D1.2 Measurements of transcripts, proteome and metabolite profiles

FUEL4ME PUBLIC

DOCUMENT FILENAME:	FUEL4ME D1.2
VERSION:	2
STATUS:	FINAL
REVISION DATE:	20/02/2017
READERSHIP:	Public
Responsible/Author:	Dr. ir. Sander Peters Wageningen Plant Research

FUEL4ME project: Grant Agreement Nº 308983 is co-funded by the European Commission, within the 7th Framework Programme, theme ENERGY 2012.3.2.1 Biofuels from microalgae and macroalgae.

Content

Executive summary

The work described for this deliverable was carried out by DLO-PRI, in collaboration with WU, and aimed to understanding lipid production in *Phaeodactylum tricornutum*. By means of transcriptome, proteome and metabolome analyses we aimed to provide insight into the one step lipid production performed by partner WU. Cultivation conditions were chosen to steer lipid profile towards most suitable composition for biofuel production. This work was divided in the following steps;

• Genome annotation for *Phaeodactylum* with specific focus on gene functions known to play a role in synthesis, cellular transport, and storage of lipids

• Transcriptome analysis of *Phaeodactylum* grown under various growth conditions and qualitative and quantitative analysis of gene expression. Expression of the abovementioned genes under specific growth conditions by quantitative transcriptomics to identify key components with high levels of induction under conditions of high TAG accumulation.

• Qualitative and quantitative analysis of both polar compounds and lipids in *Phaeodactylum* (targeted and non-targeted metabolomics).

- GO term enrichment of differentially expressed genes. This bioinformatics analysis is aimed to find possible functions involved in lipid biosynthesis *of P. tricornutum*.

Description of work

Tool development

DLO-PRI has developed a multi ~omics pathway search tool (MOPS). This generic tool maps transcriptome, metabolome, and proteome data to pathways (from e.g. KEGG) including those involved in lipid biosynthesis. By identifying the transcripts, proteins and metabolites that are associated with lipid production, and evaluating the coverage for a particular pathway a probability score can be assigned, describing the likelihood for the pathway to be involved in lipid production in *P. tricornutum*. Currently, we have selected 6 KEGG pathways for transcriptome profiling;

- Fatty acid biosynthesis elongation-Pti00061
- Glycerolipid metabolism-Pti00561
- Biosynthesis of unsaturated fatty acids-Pti01040
- Fatty acid degradation-Pti00071
- Glycerophospholipid metabolism-Pti00564

The multi omics mapping can provide further understanding of genes and their regulation in response to environmental factors affecting lipid biosynthesis in *P. tricornutum.* Furthermore, by visualizing the combination of ~omics data it is possible to find leads to manipulate lipid production in *P. tricornutum*.

Transcriptomics profiling on test samples

We have successfully implemented methods and protocols for transcriptome profiling of N-depleted and N-repleted *Phaeodactylum* samples from partner WU. Mapping of the transcriptome, established with RNAseq on polA+ RNA against the annotated genome sequence from *P. tricornutum* revealed that 10213 genes out of 10391 annotated genes were expressed, indicating the sufficient coverage of transcriptome. We have analysed expression levels of genes by mapping transcriptome data against gene sequences from known lipid biosynthesis (KEGG, DiatomCyc) pathways. This approach enables the identification of differentially expressed genes underlying lipid biosynthesis in *P. tricornutum* **(DLO-PRI)**.

Transcriptome sequencing and profiling on N-depleted and N-repleted samples

A series of *P. tricornutum* samples (one-step process, duplicate experiments, both nitrogen replete and nitrogen limited; see task 1.2, objective 8) have been processed for transcriptomics analyses. RNA extractions and library construction has been completed. Transcriptome sequencing using Illumina RNS seq has been completed **(DLO-PRI)**. The, differential expression analysis of all samples has been completed (deliverable 1.2).

Transcriptome mapping has been carried out against the *P. tricornutum* genome. We have used the ENSEMBL release 29 version to update the annotation for the *P. tricornutum* genome All ~omics data were assessed for differential expression using a Students T-test and FDR corrected p-value < 0.01 and fold change of |2.5|. Genes, proteins and metabolites that have a statistical relevant differential expression were identified, as far as was possible from available databases.

Proteomics

We have tested and evaluated several protein extraction procedures from intact *Phaeodactylum* cells with the aim of obtaining a maximum of protein identifications plus an optimal quantitative reproducibility of the extraction efficiency. 3 grinding methods and 2 extractions buffers have been tested. In conclusion, grinding wet sample gives maximum identifications. The 2 buffers both give high yield, with some complementarity. We are ready to perform detailed quantitative analysis on sample series **(DLO-PRI)**.

The extraction and LC-MS analysis of 2*6 replicate *P. tricornutum* nitrogen replete (N+) and nitrogen limited (N-) samples (using 3 different extraction protocols) have been performed. Protein identification has detected a total of 4034 protein groups, comprised of 47345 identified peptides. For the majority of these peptides quantification data have been obtained. Principal component analysis of the global dataset revealed that the extraction procedure has a greater impact than the difference between N+ and N- samples. However, filtering on ANOVA differential peptides revealed a clear and significant difference between N+ and N- samples. Multiple proteins have been detected that appear significantly differential between the two growth conditions. However, no multiple time point series (from the one-step *P. tricornutum* samples provided by partner WU; see task 1.2, objective 8) have been measured yet. This is planned to occur in Q3 of 2015.

In addition, mapping of the identified proteins on KEGG pathways from *Phaeodactylum* has demonstrated a nearly complete mapping of the fatty acid biosynthesis pathway in addition to a majority of pathways being mapped by identified proteins **(DLO-PRI).**

Mapping transcriptome and proteome data against KEG pathways

Differentially expressed features were subsequently referenced against KEGG pathways to generate biological meaning. For the matching of differentially expressed genes, proteins and metabolites against KEGG pathways a new tool, MapOmics, has been developed that is now operational in the WUR Galaxy server environment (figure 1). The output of the MapOmics tool also visualizes the coverage of specific metabolic pathways for which we have found differential features (genes, proteins and/or metabolites). We found 12179 genes and 2500 proteins, of which 451 genes and 140 proteins were significantly expressed (with a fold change >2) under lipid enhanced conditions. We could map the genes and proteins in KEGG using MAPOmics, and we found differential expression of genes and proteins that relate to key biological pathways, among them nitrogen metabolism (00910), fatty acid biosynthesis (00061), fatty acid metabolism (01212), and arginine biosynthesis (00220).

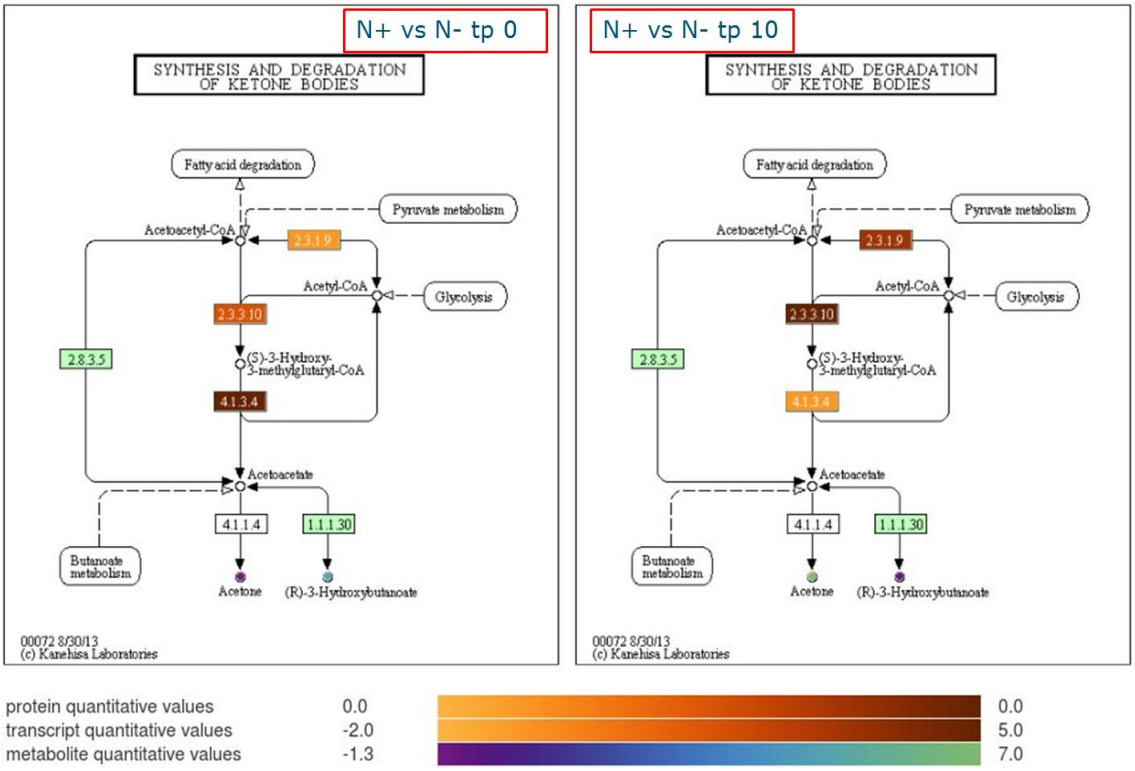

Figure 1. Pathway quantification and visualization by Mapomics. Expression of transcripts, proteins and metabolites in *Phaeodactylum* are compared between nitrogen replete (N+) and deplete (N-) growth conditions at different time points. Log fold changes as indicated in the legend below the plots are shown as colored heat maps onto the nodes of specific KEGG pathways.

Metabolomics

Methods for comprehensive GCMS and LCMS-based metabolomics analyses of both polar and apolar extracts from *Phaeodactylum* have been set up and successfully tested using test samples from N-depleted and N-repleted cultures. With either platform hundreds of metabolites, both known and yet unknown, can be detected and samples can be compared for their relative abundance of each metabolite. Both platforms have subsequently been used for analysing a series of 60 *Phaeodactylum* samples from partner WU. A total of 174 polar and 605 apolar compounds could be detected in these samples and their relative intensities across samples were compared in relation to steady state growth under N-repleted and depleted conditions and during their day/night cycle. The ability to directly map these compounds to KEGG pathways is however limited, as most of the N-limited/N-replete differential metabolites are not specified in KEGG or yet unknown. The differential polar compounds that could be annotated were mainly amino acids (decrease at N-limitation) and carbohydrates (increase at N-limitation). Amongst the most increased apolar metabolites we identified a large series of

TAGs, amongst which were EPA-containing TAG species, oxidized EPA and DMSP, while several membrane lipids were decreased in N-limited algae as compared to N-replete algae (**WPR and WU**).

GO term enrichment of differentially expressed genes

To identify GO terms associated with the differentially expressed genes, the proteomics data sets were annotated using ontology tools, including an Interpro analysis. The Gene ontology database was loaded in a local database and the expression data was linked to the GO terms in this database. The up and down regulated genes were counted for each GO term to determine GO term enrichment for the differentially expressed genes. Figure 2 shows, for example, the GO term enrichment of the main category 'Biological Process'

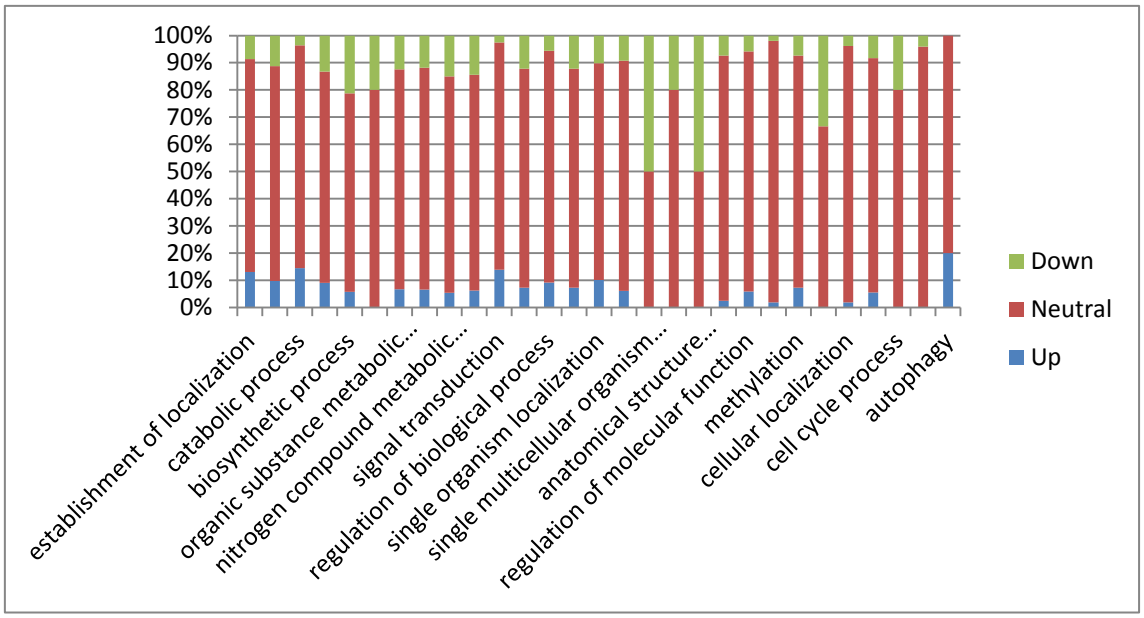

Figure 2: Go term enrichment for the main category 'Biological Process' for time point zero. Two levels down for one time point. In green the percentage of genes associated with that particular GO term which are down regulated, in red the genes which show no regulator effect and in blue the up regulated genes.