strain of *T. lutea*, using gene co-expression network construction and analysis. To characterize the putative role of these TF candidates in the establishment of the mutant phenotype, the functional annotation of their co-expressed genes was carried out. To avoid drawbacks due to sparse annotation, phenotypically relevant genes were selected, regardless of their annotation, for their links within network analysis to phenotypic features such as cell storage lipid or carbohydrate content dynamics over time. Such a selection of phenotypic parameters was intended to provide guidance to emphasize genes of interest but also the corresponding TFs and their putative roles in the mutant phenotype. Finally, for the sake of experimental validation, a RT-q-PCR approach was carried out to further characterize the expression profile of two pinpointed TF candidates.

2. Material and methods

2.1. Culture conditions and treatments

Following a previous study protocol [24], clones of *Tisochrysis lutea* CCAP 927/14 wild type (WTc1) and mutant CCAP 927/14 (2Xc1) strains were considered. The WTc1 and 2Xc1 clones were grown for 85 days in chemostat at a $0.5\,d^{-1}$ dilution rate in modified Walne's medium containing 125/125 μM N:P ratios in 10-L photobioreactors illuminated with continuous light ($150\,\mu mol \cdot m^{-2} \cdot s^{-1}$) and maintained at 27 °C and pH 7.3. The dilution rate was periodically checked by weighing the outgoing medium. Three nitrogen spikes were made at days 20, 43 and 83. Each spike consisted in the injection of 3.5 mmoles of $NaNO_3$ into the 10 L of culture.

Once the nitrogen-limited culture reached a steady state characterized by constant physiological parameters, the $NaNO_3$ injection was made [24]. Such nitrogen repletion conditions induced an increase of cell concentration (CC) and particulate carbon (PC). A N/C ratio increase indicates an uptake of microalgae nitrogen. Then, CC, PC and the N/C ratio maintained a high level and the lack of nitrogen available in the culture medium induced a decrease of this ratio, characteristic of nitrogen depleted conditions. Then, the culture reached a new steady state due to the dilution rate of the chemostat [24]. All these physiological parameters were taken from [24]. The lipid data were obtained using Nile red staining as advised in [26]. Total carbohydrate data were obtained following [27].
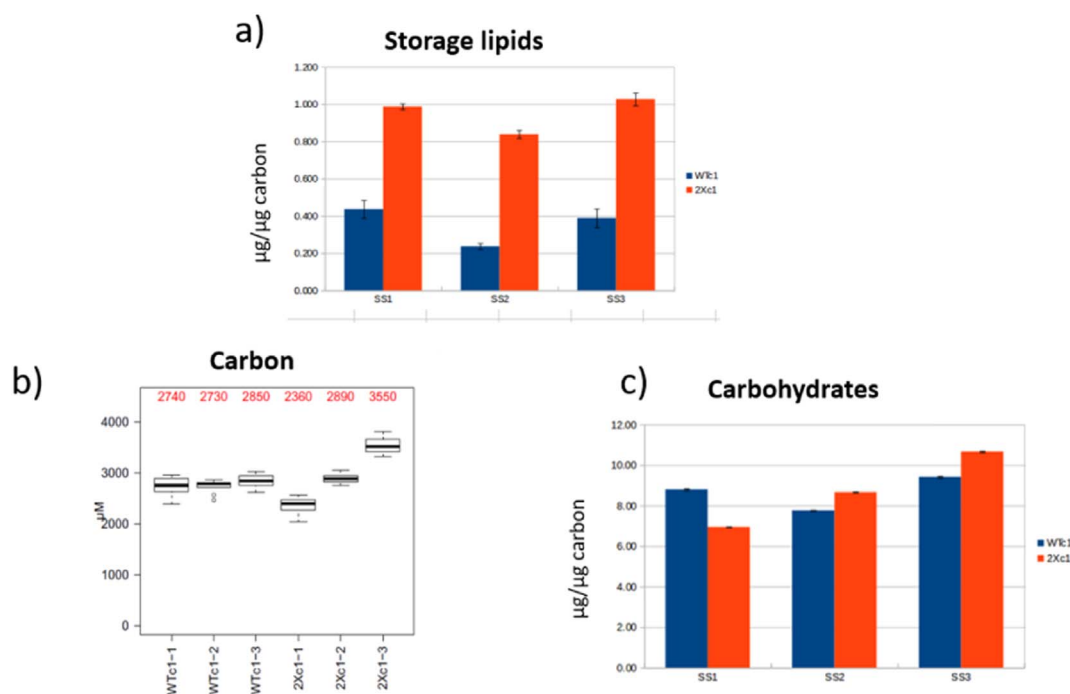

Fig. 1. The physiological state of the two strains was evaluated for the three consecutive steady states (SS1, SS2 and SS3 on each graph) during 85 days of culture [24]. The storage lipid over-accumulation of the 2Xc1 strain was confirmed (a). An increase of cellular carbon of the 2Xc1 strain after each nitrogen spike (b) was shown to be correlated to the cell carbohydrate content (c).

2.2. RNAseq library construction and data analysis

Our transcriptomic study is based on six biological samples from this experiment (Pt-27, Pt-83, Pt-90, Pt-117, Pt-125 and Pt-129) for each strain (Table S1) [24]. Pt-27 samples correspond to the steady state condition preceding the first nitrogen spike and Pt-117 to the third steady state. Pt-83 and Pt-125 samples correspond to the second and third nitrogen repletions, and samples Pt-90 and Pt-129 correspond to the second and third nitrogen depletions, respectively. These twelve samples were used for transcriptomic sequencing using Illumina technology.

Total RNA was extracted and sequencing was carried out as in [22]. For each sample, raw reads data were filtered using Cutadapt (version 1.0) [28] to remove known Illumina adapter sequences. Reads quality filter (version 1.0.0) was then used to exclude low quality reads, with a quality score threshold of 30 and a minimal length of 75 bases. Read quality was assessed using FastQC software [29]. The 352,199,068 cleaned paired reads were mapped on the reference genome of *T. lutea* (raw reads at SRA, RUN: SRR3156597) using Tophat2 (version 0.5) [30] and aligned reads for each gene were counted using htseq-count [31] (version 0.3.1) in union mode. Gene expression level was then calculated in Reads Per Kilobase per Million mapped reads (RPKM). A gene with an RPKM value > 1 in at least one of the twelve samples was considered as expressed. Among the 20,582 identified genes in the *T. lutea* genome, 15,333 were expressed in this study. Genes were annotated using BLAST against the Swissprot database and functional domains were identified using InterProScan as implemented in BLAST2GO [32]. Transcription factors were then annotated using an optimized pipeline [25].

2.3. Weighted Gene Correlation analysis

Based on the 15,333 expressed genes, an unsigned gene co-expression network was built via the WGCNA R package [33]. A soft threshold power β of 16 was used to meet the scale-free topology criterion for optimal clustering. The resulting similarity matrix was turned into an adjacency matrix, which was transformed into a topological overlap matrix (TOM). This TOM-based dissimilarity measure was coupled to average linkage hierarchical clustering using the Dynamic Tree Cut algorithm, resulting in gene module identification. To form more coherent modules, similar modules were merged on the basis of module eigengenes (i.e., the first principal component of a given module representing the expression profile of a co-expression module) using a cut height parameter of 0.25. To identify modules significantly associated with the physiological data, module eigengenes were correlated to lipid and carbohydrate data. The identification of hub genes for each module was carried out using two scores: eigengene-based intramodular connectivity (kME) value emphasizes genes of importance to the module definition, whereas the Gene Significance (GS) pinpoints genes for which expression profiles are correlated with a physiological trait. For the sake of gene classification, a gene was considered as relevant when its GS value was higher than or equal to 0.75 for (i) cell storage lipid or (ii) carbohydrate content dynamic, or its kME value was higher or equal to 0.95, respectively.

2.4. Identification of differentially expressed genes and gene co-expression networks

The differentially expressed genes (DEGs) between two strains were identified between samples corresponding to a similar sampling time (for example: the Pt-27 sample for the WTc1 strain vs Pt-27 sample for the 2Xc1 strain). The differential expression was estimated using Gfold V1.1.2, based on the GFOLD value, which was described as more biologically meaningful [34]. A gene with a |GFOLD value| > 2 was considered as differentially expressed.

A gene co-expression network (GCN) was then built for each strain.

One based on RPKM values of the 527 DEGs in the six WTc1 samples, and a second based on RPKM values of the 527 DEGs in the six 2Xc1 samples. GCNs were built using extended local similarity analysis (eLSA) [35,36]. These directed networks are based on pairwise correlation of a TF expression profile and its co-expressed genes expression profile. Such a link was considered as significant for a Spearman's Correlation Coefficient value higher than 0.8 and a *p*-value lower than 0.05. The resulting GCNs were visualized and analyzed using Gephi [37].

2.5. RNA extraction and reverse transcription

The samples were centrifuged (20 min, 5000g, 4 °C). The supernatant was discarded and the seawater was taken up to remove the salt. The pellets were resuspended in Trizol reagent (Invitrogen, Carlsbad, CA, USA) and chloroform. After centrifugation, the upper phases were collected and 0.5 volume of absolute EtOH was added. The samples were transferred to a column of the RNeasy Plant Mini Kit (Qiagen, Helden, Germany) and the manufacturer's instructions were followed thereafter. A DNase treatment (RQ1 DNase, Promega, Madison, WI, USA) was applied and total RNA was purified using the RNeasy Plant Mini Kit with the RLT buffer and EtOH. The quality and concentration of RNA were determined with a spectrophotometer (ND-1000; NanoDrop Technologies, Wilmington, DE, USA) at 260 and 280 nm wavelengths. The PCR amplification of an RNA sample was used to check for genomic DNA contamination. Total RNA was stored at −80 °C. Reverse transcription of RNA was performed using the High Capacity cDNA Reverse transcription kit (Applied Technologies, Foster, CA, USA) following the manufacturer's instructions.

The q-RT-PCR experiment was performed with Fluidigm Biomark technology. The geometric mean of the normalization using two reference genes, EF1 (translation elongation factor 1 alpha) and GAPDH (Glyceraldehyde-3-Phosphate Dehydrogenase), was then used to calculate differential expression by the $2^{-\Delta\Delta Ct}$ formula. For *MYB-2R_20*, *MYB-rel_11*, *PLAAOx*, *CSAP* and *Nrt*2.1 genes, the differential expression was assessed along the dynamic of a nitrogen spike using the previous steady state as a reference condition. Eight genes representing a broad expression range were selected to evaluate the quality of RNA-seq data. We observed an overall correlation *r* of 0.9 (*p*-value = $1.47\,e^{-9}$) between RNA-seq and q-RT-PCR experiments, indicating that the RNA-seq results were representative of the transcriptomic profiles (Fig. S1). The list of the specific primers used for the q-RT-PCR experiments are listed in Table S2

3. Results

3.1. Differentially expressed genes between WT and mutant strains include seven transcription factors

To identify genes related to the mutant phenotype, DEGs between WTc1 and the mutant strain were identified for each of the six experimental conditions of the culture dynamic. From these samples, 527 genes were differentially expressed (Table S3) between the two strains in at least one set of experimental conditions sampled using a Gfold fold change value higher than 2 for up-regulation (249 up-regulated genes) and lower than −2 for down-regulation (278 down-regulated genes). Among these gene regulatory specificities of the 2Xc1 strain, seven were TFs. One TF belonging to the MYB-2R (*MYB-2R_14*) family and a second TF belonging to the C2C2-GATA (*GATA_2*) family were not expressed in the WTc1 strain but only in the mutant strain. Given this strain-specific expression, these two TFs are good candidates for understanding the mutant phenotype. Moreover, five TFs belonging to the MYB-2R (*MYB-2R_20*), MYB-related (*MYB-rel_11*), NF-YB (*NF-YB_2*), Fungal-TRF (*Fungal-TRF_8*) and HB-other (*HB-other_9_PAS*) families were down-regulated in the mutant strain.
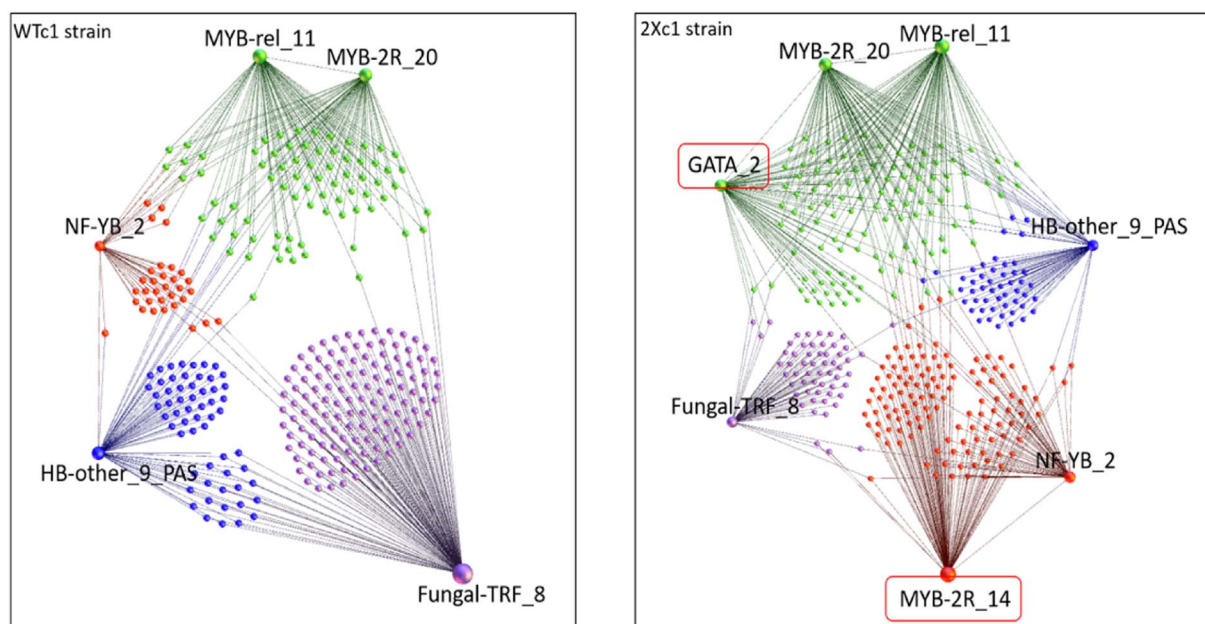
Fig. 2. Representation of the differentially expressed gene regulatory network of *T. lutea* WTc1 and 2Xc1 strains using Gephi. In these directed networks, each node is a gene, with the TFs labeled with their name. Each edge links a TF to its co-expressed genes based on co-expression. The node size for each TF is set according to their modularity (the number of co-expressed genes). The two TFs framed in red in the 2Xc1 GCN are only expressed in this strain. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2. Gene co-expression networks (GCNs) underlying the response of each strain to nitrogen deprivation

Seven TFs (*MYB-2R_14*, *GATA_2*, *MYB-2R_20*, *MYB-rel_11*, *NF-YB_2*, *Fungal-TRF_8* and *HB-other_9_PAS*) were differentially expressed in the 2Xc1 strain. They are used in a strain-specific manner to cope with the variation of nitrogen availability, which induced the mutant phenotype. This suggests an important role of these seven TFs in the establishment of the mutant phenotype under nitrogen deprivation. To identify the link between these TFs and the mutant phenotype, a GCN was constructed for each strain using eLSA software [35,36] that uses co-expression to link TFs to the others DEGs (Fig. 2). The GCNs were based on the RPKM values of the DEGs between the strains.

The WTc1 GCN is a directed graph with 298 nodes for five TFs and 293 genes linked by 412 edges (density: 0.005). Following the modularity detection implemented in Gephi, the graph was clustered into four gene clusters: green, red, purple and blue in Fig. 2. Each cluster depicts denser subgraphs. Clusters comprising only one TF (*Fungal-TRF_8*, *HB-other_9_PAS* and *NF-YB_2*) belong to the purple, blue and red clusters, respectively, whereas the green cluster includes both (i.e., *MYB-rel_11* and *MYB-2R_20*).

The 2Xc1 GCN is a directed graph with 349 nodes including seven TFs and 577 edges (density: 0.005). Again, the four clusters are emphasized and show a similar composition of TFs. The purple cluster remains composed of *Fungal-TRF_8*, whereas the blue one is composed of *HB-other_9_PAS*. The red cluster that includes *NF-YB_2* TF and its co-expressed genes in the WTc1 GCN is extended to the *MYB-2R_14* TF in the 2Xc1 GCN. Likewise, the green cluster is extended to *GATA_2*.

The fact that these genes are co-expressed together suggests that they are co-regulated. Consequently, in order to evaluate the relevance of this clustering, the promoting sequences (1000 bp upstream and 100 bp downstream of the START codon) of each co-expressed genes were retrieved. Then, the AME tool form the MEME suite (Fabian et al., 2010) was used to identify enrichment of known TF binding sites (TFBS) in the promoting sequences of co-expressed genes. If the clustering was relevant, the same TFBS should be enriched in the promoting sequences of co-expressed genes. Moreover, TFs are included in each cluster. Since these are potential regulators of the genes co-expressed

with them, more attention was paid to TFBS bond by TFs belonging to the same family than those co-expressed with the analyzed genes.

For the genes co-expressed with the Fungal-TRF_8 TF, the TFBS contained in the JASPAR fungi database [38] were used because this TF family is well characterized in fungi. Three motifs bound by TF from this family were found (RSC30 (uniprot: P38781), RSC3 (uniprot: Q06639) and CHA4 (uniprot: P43634) TFs from *Saccharomyces cerevisiae*). An alignment of the amino acid sequence of the DNA binding domains (DBDs) of these four TFs revealed that Fungal-TRF_8 is closer to CHA4 (48.15% identity) than to RSC3 (32.14% identity) or RSC30 (30% identity). CHA4 was shown to be involved in serine catabolism allowing the use of serine and threonine as nitrogen source in *S. cerevisiae* [39,40].

For the other TFs, the *Arabidopsis thaliana* dataset from O'Malley et al. [41] was used to screen the promoting sequences for TFBS enrichment. For the genes co-expressed with the MYB-2R_14 TF, which seems to be linked to cell lipid storage content, two motifs bound by MYB TFs were found (MYB55 (At4g01680) and MYB30 (At3g28910)). These two motifs were also found in the promoting sequences of the genes co-expressed with the NF-YB_2 TF. This is not surprising since 79% of the genes co-expressed with NF-YB_2 are shared with MYB-2R_14. The alignment of the DBDs of these MYB-2R_14, MYB55 and MYB30 showed that the former is closer to MYB30 (36.14% identity) than to MYB55 (32.53% identity). Moreover, MYB30 is known to be involved in very long chain fatty acid synthesis in *A. thaliana* [42].

Concerning the green cluster, which seem to be linked to cell carbohydrate content, one motif bound by a MYB-rel TF (At1g18960) was found in the promoting sequences of the genes co-expressed with MYB-rel_11 as well as with MYB-2R_20 and GATA_2 TF. This is not surprising since these three TFs shared at least the half of their co-expressed genes in the mutant strain (GATA_2 being only expressed in this strain). Likewise, one motif bound by a GATA TF (GATA1: At3g24050) and four motifs bound by MYB TFs (MYB83: At3g08500, MYB94: At3g47600 and MYB15: At3g23250) were found in the promoting sequences of the genes co-expressed with MYB-rel_11, MYB-2R_20 and GATA_2 TFs. The alignment of GATA_2 DBD to that of GATA1 showed 42.86% identity and the DBD of MYB-rel_11 was 41.46% identical to that of AT1G18960. Concerning MYB-2R_20 TF, the closest TF was MYB83

(45% identity) and, interestingly, this TF is involved in polysaccharide synthesis in *A. thaliana* [43].

3.3. WGCNA-based gene prioritization overcomes the lack of functional annotation

The purpose was then to evaluate the function of the TFs and their putative role in the mutant phenotype. An efficient strategy to assign a putative function to an unknown gene consists in analyzing annotations of the set of genes co-expressed with the unknown. Such a strategy was applied herein with a focus on genes linked to each TF in the GCNs (Fig. 2). However, due to sparse annotation, only 50% of *T. lutea* genes have been associated to putative functions, which significantly limits standard assignation procedures. The present study overcomes this weakness by prioritizing genes involved in the *T. lutea* 2Xc1 phenotype. First, genes related to lipid and carbohydrate content regardless their annotation were identified. For this purpose, a weighted gene co-expression network was built based on the twelve RNAseq samples using the WGCNA R package [33]. Based on the WGCNA data, the genes for which the expression profile was correlated to the variation of lipid or carbohydrate content were identified. The genes were then prioritized (or ranked) following this correlation coefficient. These genes are likely to be related to lipid and carbohydrate content because their expression follows the variation of these physiological parameters.

Moreover, in co-expression modules such as those identified by WGCNA, the more highly connected genes (called "hub genes") were shown to play a critical role in the cellular function of the whole module [44–46] making of these genes very interesting candidates. Consequently, the hub genes of the modules found to be significantly correlated to lipid and carbohydrate content were thought to be functionally related to these two parameters. Therefore, these hub genes were added to the gene prioritization. Following the WGCNA standard procedure, the 15,333 expressed genes were clustered into 20 modules of co-expressed genes (Fig. 3a).

The first principal component of each identified module, called the eigengene, describes the eigen expression profile of a given module. This signal is then correlated with physiological traits of interest (i.e., cell carbohydrate and storage lipid content dynamics) to decipher three modules of physiological interest (Fig. 3b). A first module, called the Lipid Related (LR) module, correlates with cell storage lipid content ($r = 0.72$, *p*-value = 8 e^{-3}), Two other modules, called Carbohydrate Related (CR1 and CR2), correlate with cell carbohydrate content ($r = 0.78$, *p*-value = 3 e^{-3} and $r = 0.86$, *p*-value = 4 e^{-4}, respectively).

Three TFs (*LIM_1*, *MYB-3R* and *MYB-2R_14*) belong to the LR module and therefore constitute interesting candidate regulators of the cell storage lipid content in *T. lutea*, of which *MYB-2R_14* TF is particularly notable as it is only expressed in the mutant strain.

In addition, a complimentary Gene Ontology (GO) enrichment was performed to identify functions significantly more represented in genes modules (*p*-value < 0.05). Eighteen GO terms were enriched in the LR, 20 in the CR1 and 24 in the CR2 module (see Table S4 for details). However, such a result should be moderated by the fact that among the gene contents of each module (i.e., 445, 115, 262 genes for the LR, CR1, and CR2 modules, respectively) on average only 39% have a putative associated function (37%, 42%, 41% for LR, CR1, CR2, respectively), due to standard mis-annotations of non-model species. It is noteworthy that these GO results also suggest diverse functions that are not only directly linked to storage lipid and carbohydrate contents, thus re-inforcing the interest of a network analysis that is, by essence, less sensitive to annotation flaws. In particular, WGCNA identifies 284 genes linked to storage lipid and carbohydrate content dynamics regardless of their annotation. Among these genes, 33% show putative functions (see Table S5 for details). For this subset of genes, the GO analysis showed an enrichment for carbohydrate metabolism (*p*-value: 2.46 e^{-2}), chlorophyll metabolism (*p*-value: 1.24 e^{-2}), ubiquitin-dependent protein degradation (*p*-value: 3.02 e^{-2}), as well as tRNA (*p*-value: 1.24 e^{-2}) and mRNA processing (*p*-value: 2.46 e^{-2}).
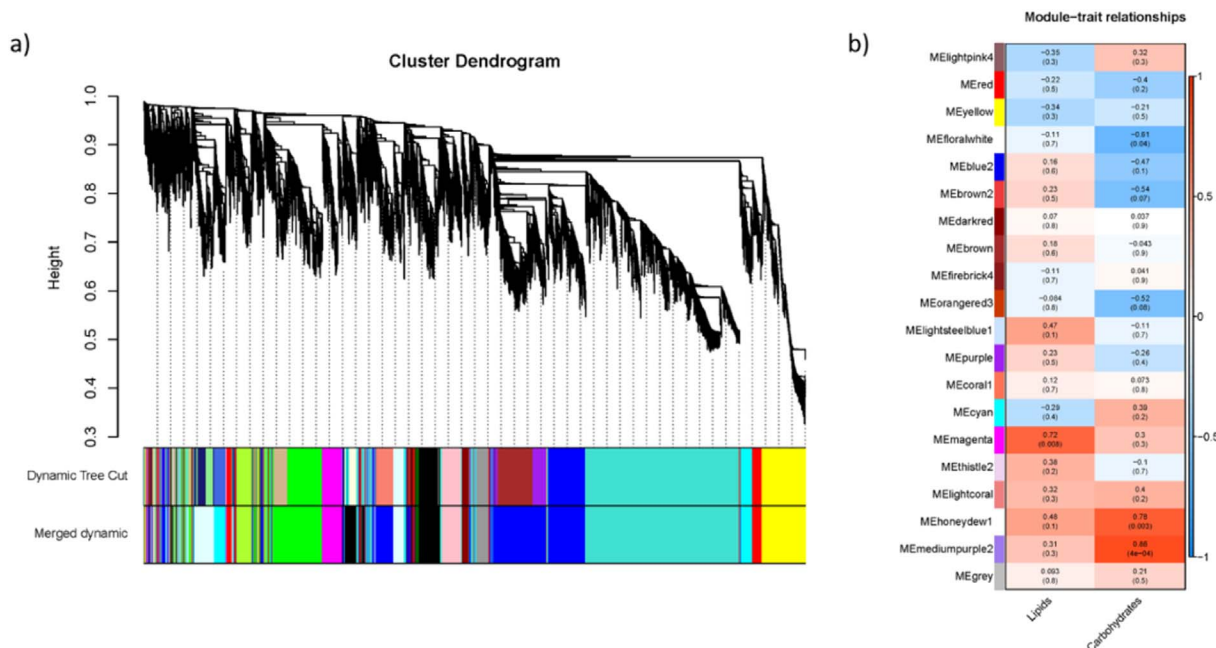

Fig. 3. RNA-seq samples and WGCNA analysis. A) Representation of the hierarchical clustering tree leading to the identification of the 20 modules of co-expressed genes. The leaves of the tree correspond to the 15,333 expressed genes included in the analysis, and the height reflects the closeness of individual genes. The lower panel represents the colors assigned to each module by the Dynamic Tree Cut method ('Grey' is for unassigned genes). The changes in module assignment consecutive to the Merged Dynamic method using a stringency threshold of 0.75 are also shown. B) Module-lipids/carbohydrates correlations with the corresponding *p*-values. Each row corresponds to a co-expression module. Each cell contains the correlation coefficient between the expression profile of a module and the variation of cell storage lipid (first column) or carbohydrate content (second column). The number between parentheses is the associated *p*-value. The table is color-coded according to the color scale on the right (from blue for − 1 to red for 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
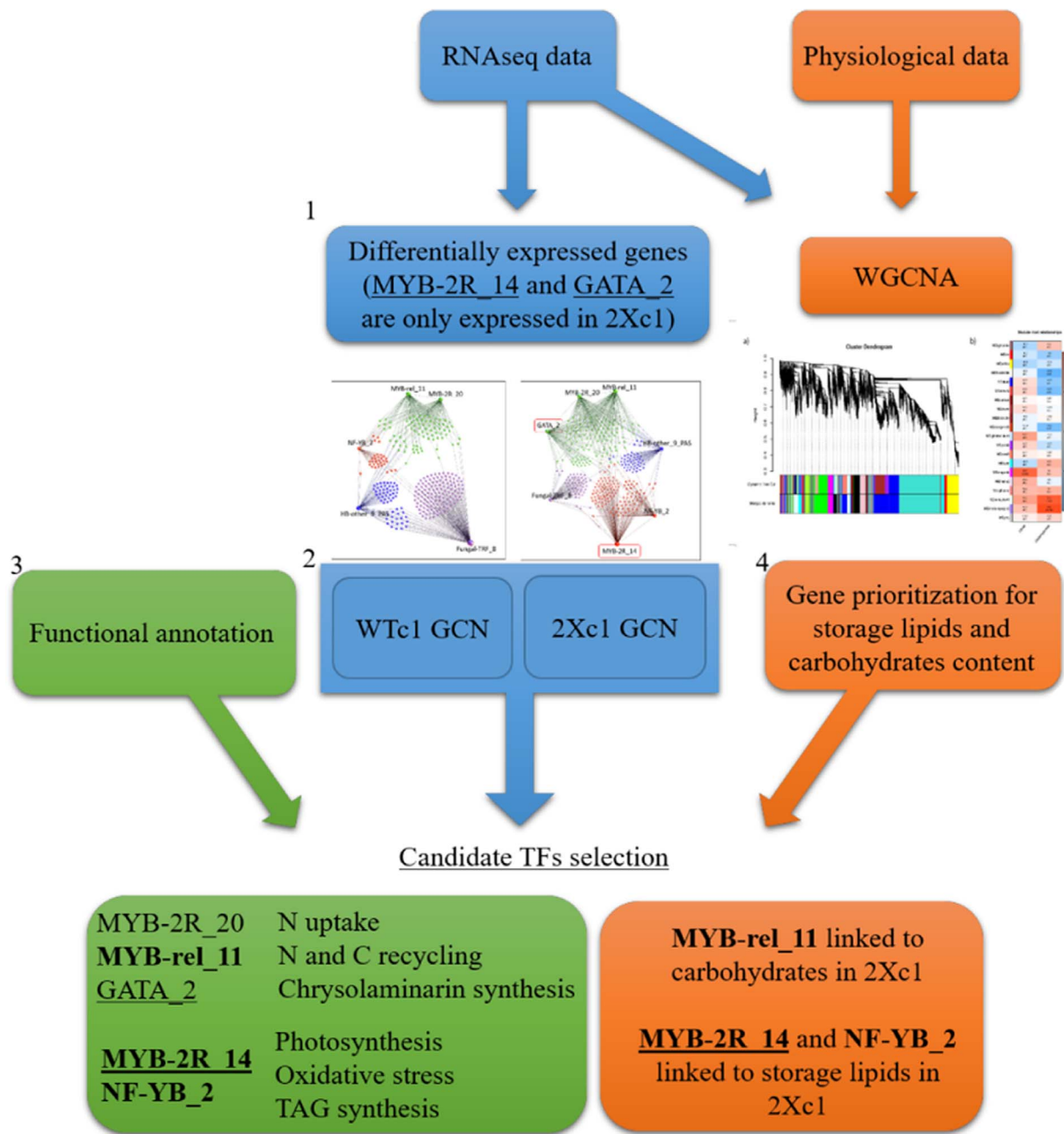
Fig. 4. Workflow representing the strategy used to identify the candidate TFs. First, (1) the differentially expressed genes (DEGs) between the two strains were identified and highlighted seven TFs, two of which were only expressed in the mutant strain. (2) These DEGs were used to build the gene co-expression network (GCN) representing the transcriptional response of each strain to nitrogen deprivation. (3) To pinpoint the putative role of the TFs in the mutant phenotype, the functional annotation of their co-expressed genes was carried out in both networks. Then, the comparison of the two GCNs highlighted some functions present in the genes co-expressed with a TF in the 2Xc1 GCN but absent from the genes co-expressed with the same TF in the WTc1 GCN. Due to standard mis-annotation inherent to non-model organisms, the preceding functional annotation was completed by (4) a gene prioritization based on weighted gene correlation network analysis (WGCNA). Identifying genes related to mutant phenotype features regardless of their functional annotation, this gene prioritization is a good complementary strategy for candidate TF selection. The enrichment of prioritized genes in the genes co-expressed with each TF was then calculated and the two GCNs compared to pinpoint some candidate TFs.

TFs with 2Xc1-specific expression are underlined in the figure. TFs pinpointed by the gene prioritization enrichment analysis are shown in bold.

3.4. Co-expression network analysis of differentially expressed genes

To identify major TFs involved in the mutant phenotype, the putative functions represented in each cluster (Tables S6 and S7) were compared between the WTc1 GCN and the 2Xc1 GCN. The proportion of prioritized genes among the genes co-expressed with a TF in the WTc1 GCN (exact test of Fisher) was compared to the proportion of prioritized genes among the genes co-expressed with the same TF in the mutant GCN. This enrichment of prioritized genes completed the functional annotation and further linked the candidate TFs to the

mutant phenotype (Fig. 4). Among the DEGs used for the GCN construction, a total of 35 genes were selected following the gene prioritization: 14 related to cell storage lipid content and 21 related to cell carbohydrate content (Table 1).

The purple cluster, driven by the *Fungal-TRF_8* TF, shows similar functions for both strains, whereas the blue, red and green clusters show functions specific to the 2Xc1 strain: the blue cluster, driven by the *HB-other_9_PAS* TF, emphasizes gene functions related to (i) transport with a magnesium ion transporter (p-value = 2.99 e^{-2}) and vacuolar protein sorting-associated protein 35 (p-value = 4.66 e^{-3})

Table 1
List of prioritized genes. Genes selected among the DEG based on either their Gene Significance (GS) for carbohydrates or lipids, or on the KME value for hub genes.

Gene name	GS for carbohydrates	GS for lipids	KME	Module	Putative function
Tiso_gene_3094	0.77		0.96	CR1 module	Nd
Tiso_gene_5367	0.70		0.96	CR1 module	Nd
Tiso_gene_5511	0.74		0.95	CR1 module	Nd
Tiso_gene_5871	0.86		0.95	CR2 module	Nd
Tiso_gene_14258	0.84		0.95	CR2 module	Nd
Tiso_gene_11673	0.89		0.95	CR2 module	Nd
Tiso_gene_9025		0.76	0.97	LR module	Acetyl-CoA synthetase-like
Tiso_gene_20182		0.70	0.96	LR module	Nd
Tiso_gene_18537		0.73	0.96	LR module	Nd
Tiso_gene_8095		0.67	0.95	LR module	MYB-2R_14
Tiso_gene_19516		0.61	0.95	LR module	Nd
Tiso_gene_19445		0.68	0.95	LR module	Nd
Tiso_gene_6707		0.70	0.95	LR module	Nd
Tiso_gene_18899		0.71	0.95	LR module	Tyrosine-protein kinase ephrin type A/B receptor-like
Tiso_gene_11571	0.92			CR2 module	Regulator of chromosome condensation RCC1
Tiso_gene_13889	0.86			CR2 module	Nd
Tiso_gene_6544	0.84			CR2 module	Nd
Tiso_gene_6760	0.84			LR module	Nd
Tiso_gene_13518	0.83			CR2 module	ATP-binding cassette transporter
Tiso_gene_7210	0.82			CR2 module	ATP-binding cassette transporter
Tiso_gene_9366	0.80			CR2 module	Saccharopine dehydrogenase
Tiso_gene_17273	0.80			CR2 module	Nd
Tiso_gene_20489	0.80				Nd
Tiso_gene_4808	0.78			CR2 module	Calcium-independent phospholipase A2-gamma
Tiso_gene_16415	0.77			CR2 module	Nd
Tiso_gene_2919	0.77			CR2 module	Neurotransmitter-gated ion-channel ligand-binding domain containing protein
Tiso_gene_20165	0.76			CR2 module	Nd
Tiso_gene_3564	0.76			CR2 module	Nd
Tiso_gene_11856	0.75			CR2 module	Nd
Tiso_gene_4035		0.91		CR2 module	Nd
Tiso_gene_19975		0.79		LR module	Nd
Tiso_gene_9964		0.77		CR1 module	Nd
Tiso_gene_13835		0.76		LR module	Nd
Tiso_gene_14115		0.76		LR module	Nd
Tiso_gene_15057		0.76		LR module	ADP-ribosylation factor
Tiso_gene_20336		0.75		LR module	Nd

involved in the endosome-to-Golgi retrieval cargo transport pathway; (ii) protein glycosylation (p-value = $1.39\,e^{-2}$) with the UDP-glucose: glycoprotein glucosyltransferase-like protein that glycosylates unfolded glycoproteins, providing a quality control for protein transport out of the ER; (iii) signaling with phosphatidylinositol 4-kinase (p-value = $3.44\,e^{-2}$) involved in cellular signaling as well as in trafficking in the Golgi apparatus and trans-Golgi network [47–49]; and (iv) calcium dependent protein degradation with a peptidase C2, calpain family protein (p-value = $4.34\,e^{-2}$). Moreover, a diacylglycerol O-acyltransferase (p-value = $2.33\,e^{-3}$), involved in triacylglycerol (TAG) synthesis in nitrogen deprivation conditions [50], as well as a mannose-6-phosphate isomerase (p-value = $2.33\,e^{-3}$) and an enolase (p-value = $1.62\,e^{-2}$) involved in carbohydrate synthesis are among the genes co-expressed with the *HB-other_9_PAS* TF. All these highlighted genes were positively correlated to *HB-other_9_PAS* TF.

In the red cluster, driven by the *MYB-2R_14* and *NF-YB_2* TFs, functions related to transport were enriched in the 2Xc1 GCN, with a gated ion-channel ligand binding domain containing protein (p-value = $3.54\,e^{-2}$) and a high affinity iron ion transporter (p-value = $1.02\,e^{-2}$), both negatively correlated to *MYB-2R_14* TF. DNA repair was enriched, with two poly(ADP-ribose) polymerases positively correlated to both *MYB-2R_14* and *NF-YB_2* TFs (p-value = $1.93\,e^{-3}$), a DNA photolyase protein and a cryptochrome DASH family protein positively correlated to *MYB-2R_14* TF (p-value = $7.05\,e^{-4}$) and involved in DNA repair with a specific action on pyrimidine dimers [51]. Photosynthesis was also enriched in the red cluster (p-value = $2.87\,e^{-4}$), with four genes involved in light harvesting positively correlated to *MYB-2R_14* TF. Very interestingly, the *MYB-2R_14* TF, which was only expressed in the 2Xc1 strain, is a hub gene (kME = 0.95) in the LR module and has an expression profile correlated with cell storage lipid

content dynamics. Moreover, storage lipid-related prioritized genes were enriched among the genes co-expressed with both *MYB-2R_14* and *NF-YB_2* TFs in the 2Xc1 GCN (Fisher's exact test, p-value < 0.05), but not in the WTc1 GCN (Table S8). More broadly, in this red cluster, 43% of *MYB-2R_14* co-expressed genes belonged to the LR module. Among the gene co-expressed with *NF-YB_2* TF, 15% belonged to the LR module in the WTc1 GCN, but this portion rose to 48% in the 2Xc1 GCN.

In the green cluster, driven by *MYB-2R_20*, *MYB-rel_11* and *GATA_2* TFs, the functions enriched in the 2Xc1 GCN were more diverse than those enriched in the blue and red clusters. However, functions related to carbohydrates were enriched, with an UDP-glucose pyrophosphorylase involved in carbohydrate synthesis and positively correlated to both MYB TFs. Three functions involved in nitrogen utilization under nitrogen deprivation conditions were among the enriched functions: nitrogen uptake with an ammonium transporter (positively correlated to both MYB TFs), and cellular nitrogen remobilization with genes involved in protein degradation (three positively correlated to both MYB TFs, one of which also negatively correlated to *GATA_2* TF) and a urea transporter (positively correlated to both MYB TFs). Moreover, in the 2Xc1 GCN, carbohydrate-related prioritized genes were enriched among the genes co-expressed with *MYB-rel_11* (p-value = $2\,e^{-2}$), although the p-value was slightly higher for *MYB-2R_20* (p-value = $9\,e^{-2}$) (Table S8). No enrichment was found in the WTc1 GCN.

Interestingly, the functions identified in the blue, red and green clusters did not overlap in the WTc1 strain (Fig. 5a), whereas, in the mutant strain, functions identified in the green cluster partly overlapped with those of the blue and red clusters (Fig. 5b). This functional overlap suggests a combinatorial role of the TFs driving these three clusters (*HB-other_9_PAS*, *MYB-2R_14*, *NF-YB_2*, *MYB-2R_20*, *MYB-rel_11*
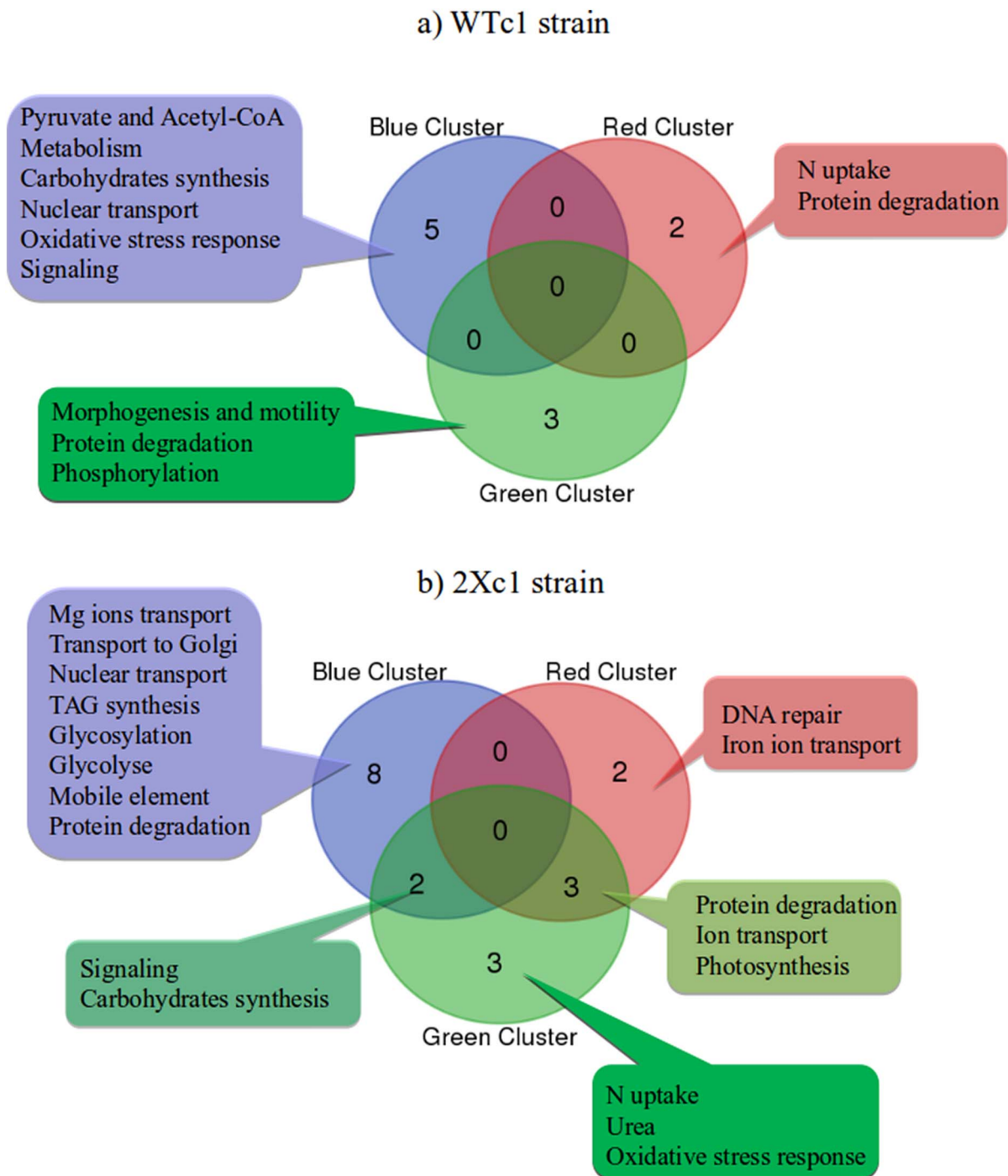
a) WTc1 strain

b) 2Xc1 strain

Fig. 5. Venn diagrams showing the functions common to the blue, red and green clusters in WTc1 (a) and 2Xc1 (b) strains. No overlapping functions were identified in the WTc1 strain (a). In the mutant strain, overlapping functions were identified between blue and green clusters as well as between red and green clusters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and *GATA_2*) in the mutant phenotype.

For microalgae grown under nitrogen deprivation, the high-affinity nitrate/nitrite transporters (*Nrt*2) are also known to be involved in the enhancement of nitrogen uptake efficiency [52–54]. Four *Nrt*2s were previously identified in *T. lutea* [55], but their nucleotide sequences are too similar to be differentiated during transcriptomic analysis, resulting in a biased expression quantification. In order to further characterize the involvement of *MYB-2R_20* and *MYB-rel_11* TFs in key mechanisms such as nitrogen uptake and recycling, a real time PCR approach was used.

3.5. MYB-2R_20 and MYB-rel_11 TF gene expression monitoring during the dynamics of a nitrogen spike

To further characterize the role of *MYB-2R_20* and *MYB-rel_11* TFs in *T. lutea* strain 2Xc1, their gene expression was monitored together with the four *Nrt*2 transporter coding genes identified in *T. lutea*. Moreover, two interesting genes were added to this analysis: Periplasmic ʟ-Amino-Acid Oxidase (*PLAAOx*) and Coccolith Scale Associated Protein (*CSAP*). Both these genes encode proteins previously identified to be up-accumulated during nitrogen limitation in the WT strain [23]. The expression of these eight genes was monitored in the 17
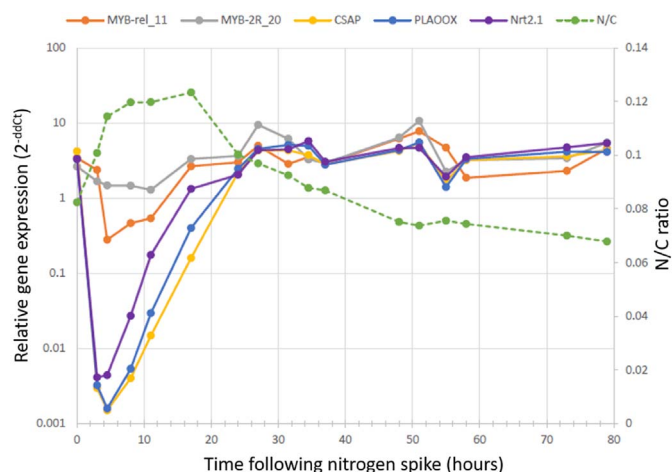
Fig. 6. Gene expression monitoring by q-RT-PCR of CSAP, PLAAOx, Nrt2.1, MYB-rel_11 and MYB-2R_20 TFs in the 2Xc1 strain during the second nitrogen spike. The nitrogen injection was carried out 30 min after the first sample represented in this graph. Gene expression was normalized using two housekeeping genes. Relative gene expression was assessed using the steady state preceding the spike as a reference condition.

Table 2

q-RT-PCR expression profile correlations. Co-expression evaluation of MYB-rel_11, MYB-2R_20, CSAP, PLAAOx and NRT2.1 genes in 2Xc1 (a) and WTc1 (b) strains. Pearson correlation coefficients (PCC) are shown with the corresponding *p*-values.

	MYB-rel_11	MYB-2R_20	CSAP	PLAOOX
a) 2Xc1 strain				
MYB-rel_11				
MYB-2R_20	0.83			
	(0.000024)			
CSAP	0.76	0.84		
	(0.000877)	(0.000058)		
PLAOOX	0.73	0.82	0.97	
	(0.001455)	(0.000135)	(0.000007)	
NRT2.1	0.74	0.79	0.92	0.93
	(0.001342)	(0.000205)	(0.000001)	(0.000002)
b) WTc1 strain				
MYB-rel_11				
MYB-2R_20	0.64			
	(0.005070)			
CSAP	0.37	0.70		
	(0.127717)	(0.001566)		
PLAOOX	0.48	0.75	0.95	
	(0.048095)	(0.000507)	(0.000001)	
NRT2.1	0.48	0.71	0.94	0.97
	(0.048546)	(0.001325)	(0.000001)	(0.000008)

samples taken for each strain during the nitrogen spike (Fig. 6).

A cross-correlation was performed to identify co-expression relationships. In both strains, the three genes *PLAAOx*, *CSAP* and *Nrt*2.1 are highly co-expressed (Table 2). In the WTc1 strain, only the *MYB-2R_20* expression profile was correlated with that of *PLAAOx*, *CSAP* and *Nrt*2.1 genes (Table 2). However, in the 2Xc1 strain, the co-expression between *MYB-2R_20* and, respectively, *PLAAOx*, *CSAP* and *Nrt*2.1 was strengthened (higher coefficient correlations and lower *p*-values, see Table 2). Interestingly, *MYB-rel_11* TF was only co-expressed with *PLAAOx*, *CSAP*, *Nrt*2.1 and *MYB-2R_20* in the mutant strain (Table 2). Moreover, *PLAAOx*, *CSAP* and *Nrt*2.1 genes were down-regulated throughout the whole spike (Table S9).

4. Discussion

In this study, we identified TFs potentially involved in the establishment of the mutant phenotype of the 2Xc1 strain of *T. lutea*. Using RNA-seq data representing changes in nitrogen availability, DEGs

between the two strains were identified and highlighted seven TFs. From these DEGs, a Gene Co-expression Network was built for each strain. Each network is a synthesis of the transcription response of a given strain to nitrogen availability variations. In order to assign a role to each TF in the establishment of the mutant phenotype, a gene prioritization was set up to complete the GO annotation. This ranking is based on WGCNA with the purpose of identifying genes linked to the features of the mutant phenotype (i.e., storage lipids and carbohydrates or content dynamics over time), regardless of their putative function. This complementary strategy allowed us to identify functions specific to the 2Xc1 strain.

4.1. TF candidates for establishing the mutant phenotype

The TF families for which members were differentially expressed in the mutant strain (MYB, NF-Y, GATA, HB and Fungal TRF) were described to be involved in nitrogen deprivation-related processes. Members of MYB-related, HB and GATA families were found to be regulatory hub genes to nitrogen deprivation in the green alga *Chlamydomonas reinhardtii* [56]. More precisely, two MYB-related TFs were described to be related to photosynthesis, carbohydrate metabolism, amino acids and lipid metabolism through TAG synthesis. A GATA TF was linked to amino acids and central metabolism [56]. An HB TF was linked to protection against photo-oxidative stress, carbohydrate metabolism, citrate and glyoxylate cycles, and lipid metabolism [56]. Moreover, in the same study, members of the NF-YB family were found to be activated in response to nitrogen deprivation [56]. In the red alga *Cyanidioschyzon merolae*, an MYB-2R TF was shown to be involved in the regulation of N assimilation under N-depleted conditions [57]. A few studies have identified the Fungal TRF family in algae [25,58]. This family was first identified in fungi and consequently extensively studied in these organisms. One member of this group was identified as a regulator of a lipase coding gene in *Aspergillus oryzae* [59]. Other fungal TRFs have been linked to nitrogen metabolism and assimilation in yeast and fungi in general [60–62].

The two GCNs built from the DEGs illustrate the transcriptional regulations underlying the response of each strain to nitrogen deprivation. Consequently, the 2Xc1 GCN driven by TFs belonging to families known to be involved in nitrogen-related processes is likely to be at least partly responsible for the mutant phenotype. To identify the potential role of each TF in the establishment of this phenotype, the functional annotation of the co-expressed genes was completed by gene prioritization, based on WGCNA data. The functions enriched in these three modules were diverse.

Such a functional diversity is explained by the overall reorganization of the metabolism driven by changes in nitrogen availability. These conditions are already known to induce various changes in gene expression and metabolic re-modelling, which are not necessarily directly related to lipid or carbohydrate biosynthesis [56,63,64].

4.2. WGCNA-based gene prioritization, a way to identify phenotypically relevant genes

To evaluate the relevance of gene prioritization via WGCNA, a GO enrichment was carried out for the 284 prioritized genes (Table S5). This analysis showed an enrichment for carbohydrate and chlorophyll metabolism, which are known to be impacted at gene, protein and chlorophyll content levels in algae grown under nitrogen deprivation [50,56,64–66]. Ubiquitin-dependent protein degradation as well as tRNA and mRNA processing were also enriched. These functions illustrate the impact of nitrogen deprivation on cell mRNA and protein content, which are largely impacted in order to limit nitrogen utilization [64,66,67]. Finally, the protein-arginine deiminase identified in the *T. lutea* genome was selected by the gene prioritization. This enzyme produces ammonia from arginine and, interestingly, arginine content was decreased in *C. reinhardtii* cells grown under nitrogen

deprivation [64]. A nutrient-limiting growth environment creates selective pressure so that cellular demand for the limiting nutrient is minimized [68]. Therefore, this mechanism is a part of an optimization of the proteome amino acid content under nitrogen deprivation since arginine contains three nitrogen molecules in its side chain [69].

The prioritized genes are related to storage lipids and carbohydrate content dynamics in *T. lutea*, which are induced by nitrogen deprivation. The fact that the functions enriched are known to be involved in the algal nitrogen deprivation response suggests that the WGCNA-based gene prioritization is a relevant strategy to identify genes related to a physiological feature. Consequently, network analysis appears to be a relevant strategy to emphasize biologically meaningful results, coping with functional annotation issues that occur in non-model organisms.

4.3. Analysis of the co-expression network of differentially expressed genes identified TFs potentially involved in mutant phenotype-specific functions

The complementary association of the functional annotation coupled with the gene prioritization allowed us to identify functions specific to the 2Xc1 GCN strain (Fig. 4). This assigned a putative role in the mutant phenotype establishment to six TF candidates. The *HB-other_9_PAS* TF was linked to functions involved in protein degradation, maturation and trafficking. Under nitrogen deprivation, algal proteomes are largely impacted, in particular through the decrease of protein synthesis and the increase of proteolysis [66,67,74]. These mechanisms make it possible to reduce nitrogen utilization and supply energy to the synthesis of carbon storage compounds [66,74,75]. The presence among the genes co-expressed with this TF of a diacylglycerol O-acyltransferase involved in TAG synthesis in nitrogen deprivation conditions [50], as well as an enolase and a mannose-6-phosphate isomerase involved in carbohydrate metabolism, strongly point towards proteome remodelling and carbon storage compound synthesis.

For *MYB-2R_14* and *NF-YB_2* TFs, which drive the red cluster, functions related to DNA repair with a specific action on pyrimidine dimers were identified. These pyrimidine dimers are induced by UV-light irradiation and also by oxidative stress [76], which is induced by nitrogen deprivation in algae [77]. Moreover, oxidative stress is thought to mediate lipid accumulation under nitrogen deprivation in green algae [78,79], but the underlying mechanisms are still unclear. One hypothesis concerning this relationship is that TAG synthesis may consume excessive photoassimilates that are unused because of the nitrogen deprivation, thus preventing oxidative stress [80,81]. Interestingly, photosynthesis is also enriched in the red cluster with four proteins involved in light harvesting. Under nitrogen deprivation, the photosynthesis is largely impacted at mRNA, protein and physiological levels [50,56,64–66]. These functions converge towards the linking of *MYB-2R_14* and *NF-YB_2* TFs to photosynthesis and oxidative stress response, possibly through TAG synthesis. This was reinforced by the enrichment of genes prioritized for their link to cell storage lipid content dynamics among the genes co-expressed with these TFs in the mutant strain.

MYB-2R_20, *MYB-rel_11* and *GATA_2* TFs, driving the green cluster, were linked to carbohydrates with an UDP-glucose pyrophosphorylase. In haptophytes, carbohydrates can be stored in the form of chrysolaminarin [82,83]. Very interestingly, UDP-glucose pyrophosphorylase is thought to be a rate-limiting enzyme playing an important role in carbon allocation and chrysolaminarin synthesis in the diatom *P. tricornutum* [84]. In addition to carbohydrate synthesis, an ammonium transporter, a urea transporter and genes involved in proteolysis were linked to *MYB-2R_20*, *MYB-rel_11* and *GATA_2* TFs. Protein degradation allows recycling of the nitrogen from proteins and is assumed to feed carbon storage compound synthesis in response to nitrogen stress in microalgae [66,74,75]. The urea cycle is also used for intracellular nitrogen recycling in diatoms [85] as well as in the model haptophyte *Emiliania huxleyi* [86].

These results strongly suggest an involvement of *GATA_2* and,

especially, *MYB-2R_20* and *MYB-rel_11* TFs in carbon recycling, feeding chrysolaminarin synthesis in the mutant strain. This idea is strengthened by an enrichment of prioritized genes with a link to cell carbohydrate content among the genes co-expressed with these TFs. An involvement in nitrogen uptake and recycling, key mechanisms in algal response to nitrogen deprivation, is also strongly suggested.

Since a functional confirmation is not possible by knock-down or knock-out because these tools are not developed yet in *T. lutea*, we used a bioinformatic approach to provide more clues concerning the reliability of the GCNs. With this purpose, the identification of motifs corresponding to known TFBS of TFs belonging to the same families of GATA_2, Fungal-TRF_8, MYB-related_11, MYB-2R_14, MYB-2R_20 and NF-YB_2 in the promoting sequences of their respective co-expressed genes suggests (1) that the GCNs and their subsequent clustering are reliable, and (2) that some of the genes co-expressed with the TFs may be regulated by these TFs. Moreover, the motif found in the promoting sequences of MYB-2R_14 putative target genes is bound by the MYB30 TF of *A. thaliana* which is known to be involved in lipid synthesis [42]. Given the suggestion that some of the genes co-expressed with this TF could be targeted by it, this function strengthens the link of MYB-2R_14 to lipid content in the mutant strain. The same functional convergence is found for the MYB-2R_20 TF for which a motif bound by the MYB83, involved in polysaccharide synthesis in *A. thaliana* [43] was identified.

However, since the relations between the TFs and putative target genes are inferred from co-expression (i.e., correlation based), they require further confirmation. Nevertheless, the advantage of using co-expression network is that the modules identified group co-expressed genes that are often involved in the same pathway or response [87]. Such analysis allows the understanding of mechanisms involved in a process or a phenotype [88–90]. Thus, co-expression networks were successfully used to identify TFs involved in the response of the microalga *C. reinhardtii* to nitrogen deprivation [56]. Target genes of the mouse TF Hoxc8 [91], as well as TFs involved in seed longevity of *Medicago truncatula* [92] were also identified using co-expression. Such studies prove that co-expression network analysis is an accurate and meaningful method to decipher key mechanisms and their cognate regulators in the context of the present work.

It should also be noted that gene expression regulation is a very complex process involving many actors working in combination [93]. The impact of these gene expression modifications could be modulated by post-transcriptional modifications involving, for example, microRNA [94,95]. Post-translational modifications such as riboswitches [96], RNA-proteins interactions [97,98] or tRNAs disponibility [99] could also modulate proteins production. Concerning the impact of TFs on the mutant phenotype, the regulation of their activity could also have a role. TFs activity is regulated by post-translational modifications such as acetylation [100], phosphorylation [101,102], sumoylation or ubiquitination [103,104]. An upregulation at the transcriptional or translational level of a TF would not have an impact on the cell functioning in the case of activation or inactivation by such regulatory modifications [105,106].

4.4. MYB-2R_20 and MYB-rel_11 TFs are linked to nitrogen uptake and recycling under nitrogen deprivation

The q-RT-PCR analysis showed a co-expression of *MYB-2R_20* and *MYB-rel_11* TFs with the *Nrt*2.1 coding gene. This confirmed the potential link of these two TFs to nitrogen uptake in the mutant strain. Interestingly, the *MYB-rel_11* TF was only co-expressed with *PLAAOx*, *CSAP*, *Nrt*2.1 and *MYB-2R_20* in the mutant strain (Table 2), suggesting a specific role of this TF in *T. lutea* 2Xc1. Moreover, *PLAAOx*, *CSAP* and *Nrt*2.1 genes were down-regulated throughout the whole spike (Table S9). This differential expression suggests a different role of these three genes in the response of the two strains to nitrogen deprivation.

The high co-expressions of *PLAAOx*, *CSAP* and *Nrt*2.1 with each other, as well as their respective co-expression with *MYB-2R_20* TF in
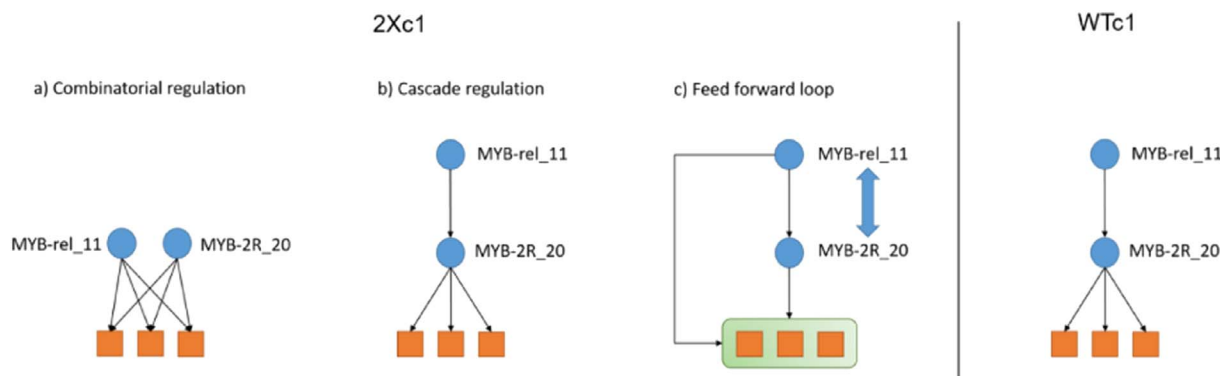
Fig. 7. Hypotheses for the regulatory action of MYB-rel_11 in the 2Xc1 strain. MYB-rel_11 and MYB-2R_20 are represented by blue circles and the potential target genes PLAAOx, CSAP and Nrt2.1 by orange squares. a) Combinatorial regulation of the target genes by the two TFs. b) Cascade regulation in which MYB-rel_11 regulates the three target genes through that of MYB-2R_20. c) Feed forward loop in which MYB-rel_11 regulates the three target genes both directly and through the regulation of MYB-2R_20. In the feed forward loop, the role of the two TFs could be reversed. Comparatively, in the WTc1 strain, MYB-2R_20 seem to be the only direct regulator despite the co-expression of both TFs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

both strains, may suggest that this TF could be involved in the co-regulation of these three genes in both WTc1 and 2Xc1 strains. However, *MYB-rel_11* TF was only co-expressed with *PLAAOx*, *CSAP* and *Nrt*2.1 genes in the 2Xc1 strain. This strain-specific co-expression may suggest that *MYB-rel_11* TF is involved in the differential expression of these three genes in the mutant strain. *MYB-rel_11* could exert this effect by: (i) the direct gene expression regulation of *PLAAOx*, *CSAP* and *Nrt*2.1 coding genes in coordination with *MYB-2R_20* (Fig. 7a); (ii) the regulation of *MYB-2R_20* TF gene expression which, in turn, regulates the expression of *PLAAOx*, *CSAP* and *Nrt*2.1 coding genes in a regulatory cascade (Fig. 7b); or (iii) the expression regulation of *MYB-2R_20* together with *PLAAOx*, *CSAP* and *Nrt*2.1 coding genes in a feed forward loop, strengthening the regulation of these three genes (Fig. 7c). In this later hypothesis, the role of the two TFs could be reversed in the feed forward loop: *MYB-2R_20* regulating the expression of *PLAAOx*, *CSAP* and *Nrt*2.1 coding genes directly and via the regulation of *MYB-rel_11*.

Since the relations between the TFs and their putative target genes are inferred from co-expression, they remain subject of further investigations. It should be also noted that gene expression regulation is a very complex process involving many actors working in combination [68]. Consequently, the role of the *MYB-2R_20* and *MYB-rel_11* TFs is likely to be a part of the picture but not the only mechanism of *Nrt*2.1, *PLAAOx* and *CSAP* gene expression regulation.

In addition to these regulatory specificities, the expression profiles of *PLAAOx*, *CSAP* and *Nrt*2.1 coding genes are specific and seem to be dependent on the physiological state of the microalgae. Their expression is decreased or repressed quickly following the nitrogen spike (Fig. 6) and then increases gradually as the microalgae absorb the injected nitrogen (nitrogen repletion conditions) and the N/C ratio stabilizes (Fig. 6).

This high N/C ratio is the consequence of the absorption of the injected nitrogen by microalgae. Consequently, as the N/C ratio increases, nitrogen availability decreases. This expression dependent on the nitrogen availability was already shown for the *Nrt*2.1 transporter in *T. lutea* [55]. These high-affinity transporters allow an increase of nitrogen absorption during its scarcity in the medium and, consequently, provide a nitrogen supply. The fact that the *PLAAOx* and *CSAP* coding genes have the same expression behaviour suggests a function related to the nitrogen availability. This function has also been suggested by other studies, since a homologous protein of the *PLAAOx* was involved in the response of *C. reinhardtii* to nitrogen deprivation [107,108] and was thought to provide ammonia from amino-acid deamination. Moreover, a homologous gene of *CSAP* was differentially expressed in *P. tricornutum* under nitrogen deprivation and thought to be involved in carbon homeostasis because of its putative decarboxylase function [109].

5. Conclusions

The analysis and comparison of the differentially co-expressed gene network of *T. lutea* strains WTc1 and 2Xc1 identified candidate TFs potentially involved in the phenotype establishment of the mutant strain. In particular, the *MYB-2R_14* and *NF-YB_2* TFs seem to be involved in photosynthesis through light harvesting as well as oxidative stress responses and prevention, possibly through TAG synthesis. The *GATA_2*, *MYB-rel_11* and *MYB-2R_20* TFs are likely to be involved in nitrogen uptake and carbon and nitrogen recycling, feeding carbohydrate synthesis in the form of chrysolaminarin. The real time PCR analysis further characterizes the role of *MYB-rel_11* and *MYB-2R_20* TFs in nitrogen uptake and their potential involvement in carbon and nitrogen recycling.

The *MYB-2R_20* TF is likely to regulate *PLAAOx*, *CSAP* and *Nrt*2.1 gene expression in both strains, participating in their co-expression. In the 2Xc1 strain, *MYB-rel_11* TF may act either coordinately with *MYB-2R_20* TF or as a master regulator (Fig. 7) to regulate the expression of *PLAAOx*, *CSAP* and *Nrt*2.1 genes, resulting in their differential expression. The fact that these three genes are co-expressed and have an expression profile dependent on nitrogen availability suggests an involvement in the nitrogen deprivation response. The action of these three genes could provide nitrogen as well as carbon resources, allowing the microalgae to cope with nitrogen deprivation. Moreover, since *MYB-rel_11* TF was linked to cell carbohydrate content by the gene prioritization and functional annotation of these co-expressed genes, the carbon resources provided may be directed towards carbohydrate synthesis. Finally, the q-RT-PCR confirmation of the putative role of *MYB-2R_20* and *MYB-rel_11* TFs shows that our co-expression network-based strategy is sufficiently accurate to identify TFs potentially involved in the establishment of a mutant phenotype and characterize their putative function. Although the putative role of these TFs needs to be confirmed by further functional investigations and modelling that consider gene regulations per se, the *MYB-rel_11* TF is nonetheless an interesting candidate in the response of *T. lutea* to nitrogen deprivation.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2017.12.011.

Abbreviations

CSAP	Coccolith scale associated protein
DBD	DNA binding domain
DEGs	Differentially expressed genes
Fungal-TRF	Fungal transcriptional regulatory factor
GCNs	Gene Co-expression networks
GO	Gene Ontology
HB	Homeobox
MYB-rel	MYB related
NF-YB	Nuclear factor Y, B subunit
NRT	Nitrate transporter
PLAAOx	Periplasmic L-Amino-Acid Oxidase
TAG	Triacylglycerol
TF	Transcription factor
TFBS	Transcription factor binding site
WGCNA	Weighted gene co-expression analysis

Author contributions

STR designed the strategy, analyzed and interpreted the data, designed and validated the q-RT-PCR primers, and drafted the manuscript. GC participated in the coordination of the study, the assemblage of the genome and the drafting of the manuscript. CT and DE participated in the design of the study and helped to draft the manuscript. BC, BS and BSJ participated in the design and coordination of the study and helped to draft the manuscript. JPC and GB participated in the coordination of the study. All authors read and approved the finalised manuscript.

References

[1] C.B. Field, M.J. Behrenfeld, J.T. Randerson, P. Falkowski, Primary production of the biosphere: integrating terrestrial and oceanic components, Science 281 (1998) 237–240, http://dx.doi.org/10.1126/science.281.5374.237.

[2] P.G. Falkowski, M.E. Katz, A.H. Knoll, A. Quigg, J.A. Raven, O. Schofield, F.J.R. Taylor, The evolution of modern eukaryotic phytoplankton, Science 305 (2004) 354–360, http://dx.doi.org/10.1126/science.1095964.

[3] M.D. Guiry, How many species of algae are there? J. Phycol. 48 (2012) 1057–1063, http://dx.doi.org/10.1111/j.1529-8817.2012.01222.x.

[4] F. Not, R. Siano, W.H.C.F. Kooistra, N. Simon, D. Vaulot, I. Probert, Diversity and Ecology of Eukaryotic Marine Phytoplankton, Adv. Bot. Res. Elsevier, 2012, pp. 1–53 http://linkinghub.elsevier.com/retrieve/pii/B9780123914996000013 (accessed June 28, 2016).

[5] P. Spolaore, C. Joannis-Cassan, E. Duran, A. Isambert, Commercial applications of microalgae, J. Biosci. Bioeng. 101 (2006) 87–96, http://dx.doi.org/10.1263/jbb. 101.87.

[6] Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, A. Darzins, Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, Plant J. Cell Mol. Biol. 54 (2008) 621–639, http://dx.doi.org/10.1111/ j.1365-313X.2008.03492.x.

[7] J.-P. Cadoret, M. Garnier, B. Saint-Jean, G. Piganeau (Ed.), Chapter Eight - Microalgae, Functional Genomics and Biotechnology, Adv. Bot. Res. Academic Press, 2012, pp. 285–341 http://www.sciencedirect.com/science/article/pii/ B9780123914996000086 (accessed April 12, 2016).

[8] V. Mimouni, L. Ulmann, V. Pasquet, M. Mathieu, L. Picot, G. Bougaran, J.-P. Cadoret, A. Morant-Manceau, B. Schoefs, The potential of microalgae for the production of bioactive molecules of pharmaceutical interest, Curr. Pharm. Biotechnol. 13 (2012) 2733–2750.

[9] H. Gateau, K. Solymosi, J. Marchand, B. Schoefs, Carotenoids of microalgae used in food industry and medicine, Mini-Rev. Med. Chem. 17 (2016) 1140–1172.

[10] N.K. Kang, S. Jeon, S. Kwon, H.G. Koh, S.-E. Shin, B. Lee, G.-G. Choi, J.-W. Yang, B.-R. Jeong, Y.K. Chang, Effects of overexpression of a bHLH transcription factor on biomass and lipid production in Nannochloropsis salina, Biotechnol. Biofuels. 8 (2015) 200, , http://dx.doi.org/10.1186/s13068-015-0386-9.

[11] G. Perin, A. Bellan, A. Segalla, A. Meneghesso, A. Alboresi, T. Morosinotto, Generation of random mutants to improve light-use efficiency of Nannochloropsis

[12] gaditana cultures for biofuel production, Biotechnol. Biofuels. 8 (2015) 161, , http://dx.doi.org/10.1186/s13068-015-0337-5.

[12] J. Anthony, V.R. Rangamaran, D. Gopal, K.T. Shivasankarasubbiah, M.L.J. Thilagam, M. Peter Dhassiah, D.S.M. Padinjattayil, V.N. Valsalan, V. Manambrakat, S. Dakshinamurthy, S. Thirunavukkarasu, K. Ramalingam, Ultraviolet and 5′fluorodeoxyuridine induced random mutagenesis in Chlorella vulgaris and its impact on fatty acid profile: a new insight on lipid-metabolizing genes and structural characterization of related proteins, Mar. Biotechnol. (N.Y.) 17 (2015) 66–80, http://dx.doi.org/10.1007/s10126-014-9597-5.

[13] P.D. Weyman, K. Beeri, S.C. Lefebvre, J. Rivera, J.K. McCarthy, A.L. Heuberger, G. Peers, A.E. Allen, C.L. Dupont, Inactivation of Phaeodactylum tricornutum urease gene using transcription activator-like effector nuclease-based targeted mutagenesis, Plant Biotechnol. J. 13 (2015) 460–470, http://dx.doi.org/10.1111/pbi. 12254.

[14] H. Liu, I. Probert, J. Uitz, H. Claustre, S. Aris-Brosou, M. Frada, F. Not, C. de Vargas, Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12803–12808, http://dx.doi.org/10.1073/pnas.0905841106.

[15] I.T. Marlowe, J.C. Green, A.C. Neal, S.C. Brassell, G. Eglinton, P.A. Course, Long chain (n-C37–C39) alkenones in the Prymnesiophyceae. Distribution of alkenones and other lipids and their taxonomic significance, Br. Phycol. J. 19 (1984) 203–2016.

[16] S. Theroux, W.J. D'Andrea, J. Toney, L. Amaral-Zettler, Y. Huang, Phylogenetic diversity and evolutionary relatedness of alkenone-producing haptophyte algae in lakes: implications for continental paleotemperature reconstructions, Earth Planet. Sci. Lett. 300 (2010) 311–320, http://dx.doi.org/10.1016/j.epsl.2010.10.009.

[17] T.M. Mata, A.A. Martins, N.S. Caetano, Microalgae for biodiesel production and other applications: a review, Renew. Sust. Energ. Rev. 14 (2010) 217–232, http:// dx.doi.org/10.1016/j.rser.2009.07.020.

[18] A. Sukenik, R. Wahnon, Biochemical quality of marine unicellular algae with special emphasis on lipid composition. I. Isochrysis galbana, Aquaculture 97 (1991) 61–72, http://dx.doi.org/10.1016/0044-8486(91)90279-G.

[19] M.L. Eltgroth, R.L. Watwood, G.V. Wolfe, Production and cellular localization of neutral long-chain lipids in the haptophyte algae Isochrysis galbana and Emiliania huxleyi, J. Phycol. 41 (2005) 1000–1009, http://dx.doi.org/10.1111/j.1529-8817. 2005.00128.x.

[20] L. Ulmann, V. Blanckaert, V. Mimouni, M.X. Andersson, B. Schoefs, B. Chénais, Microalgal fatty acids and their implication in health and disease, Mini-Rev. Med. Chem. 17 (2016) 1112–1123.

[21] G. Bougaran, C. Rouxel, N. Dubois, R. Kaas, S. Grouas, E. Lukomska, J.-R. Le Coz, J.-P. Cadoret, Enhancement of neutral lipid productivity in the microalga Isochrysis affinis Galbana (T-Iso) by a mutation-selection procedure, Biotechnol. Bioeng. 109 (2012) 2737–2745, http://dx.doi.org/10.1002/bit.24560.

[22] G. Carrier, M. Garnier, L. Le Cunff, G. Bougaran, I. Probert, C. De Vargas, E. Corre, J.-P. Cadoret, B. Saint-Jean, Comparative transcriptome of wild type and selected strains of the microalgae Tisochrysis lutea provides insights into the genetic basis, lipid metabolism and the life cycle, PLoS One 9 (2014) e86889, , http://dx.doi. org/10.1371/journal.pone.0086889.

[23] M. Garnier, G. Carrier, H. Rogniaux, E. Nicolau, G. Bougaran, B. Saint-Jean, J.P. Cadoret, Comparative proteomics reveals proteins impacted by nitrogen deprivation in wild-type and high lipid-accumulating mutant strains of Tisochrysis lutea, J. Proteome 105 (2014) 107–120, http://dx.doi.org/10.1016/j.jprot.2014. 02.022.

[24] M. Garnier, G. Bougaran, M. Pavlovic, J.-B. Berard, G. Carrier, A. Charrier, F. Le Grand, E. Lukomska, C. Rouxel, N. Schreiber, J.-P. Cadoret, H. Rogniaux, B. Saint-Jean, Use of a lipid rich strain reveals mechanisms of nitrogen limitation and carbon partitioning in the haptophyte Tisochrysis lutea, Algal Res. 20 (2016) 229–248, http://dx.doi.org/10.1016/j.algal.2016.10.017.

[25] S. Thiriet-Rupert, G. Carrier, B. Chénais, C. Trottier, G. Bougaran, J.-P. Cadoret, B. Schoefs, B. Saint-Jean, Transcription factors in microalgae: genome-wide prediction and comparative analysis, BMC Genomics 17 (2016) 282, , http://dx.doi. org/10.1186/s12864-016-2610-9.

[26] J. Rumin, H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J.-P. Cadoret, G. Bougaran, The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae, Biotechnol. Biofuels. 8 (2015) 42, http://dx.doi.org/ 10.1186/s13068-015-0220-4.

[27] M. DuBois, K.A. Gilles, J.K. Hamilton, P.A. Rebers, F. Smith, Colorimetric method for determination of sugars and related substances, Anal. Chem. 28 (1956) 350–356, http://dx.doi.org/10.1021/ac60111a017.

[28] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet. Journal 17 (2011) 10–12.

[29] S. Andrews, FastQC: a quality control tool for high throughput sequence data, (2010) http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 7, 2017).

[30] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinforma. Oxf. Engl. 25 (2009) 1105–1111, http://dx.doi.org/10. 1093/bioinformatics/btp120.

[31] S. Anders, P.T. Pyl, W. Huber, HTSeq—a python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169, http://dx.doi. org/10.1093/bioinformatics/btu638.

[32] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinforma. Oxf. Engl. 21 (2005) 3674–3676, http://dx.doi.org/10. 1093/bioinformatics/bti610.

[33] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinformatics. 9 (2008) 559, http://dx.doi.org/10.1186/1471-

2105-9-559.

[34] J. Feng, C.A. Meyer, Q. Wang, J.S. Liu, X. Shirley Liu, Y. Zhang, GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data, Bioinforma. Oxf. Engl. 28 (2012) 2782–2788, http://dx.doi.org/10.1093/bioinformatics/bts515.

[35] L.C. Xia, D. Ai, J. Cram, J.A. Fuhrman, F. Sun, Efficient statistical significance approximation for local similarity analysis of high-throughput time series data, Bioinforma. Oxf. Engl. 29 (2013) 230–237, http://dx.doi.org/10.1093/bioinformatics/bts668.

[36] L.C. Xia, J.A. Steele, J.A. Cram, Z.G. Cardon, S.L. Simmons, J.J. Vallino, J.A. Fuhrman, F. Sun, Extended local similarity analysis (eLSA) of microbial community and other time series data with replicates, BMC Syst. Biol. 5 (Suppl. 2) (2011) S15, http://dx.doi.org/10.1186/1752-0509-5-S2-S15.

[37] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, in: Third Int. AAAI Conf. Weblogs Soc. Media, http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154, (2009) , Accessed date: 23 May 2016.

[38] A. Mathelier, O. Fornes, D.J. Arenillas, C.-Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A.W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W.W. Wasserman, JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles, Nucleic Acids Res. 44 (2016) D110–115, http://dx.doi.org/10.1093/nar/gkv1176.

[39] S. Holmberg, P. Schjerling, Cha4p of Saccharomyces cerevisiae activates transcription via serine/threonine response elements, Genetics 144 (1996) 467–478.

[40] J.G. Petersen, M.C. Kielland-Brandt, T. Nilsson-Tillgren, C. Bornaes, S. Holmberg, Molecular genetics of serine and threonine catabolism in Saccharomyces cerevisiae, Genetics 119 (1988) 527–534.

[41] R.C. O'Malley, S.-S.C. Huang, L. Song, M.G. Lewsey, A. Bartlett, J.R. Nery, M. Galli, A. Gallavotti, J.R. Ecker, Cistrome and epicistrome features shape the regulatory DNA landscape, Cell 165 (2016) 1280–1292, http://dx.doi.org/10.1016/j.cell.2016.04.038.

[42] S. Raffaele, F. Vailleau, A. Léger, J. Joubès, O. Miersch, C. Huard, E. Blée, S. Mongrand, F. Domergue, D. Roby, A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in Arabidopsis, Plant Cell 20 (2008) 752–767, http://dx.doi.org/10.1105/tpc.107.054858.

[43] R. Zhong, Z.-H. Ye, MYB46 and MYB83 bind to the SMRE sites and directly activate a suite of transcription factors and secondary wall biosynthetic genes, Plant Cell Physiol. 53 (2012) 368–380, http://dx.doi.org/10.1093/pcp/pcr185.

[44] H. Jeong, S.P. Mason, A.-L. Barabási, Z.N. Oltvai, Lethality and centrality in protein networks, Nature 411 (2001) 41–42, http://dx.doi.org/10.1038/35075138.

[45] S.L. Carter, C.M. Brechbühler, M. Griffin, A.T. Bond, Gene co-expression network topology provides a framework for molecular characterization of cellular state, Bioinformatics 20 (2004) 2242–2250, http://dx.doi.org/10.1093/bioinformatics/bth234.

[46] T.F. Cooper, A.P. Morby, A. Gunn, D. Schneider, Effect of random and hub gene disruptions on environmental and mutational robustness in Escherichia coli, BMC Genomics 7 (2006) 1.

[47] B.-H. Kang, E. Nielsen, M.L. Preuss, D. Mastronarde, L.A. Staehelin, Electron tomography of RabA4b- and PI-4Kβ1-labeled trans Golgi network compartments in Arabidopsis, Traffic Cph. Den. 12 (2011) 313–329, http://dx.doi.org/10.1111/j.1600-0854.2010.01146.x.

[48] M. Fujimoto, Y. Suda, S. Vernhettes, A. Nakano, T. Ueda, Phosphatidylinositol 3-kinase and 4-kinase have distinct roles in intracellular trafficking of cellulose synthase complexes in Arabidopsis thaliana, Plant Cell Physiol. 56 (2015) 287–298, http://dx.doi.org/10.1093/pcp/pcu195.

[49] V. Ghugtyal, R. Garcia-Rodas, A. Seminara, S. Schaub, M. Bassilana, R.A. Arkowitz, Phosphatidylinositol-4-phosphate-dependent membrane traffic is critical for fungal filamentous growth, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 8644–8649, http://dx.doi.org/10.1073/pnas.1504259112.

[50] J. Msanne, D. Xu, A.R. Konda, J.A. Casas-Mollano, T. Awada, E.B. Cahoon, H. Cerutti, Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae Chlamydomonas reinhardtii and Coccomyxa sp. C-169, Phytochemistry 75 (2012) 50–59, http://dx.doi.org/10.1016/j.phytochem.2011.12.007.

[51] C.P. Selby, A. Sancar, A cryptochrome/photolyase class of enzymes with single-stranded DNA-specific photolyase activity, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 17696–17700, http://dx.doi.org/10.1073/pnas.0607993103.

[52] M. Hildebrand, K. Dahlin, Nitrate transporter genes from the diatom cylindrotheca fusiformis (Bacillariophyceae): mRna levels controlled by nitrogen source and by the cell cycle, J. Phycol. 36 (2000) 702–713, http://dx.doi.org/10.1046/j.1529-8817.2000.99153.x.

[53] B. Song, B.B. Ward, Molecular cloning and characterization of high-affinity nitrate transporters in marine phytoplankton, J. Phycol. 43 (2007) 542–552, http://dx.doi.org/10.1111/j.1529-8817.2007.00352.x.

[54] L.-K. Kang, S.-P.L. Hwang, G.-C. Gong, H.-J. Lin, P.-C. Chen, J. Chang, Influences of nitrogen deficiency on the transcript levels of ammonium transporter, nitrate transporter and glutamine synthetase genes in Isochrysis galbana (Isochrysidales, Haptophyta), Phycologia 46 (2007) 521–533, http://dx.doi.org/10.2216/06-44.1.

[55] A. Charrier, J.-B. Bérard, G. Bougaran, G. Carrier, E. Lukomska, N. Schreiber, F. Fournier, A.F. Charrier, C. Rouxel, M. Garnier, J.-P. Cadoret, B. Saint-Jean, High-affinity nitrate/nitrite transporter genes (Nrt2) in Tisochrysis lutea: identification and expression analyses reveal some interesting specificities of Haptophyta microalgae, Physiol. Plant. 154 (2015) 572–590, http://dx.doi.org/10.1111/ppl.12330.

[56] M. Gargouri, J.-J. Park, F.O. Holguin, M.-J. Kim, H. Wang, R.R. Deshpande,

Y. Shachar-Hill, L.M. Hicks, D.R. Gang, Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii, J. Exp. Bot. 66 (2015) 4551–4566, http://dx.doi.org/10.1093/jxb/erv217.

[57] S. Imamura, Y. Kanesaki, M. Ohnuma, T. Inouye, Y. Sekine, T. Fujiwara, T. Kuroiwa, K. Tanaka, R2R3-type MYB transcription factor, CmMYB1, is a central nitrogen assimilation regulator in Cyanidioschyzon merolae, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 12548–12553, http://dx.doi.org/10.1073/pnas.0902790106.

[58] E. Rayko, F. Maumus, U. Maheswari, K. Jabbari, C. Bowler, Transcription factor families inferred from genome sequences of photosynthetic stramenopiles, New Phytol. 188 (2010) 52–66, http://dx.doi.org/10.1111/j.1469-8137.2010.03371.x.

[59] S.M. Garrido, N. Kitamoto, A. Watanabe, T. Shintani, K. Gomi, Functional analysis of FarA transcription factor in the regulation of the genes encoding lipolytic enzymes and hydrophobic surface binding protein for the degradation of biodegradable plastics in Aspergillus oryzae, J. Biosci. Bioeng. 113 (2012) 549–555, http://dx.doi.org/10.1016/j.jbiosc.2011.12.014.

[60] G.A. Marzluf, Genetic regulation of nitrogen metabolism in the fungi, Microbiol. Mol. Biol. Rev. MMBR. 61 (1997) 17–32.

[61] J. Avila, C. González, N. Brito, F. Machín, M.D. Pérez, J.M. Siverio, A second Zn(II)(2)Cys(6) transcriptional factor encoded by the YNA2 gene is indispensable for the transcriptional activation of the genes involved in nitrate assimilation in the yeast Hansenula polymorpha, Yeast Chichester Engl. 19 (2002) 537–544, http://dx.doi.org/10.1002/yea.847.

[62] J. Avila, C. González, N. Brito, J.M. Siverio, Clustering of the YNA1 gene encoding a Zn(II)2Cys6 transcriptional factor in the yeast Hansenula polymorpha with the nitrate assimilation genes YNT1, YNI1 and YNR1, and its involvement in their transcriptional activation, Biochem. J. 335 (Pt 3) (1998) 647–652.

[63] L. Valledor, T. Furuhashi, L. Recuenco-Muñoz, S. Wienkoop, W. Weckwerth, System-level network analysis of nitrogen starvation and recovery in Chlamydomonas reinhardtii reveals potential new targets for increased lipid accumulation, Biotechnol. Biofuels. 7 (2014) 171, , http://dx.doi.org/10.1186/s13068-014-0171-1.

[64] S. Schmollinger, T. Mühlhaus, N.R. Boyle, I.K. Blaby, D. Casero, T. Mettler, J.L. Moseley, J. Kropat, F. Sommer, D. Strenkert, D. Hemme, M. Pellegrini, A.R. Grossman, M. Stitt, M. Schroda, S.S. Merchant, Nitrogen-sparing mechanisms in chlamydomonas affect the transcriptome, the proteome, and photosynthetic metabolism, Plant Cell 26 (2014) 1410–1435, http://dx.doi.org/10.1105/tpc.113.122523.

[65] M.T. Juergens, R.R. Deshpande, B.F. Lucker, J.-J. Park, H. Wang, M. Gargouri, F.O. Holguin, B. Disbrow, T. Schaub, J.N. Skepper, D.M. Kramer, D.R. Gang, L.M. Hicks, Y. Shachar-Hill, The regulation of photosynthetic structure and function during nitrogen deprivation in Chlamydomonas reinhardtii, Plant Physiol. 167 (2015) 558–573, http://dx.doi.org/10.1104/pp.114.250530.

[66] L. Alipanah, J. Rohloff, P. Winge, A.M. Bones, T. Brembu, Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum, J. Exp. Bot. (2015) erv340, http://dx.doi.org/10.1093/jxb/erv340.

[67] A. López García de Lomana, S. Schäuble, J. Valenzuela, S. Imam, W. Carter, D.D. Bilgin, C.B. Yohn, S. Turkarslan, D.J. Reiss, M.V. Orellana, N.D. Price, N.S. Baliga, Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii, Biotechnol. Biofuels. 8 (2015) 207, , http://dx.doi.org/10.1186/s13068-015-0391-z.

[68] P. Baudouin-Cornu, Y. Surdin-Kerjan, P. Marlière, D. Thomas, Molecular evolution of protein atomic composition, Science 293 (2001) 297–300, http://dx.doi.org/10.1126/science.1061052.

[69] S.S. Merchant, J.D. Helmann, Elemental economy: microbial strategies for optimizing growth in the face of nutrient limitation, Adv. Microb. Physiol. 60 (2012) 91–210, http://dx.doi.org/10.1016/B978-0-12-398264-3.00002-4.

[74] H.-P. Dong, E. Williams, D. Wang, Z.-X. Xie, R. Hsia, A. Jenck, R. Halden, J. Li, F. Chen, A.R. Place, Responses of Nannochloropsis oceanica IMET1 to long-term nitrogen starvation and recovery, Plant Physiol. 162 (2013) 1110–1126, http://dx.doi.org/10.1104/pp.113.214320.

[75] A.F. da Silva, S.O. Lourenço, R.M. Chaloub, Effects of nitrogen starvation on the photosynthetic physiology of a tropical marine microalga Rhodomonas sp. (Cryptophyceae), Aquat. Bot. 91 (2009) 291–297, http://dx.doi.org/10.1016/j.aquabot.2009.08.001.

[76] M. Hochberg, R. Kohen, C.D. Enk, Role of antioxidants in prevention of pyrimidine dimer formation in UVB irradiated human HaCaT keratinocytes, Biomed. Pharmacother. Bioméd. Pharmacothérapie. 60 (2006) 233–237, http://dx.doi.org/10.1016/j.biopha.2006.04.008.

[77] W. Liu, Z. Huang, P. Li, J. Xia, B. Chen, Formation of triacylglycerol in Nitzschia closterium f. minutissima under nitrogen limitation and possible physiological and biochemical mechanisms, J. Exp. Mar. Biol. Ecol. 418–419 (2012) 24–29, http://dx.doi.org/10.1016/j.jembe.2012.03.005.

[78] Y.-M. Zhang, H. Chen, C.-L. He, Q. Wang, Nitrogen starvation induced oxidative stress in an oil-producing green alga Chlorella sorokiniana C3, PLoS One 8 (2013) e69225, , http://dx.doi.org/10.1371/journal.pone.0069225.

[79] K. Yilancioglu, M. Cokol, I. Pastirmaci, B. Erman, S. Cetiner, Oxidative stress is a mediator for increased lipid accumulation in a newly isolated Dunaliella salina strain, PLoS One 9 (2014), http://dx.doi.org/10.1371/journal.pone.0091957.

[80] A.E. Solovchenko, Physiological role of neutral lipid accumulation in eukaryotic microalgae under stresses, Russ. J. Plant Physiol. 59 (2012) 167–176, http://dx.doi.org/10.1134/S1021443712020161.

[81] Y. Lemoine, B. Schoefs, Secondary ketocarotenoid astaxanthin biosynthesis in algae: a multifunctional response to stress, Photosynth. Res. 106 (2010) 155–177, http://dx.doi.org/10.1007/s11120-010-9583-3.

[82] I. Sadovskaya, A. Souissi, S. Souissi, T. Grard, P. Lencel, C.M. Greene, S. Duin, P.S. Dmitrenok, A.O. Chizhov, A.S. Shashkov, A.I. Usov, Chemical structure and

biological activity of a highly branched (1 → 3,1 → 6)-β-D-glucan from *Isochrysis galbana*, Carbohydr. Polym. 111 (2014) 139–148, http://dx.doi.org/10.1016/j.carbpol.2014.04.077.

[83] H.-T. Wang, C.-H. Yao, J.-N. Ai, X.-P. Cao, S. Xue, W. Wang, Identification of carbohydrates as the major carbon sink of the marine microalga *Isochrysis zhangjiangensis* (Haptophyta) and optimization of its productivity by nitrogen manipulation, Bioresour. Technol. 171 (2014) 298–304, http://dx.doi.org/10.1016/j.biortech.2014.08.090.

[84] B.-H. Zhu, H.-P. Shi, G.-P. Yang, N.-N. Lv, M. Yang, K.-H. Pan, Silencing UDP-glucose pyrophosphorylase gene in *Phaeodactylum tricornutum* affects carbon allocation, New Biotechnol. 33 (2016) 237–244, http://dx.doi.org/10.1016/j.nbt.2015.06.003.

[85] A.E. Allen, C.L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J.P. McCrow, H. Zheng, D.A. Johnson, H. Hu, A.R. Fernie, C. Bowler, Evolution and metabolic significance of the urea cycle in photosynthetic diatoms, Nature 473 (2011) 203–207, http://dx.doi.org/10.1038/nature10074.

[86] S.D. Rokitta, L.J. de Nooijer, S. Trimborn, C. de Vargas, B. Rost, U. John, Transcriptome analyses reveal differential gene expression patterns between the life-cycle stages of *Emiliania huxleyi* (haptophyta) and reflect specialization to different ecological Niches1, J. Phycol. 47 (2011) 829–838, http://dx.doi.org/10.1111/j.1529-8817.2011.01014.x.

[87] K. Aoki, Y. Ogata, D. Shibata, Approaches for extracting practical information from gene co-expression networks in plant biology, Plant Cell Physiol. 48 (2007) 381–390, http://dx.doi.org/10.1093/pcp/pcm013.

[88] L.J.A. Kogelman, S. Cirera, D.V. Zhernakova, M. Fredholm, L. Franke, H.N. Kadarmideen, Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA sequencing in a porcine model, BMC Med. Genet. 7 (2014) 57, , http://dx.doi.org/10.1186/1755-8794-7-57.

[89] C.A. Hollender, C. Kang, O. Darwish, A. Geretz, B.F. Matthews, J. Slovin, N. Alkharouf, Z. Liu, Floral transcriptomes in woodland strawberry uncover developing receptacle and anther gene networks1[W][OPEN], Plant Physiol. 165 (2014) 1062–1075, http://dx.doi.org/10.1104/pp.114.237529.

[90] I. El-Sharkawy, D. Liang, K. Xu, Transcriptome analysis of an apple (Malus × domestica) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation, J. Exp. Bot. 66 (2015) 7359–7376, http://dx.doi.org/10.1093/jxb/erv433.

[91] R. Kalyani, J.-Y. Lee, H. Min, H. Yoon, M.H. Kim, Genes frequently coexpressed with Hoxc8 provide insight into the discovery of target genes, Mol. Cell 39 (2016) 395–402, http://dx.doi.org/10.14348/molcells.2016.2311.

[92] K. Righetti, J.L. Vu, S. Pelletier, B.L. Vu, E. Glaab, D. Lalanne, A. Pasha, R.V. Patel, N.J. Provart, J. Verdier, O. Leprince, J. Buitink, Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways, Plant Cell 27 (2015) 2692–2708, http://dx.doi.org/10.1105/tpc.15.00632.

[93] P. Heydarizadeh, J. Marchand, B. Chenais, M.R. Sabzalian, M. Zahedi, B. Moreau, B. Schoefs, Functional investigations in diatoms need more than a transcriptomic approach, Diatom Res. 29 (2014) 75–89, http://dx.doi.org/10.1080/0269249X.2014.883727.

[94] S. Arora, R. Rana, A. Chhabra, A. Jaiswal, V. Rani, miRNA–transcription factor interactions: a combinatorial regulation of gene expression, Mol. Gen. Genomics. 288 (2013) 77–87, http://dx.doi.org/10.1007/s00438-013-0734-z.

[95] A. Rogato, H. Richard, A. Sarazin, B. Voss, S. Cheminant Navarro, R. Champeimont, L. Navarro, A. Carbone, W.R. Hess, A. Falciatore, The diversity of small non-coding RNAs in the diatom *Phaeodactylum tricornutum*, BMC Genomics 15 (2014) 698, http://dx.doi.org/10.1186/1471-2164-15-698.

[96] A.D. Garst, A.L. Edwards, R.T. Batey, Riboswitches: structures and mechanisms, Cold Spring Harb. Perspect. Biol. 3 (2011), http://dx.doi.org/10.1101/cshperspect.a003533.

[97] T. Glisovic, J.L. Bachorik, J. Yong, G. Dreyfuss, RNA-binding proteins and post-transcriptional gene regulation, FEBS Lett. 582 (2008) 1977–1986, http://dx.doi.org/10.1016/j.febslet.2008.03.004.

[98] E. Szostak, F. Gebauer, Translational control by 3′-UTR-binding proteins, Brief. Funct. Genomics 12 (2013) 58–65, http://dx.doi.org/10.1093/bfgp/els056.

[99] H. Gingold, Y. Pilpel, Determinants of translation efficiency and accuracy, Mol. Syst. Biol. 7 (2011) 481, http://dx.doi.org/10.1038/msb.2011.14.

[100] T. Kouzarides, Acetylation: a regulatory modification to rival phosphorylation? EMBO J. 19 (2000) 1176–1179, http://dx.doi.org/10.1093/emboj/19.6.1176.

[101] A.H. Brivanlou, J.E. Darnell, Signal transduction and the control of gene expression, Science 295 (2002) 813–818, http://dx.doi.org/10.1126/science.1066355.

[102] R. Gao, A.M. Stock, Temporal hierarchy of gene expression mediated by transcription factor binding affinity and activation dynamics, MBio 6 (2015) e00686-00615, , http://dx.doi.org/10.1128/mBio.00686-15.

[103] R.C. Conaway, C.S. Brower, J.W. Conaway, Emerging roles of ubiquitin in transcription regulation, Science 296 (2002) 1254–1258, http://dx.doi.org/10.1126/science.1067466.

[104] G. Gill, Something about SUMO inhibits transcription, Curr. Opin. Genet. Dev. 15 (2005) 536–541, http://dx.doi.org/10.1016/j.gde.2005.07.004.

[105] W. Qian, X. Gang, T. Zhang, L. Wei, X. Yang, Z. Li, Y. Yang, L. Song, P. Wang, J. Peng, D. Cheng, Q. Xia, Protein kinase A-mediated phosphorylation of the Broad-Complex transcription factor in silkworm suppresses its transcriptional activity, J. Biol. Chem. 292 (2017) 12460–12470, http://dx.doi.org/10.1074/jbc.M117.775130.

[106] R. Soares-dos-Reis, A.S. Pessoa, M.R. Matos, M. Falcão, V.M. Mendes, B. Manadas, F.A. Monteiro, D. Lima, C. Reguenga, Ser[119] phosphorylation modulates the activity and conformation of PRRXL1, a homeodomain transcription factor, Biochem. J. 459 (2014) 441–453, http://dx.doi.org/10.1042/BJ20131014.

[107] N. Wase, P.N. Black, B.A. Stanley, C.C. DiRusso, Integrated quantitative analysis of nitrogen stress response in *Chlamydomonas reinhardtii* using metabolite and protein profiling, J. Proteome Res. 13 (2014) 1373–1396, http://dx.doi.org/10.1021/pr400952z.

[108] M. Aksoy, W. Pootakham, A.R. Grossman, Critical function of a *Chlamydomonas reinhardtii* putative polyphosphate polymerase subunit during nutrient deprivation, Plant Cell 26 (2014) 4214–4229, http://dx.doi.org/10.1105/tpc.114.129270.

[109] J. Valenzuela, A. Mazurie, R.P. Carlson, R. Gerlach, K.E. Cooksey, B.M. Peyton, M.W. Fields, Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricornutum*, Biotechnol. Biofuels 5 (2012) 40, http://dx.doi.org/10.1186/1754-6834-5-40.