International Journal of Environmental Engineering– IJEE
Volume 2 : Issue 1 [ISSN : 2374-1724]
Publication Date: 30 April, 2015

Response to Oxidative Stress Enables Fistulifera solaris to Efficiently Produce Biofuel

Pui Shan Wong, Michihiro Tanaka, Yoshihiko Sunaga, Yoshiaki Maeda, Masayoshi Tanaka, Takeaki Taniguchi, Tomoko Yoshino, Tsuyoshi Tanaka and Sachiyo Aburatani

***Abstract*—Biofuel can be produced sustainably by oleaginous microalgae when they are grown in nitrogen limited conditions. *Fistulifera solaris* JPCC DA0580 demonstrates higher lipid accumulation as a percentage of cell weight than other diatoms in nitrogen limited conditions. These features make it a solid candidate for biofuel production so we compared the gene expression of *F. solaris* with its close relative and a model diatom, *Phaeodactylum tricornutum*. The gene expression of the two diatoms cultured in low nitrogen conditions were used to calculate the difference in fold change expression. We selected genes exhibiting significantly different fold change differences and divided them into groups based on the direction of the fold change. These groups were characterised by the gene ontologies of known genes within their respective groups and the degree of fold change differences between the two diatoms in each group. The gene ontologies and fold change differences helped to highlight the differences in the way *P. tricornutum* adjusted to the low nitrogen conditions compared with *F. solaris*. Our results indicate that *F. solaris* maintains the expression of genes related to oxidative stress response. In contrast, the up-regulated *P. tricornutum* genes were related to nitrogen metabolism, and energy metabolism.**

***Keywords*— *Fistulifera solaris*, diatom, biofuel**

I. Introduction

The search for sustainable and efficient fuel sources is experiencing rapid growth as research into climate change progresses. In biology, efforts are focused on biofuel compounds produced by photosynthetic organisms thanks to their renewable characteristics. There have been advances in research using food crops and organic waste however there are issues with their dependency on local climate, as well as competition with food production. Oleaginous microalgae circumvent these issues as they can be farmed indoors and are not in direct competition with food crops for land space. They produce biofuel compounds in low nitrogen conditions so that, combined with their high rates of CO2 fixation and large biomass yields of neutral lipids, they are some of the most promising biofuel candidate organisms [1] [2]. Some of these candidates are heterokonts, chlorophytes, dinoflagellates, haptophytes and rhodophytes [3] [4] [5] [6]. The pennate diatom, *Phaeodactylum tricornutum*, is one of the main model organisms and is studied for its ease of transformation and annotated genome information [7] [8] [9].

We have focused our study on to the diatom *Fistulifera solaris* JPCC DA0580. It has an important characteristic required for efficient biofuel production. It can reach a high neutral lipid content of 40% to 60%, w/w [10] [11] [12]. The lipids produced by *F. solaris* are mainly composed of methyl palmitate (C16:0) and methyl palmitoleate (C16:1), both of which can be used as biodiesel fuel [13]. The unique characteristics of *F. solaris* make it a favourable candidate as a biofuel source over other oleaginous microalgae such as *Nannochloropsis* sp. and *Neochloris oleoabundans* [5] [14].

There are 20,470 genes in the *F. solaris* genome compared to the 10,402 genes in *P. tricornutum* and 11,776 genes in another diatom, *Thalassiosira pseudonana* [9] [15]. It has been observed in former investigations that some of the *F. solaris* genes are duplicates of highly conserved but not identical sequences. A majority of *F. solaris* genes are homologous with its closest relative, *P. tricornutum*, including genes that are known to affect lipid accumulation such as those in photosystem I and II [16]. Additionally, the lipid metabolism

Pui Shan Wong, Michihiro Tanaka, Sachiyo Aburatani
Computational Biology Research Center, National Institute of AIST
Tokyo, Japan

Michihiro Tanaka
Center for iPS Research and Application, Kyoto University
Kyoto, Japan

Takeaki Taniguchi
Mitsubishi Research Institute, Inc.
Tokyo, Japan

Yoshihiko Sunaga, Yoshiaki Maeda, Masayoshi Tanaka, Tomoko Yoshino, Tsuyoshi Tanaka
Institute of Engineering, Tokyo University of Agriculture and Technology
Tokyo, Japan

Yoshihiko Sunaga, Tomoko Yoshino, Tsuyoshi Tanaka
Japan Science and Technology Agency, CREST
Tokyo, Japan

SEEK DIGITAL LIBRARY

International Journal of Environmental Engineering– IJEE
Volume 2 : Issue 1 [ISSN : 2374-1724]
Publication Date: 30 April, 2015

of *P. tricornutum* and other microalgae share many of the same genes and pathways [9] [17]. The similarity between *F. solaris* and *P. tricornutum* has been used in comparative studies to help determine putative pathways for fatty acid desaturation in *F. solaris* showing that the underlying mechanisms for lipid accumulation is similar between the two diatoms [18]. There are, however, differences in lipid accumulation between *F. solaris* and *P. tricornutum*, suggesting that lipid metabolism genes are being expressed but under the effect of different regulation mechanisms.

By comparing the expression data of *F. solaris* and *P. tricornutum* as they are cultured in low nitrogen conditions, we can identify genes that contribute to the different levels of expression between the two diatoms. Those genes can be sorted into groups which can then be characterized more generally to describe the genes contained within them. The identification and examination of these groups can give us an insight into the patterns of gene expression of *F. solaris*.

II. Results

A. Comparative Genomics

We observed that 13,898 out of 20,470 *F. solaris* genes were most similar to *P. tricornutum* genes (68%) (Fig. 1). This was the largest majority and was followed by unmatched or novel genes, and *T. pseudonana* genes. The remaining genes matched those from other algae and microorganisms. As the majority of the matches were *P. tricornutum* genes, the 13,898 *F. solaris* genes were marked as homologs with the matching 2,817 *P. tricornutum* genes.

An exploratory analysis was carried out to examine the difference in gene expression of *F. solaris* and *P. tricornutum* grown in low nitrogen and control conditions. We examined the difference in gene expression across species, condition and homology using an unweighted ANOVA and found a significant interaction between species and homology (p-value < 0.05). We looked at the differences between the means of species and homology in more detail and found no statistical significance in gene expression between the homologous genes of both diatoms (p-value < 0.05), and homologous *F. solaris* genes and nonhomologous *P. tricornutum* genes (Fig. 2).

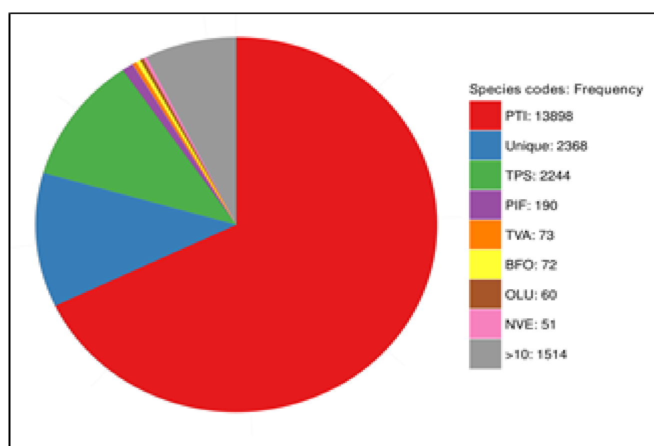

Figure 1. The species assignment of 20,470 *F. solaris* genes from searching the KEGG database. The species codes corresponds to the following species, in order from top to bottom: *Phaeodactylum tricornutum* (PTI), *Thalassiosira pseudonana* (TPS), *Phytophthora infestans* (PIF), *Trichomonas vaginalis* (TVA), *Branchiostoma floridae* (BFO), *Ostreococcus lucimarinus* (OLU) and *Nematostella vectensis* (NVE). Species with less than 10 matched genes were consolidated into one group for clarity.

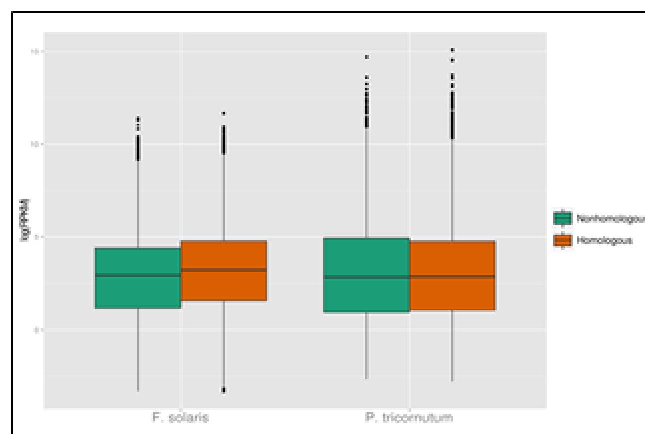

Figure 2. A comparison of expression data from *F. solaris* and *P. tricornutum* separated by experiment condition and homology. An unweighted ANOVA showed that there was a significant interaction effect between species and homology (p-value < 0.05). A pairwise comparison showed that there was no significant difference between the homologous genes of both diatoms, and no significant difference between homologous *F. solaris* genes and nonhomologous *P. tricornutum* genes (p-value < 0.05).

B. Significance Test

We examined the difference in fold change between homologous *F. solaris* genes and *P. tricornutum* genes to identify homologous genes exhibiting a marked difference in fold change in *F. solaris* compared to *P. tricornutum*. From our analysis, we identified 194 *F. solaris* genes which are homologous with 91 *P. tricornutum* genes. The genes were categorised into four groups according to the direction of the fold change of the genes in *F. solaris* and *P. tricornutum*. Those with positive fold change in both diatoms were placed in group 1, those with positive fold change in *F. solaris* and negative fold change in *P. tricornutum* were placed in group 2, those with negative fold change in *F. solaris* and positive fold change in *P. tricornutum* were placed in group 3, and those with negative fold change in both diatoms were placed in group 4. The largest group was group 3 followed by group 2, 4 and 1.

The gene expressions of each group were plotted to inspect the degree of differences in fold change (Fig. 3). The expressions of genes in group 1 were distinguished by a larger fold change in *P. tricornutum* compared to *F. solaris*. The *P. tricornutum* genes also had a smaller range of gene expression than *F. solaris* genes, although this could have been attributed to the smaller number of *P. tricornutum* genes. The most relevant group to our interest was group 2 as it was composed of up-regulated *F. solaris* genes and down-regulated

SEEK DIGITAL LIBRARY

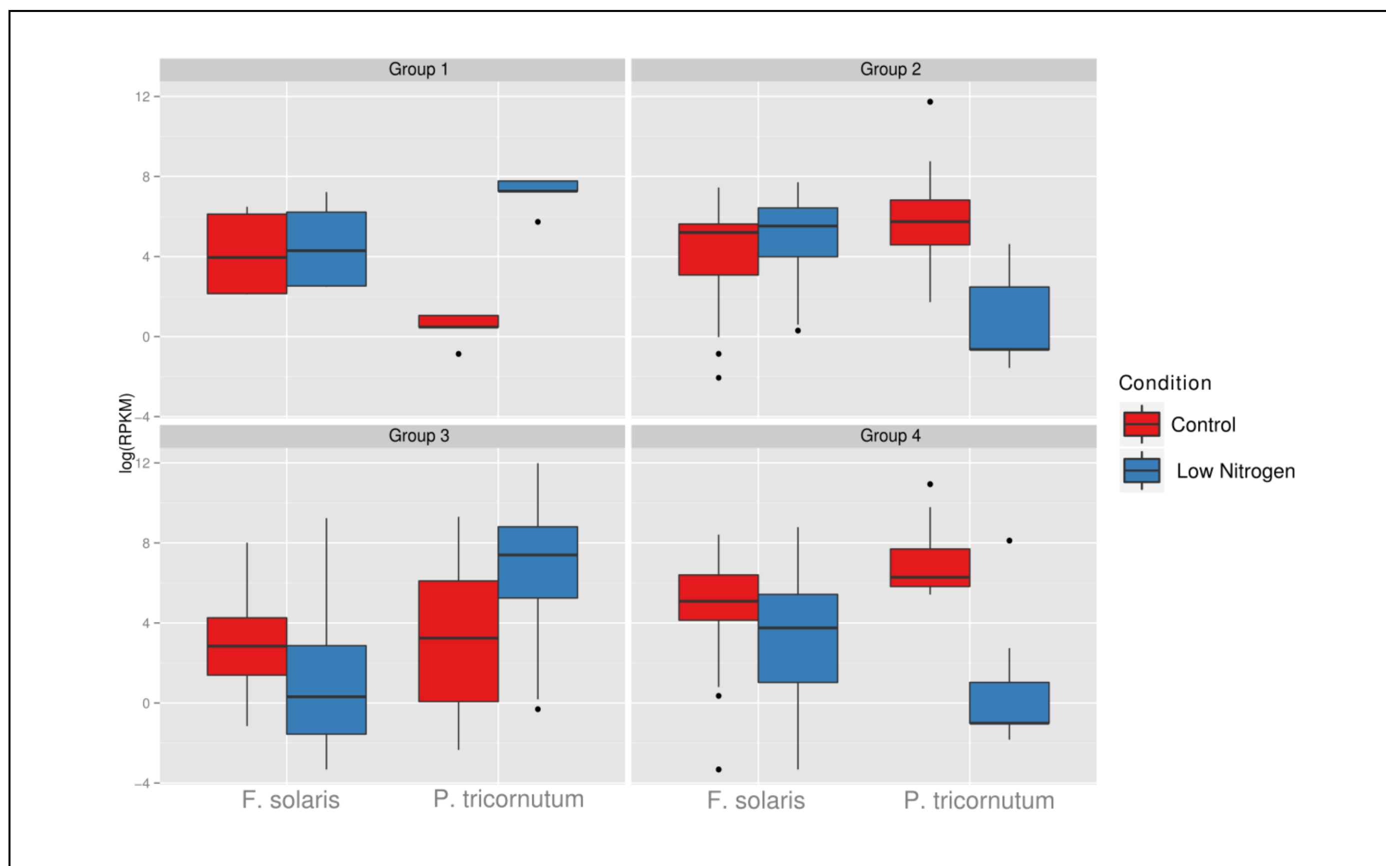

Figure 3. The expression data of the 194 significant *F. solaris* genes and 91 *P. tricornutum* genes. Group 1 contains genes that were up-regulated in both diatoms and has 6 *F. solaris* genes and 3 *P. tricornutum* genes. Group 2 contains genes that were up-regulated in *F. solaris* and down-regulated in *P. tricornutum* and has 35 *F. solaris* genes and 16 *P. tricornutum* genes. Group 3 contains genes that were down-regulated in *F. solaris* and up-regulated in *P. tricornutum* and has 121 *F. solaris* genes and 57 *P. tricornutum* genes. Group 4 contains genes that were down-regulated in both diatoms and has 32 *F. solaris* genes and 15 *P.tricornutum* genes.

P. tricornutum genes. Here, the absolute fold change of *P. tricornutum* genes were larger than *F. solaris* genes but in the opposite direction to group 1. Additionally, the mean gene expressions of *F. solaris* genes and the mean control expression of *P. tricornutum* genes were more similar than the mean treatment expression of *P. tricornutum*. In contrast to group 2, group 3 consisted of down-regulated *F. solaris* genes and up-regulated *P. tricornutum* genes. Although the mean fold change directions are reversed, the degrees of differences are similar in value. The gene expressions in group 4 were similar to group 1 where the *P. tricornutum* fold change values were larger than *F. solaris* fold change values. However, the gene expressions for both diatoms had a larger range than those in group 1.

C. *Characterization using Gene Ontology*

Each group was characterised by the over-represented gene ontologies found by testing the terms using the hypergeometric test (Table I).

Group 1 contained 6 *F. solaris* genes and 3 *P. tricornutum* genes that shared 6 gene ontologies between them. The significant over-represented gene ontologies that were identified were 4-aminobutyrate transaminase activity, pyridoxal phosphate binding, transferase activity (transferring nitrogenous groups), amino acid transport, organic acid transport and organic anion transport (Table I).

Group 2 was larger than Group 1 and contained 35 *F. solaris* genes and 16 *P. tricornutum* genes represented by 20 gene ontologies. The hypergeometric test narrowed the list of gene ontologes down to 8 terms (Table I). In general, they are related to cAMP-dependent protein kinases for regulating glycogen, sugar and lipid metabolism, hydrolase activity, isomerase activity, GTP cyclohydrolase activity, protein phosphorylation and transferase complex.

Group 3 was the largest group with 121 *F. solaris* genes and 57 *P. tricornutum* genes and it had the highest number of representative gene ontologies at 94 terms. As a result, some of the significant gene ontologies consisted of broad terms such as localization and membrane. The full list contained seven types of dehydrogenase activities, three reductase activities and three transmembrane transporter activities (Table I). There were four specific singular terms, N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase activity, methylcrotonoyl-CoA carboxylase activity, oxygen-dependent protoporphyrinogen oxidase activity, phytochromobilin:ferredoxin oxidoreductase activity, and

SEEK DIGITAL LIBRARY

TABLE I. SIGNIFICANT GENE ONTOLOGY TERMS BY GROUP

Group	Gene Ontology	Corrected P-value
1	4-Aminobutyrate transaminase activity	3.86×10^{-5}
1	Pyridoxal phosphate binding	4.25×10^{-4}
1	Transferase activity, transferring nitrogenous groups	9.941×10^{-3}
1	Amino acid transport	0.0133
1	Organic acid transport	0.0157
1	Organic anion transport	0.0197
2	GTP cyclohydrase I activity	6.76×10^{-4}
2	cAMP-dependent protein kinase complex	9.82×10^{-4}
2	cAMP-dependent protein kinase regulator activity	2.59×10^{-3}
2	Phosphomannomutase activity	4.04×10^{-3}
2	Cyclohydrolase activity	4.04×10^{-3}
2	Transferase complex	0.0120
2	Protein phosphorylation	0.0130
2	Kinase regulator activity	0.0346
3	Epoxide dehydrogenase activity	4.21×10^{-5}
3	5-Exo-hydroxycamphor dehydrogenase activity	4.21×10^{-5}
3	2-Hydroxytetrahydrofuran dehydrogenase activity	4.21×10^{-5}
3	Mevaldate reductase activity	4.21×10^{-5}
3	3-Keto sterol reductase activity	4.21×10^{-5}
3	3-Ketoglucose-reductase activity	4.21×10^{-5}
3	Membrane	1.16×10^{-4}
3	Gluconate dehydrogenase activity	1.25×10^{-4}
3	C-3 sterol dehydrogenase (C-4 sterol decarboxylase) activity	1.25×10^{-4}
3	Isocitrate dehydrogenase activity	1.25×10^{-4}
3	Ammonium transmembrane transporter activity	2.90×10^{-4}
3	Steroid dehydrogenase activity	1.04×10^{-3}
3	Transport	1.60×10^{-3}
3	Localization	1.91×10^{-3}
3	Anion transmembrane transporter activity	6.48×10^{-3}
3	Lactate transmembrane transporter activity	8.53×10^{-3}
3	Lactate transport	8.53×10^{-3}
3	N2-acetyl-L-ornithine:2-oxoglutarate 5-aminotransferase activity	8.53×10^{-3}
3	Steroid binding	8.53×10^{-3}
3	Methylcrotonoyl-CoA carboxylase activity	8.53×10^{-3}
3	Oxygen-dependent protoporphyrinogen oxidase activity	8.53×10^{-3}
3	Phytochromobilin:ferredoxin oxidoreductase activity	8.53×10^{-3}
3	Tetrapyrrole metabolic process	0.0121
4	L-ascorbate peroxidase activity	7.53×10^{-10}
4	Oxidoreductase activity, acting on peroxide as acceptor	2.01×10^{-5}
4	Antioxidant activity	5.38×10^{-5}
4	Response to oxidative stress	3.20×10^{-4}
4	Photosynthesis, light harvesting	7.60×10^{-4}
4	Photosynthesis	1.40×10^{-3}
4	Inorganic anion exchanger activity	2.12×10^{-3}

The gene ontology terms were selected from 194 *F. solaris* genes and 91 *P. tricornutum* genes. The p-values were corrected for multiple testing using Bonferroni's correction method and gene ontologies with p -value < 0.05 were marked to represent their respective group. The group numbers correspond to those shown in Fig. 3.

three broader terms, steroid binding, lactate transport and tetrapyrrole metabolic process. Also of note are the presence of terms related to mitochondria and chlorophyll in this group.

Group 4 was smaller than groups 1 and 2 and was made up of 32 *F. solaris* genes and 15 *P. tricornutum* gene. Altogether there were 16 gene ontologies between them. The significant gene ontologies were L-ascorbate peroxidase activity, oxidoreductase activity (acting on peroxide as acceptor), antioxidant activity, inorganic anion exchanger activity, response to oxidative stress, photosynthesis (light harvesting) and photosynthesis (Table I). A few terms were related to

SEEK DIGITAL LIBRARY

oxidative stress and photosynthesis while others were related to transporter and peroxidase activity.

III. Discussion

We compared the gene expression of *F. solaris* and *P. tricornutum* grown in low nitrogen and control conditions, and found that the expressions of homologous genes are generally similar across species. Surprisingly, the gene expression of homologous *F. solaris* genes are more similar to the gene expression of *P. tricornutum* genes than its own nonhomologous genes. It indicates that some of the nonhomologous *F. solaris* genes may have originated from another diatom such as *T. pseudonana*, and that they were most likely influenced by different regulatory mechanisms than *P. tricornutum* genes. Some of those genes would have also included the novel genes as previously identified in Fig 1. While those genes may hold a key component responsible for the difference in phenotype, their functions would require lengthy and expensive transformation experiments for the large number of genes.

In order to understand the metabolic differences between *F. solaris* and *P. tricornutum*, we concentrated on identifying genes that showed different responses to the low nitrogen and control conditions between the two diatoms. The gene expression and gene ontologies of these genes can be used to characterise the differences and aid in interpretation. Using our analysis, we identified 194 *F. solaris* genes and assigned them to groups based on their fold change values in *F. solaris* and *P. tricornutum*.

The first group of genes that we were interested in are in Group 2 where the genes were up-regulated in *F. solaris* and down-regulated in *P. tricornutum*. Some of the gene ontologies of the genes in this group were related to cell regulation such as cAMP-dependent protein kinase complex and protein phosphorylation. The other distinguishing gene ontologies were related to hydrolase, isomerase and transferase activity through parent terms like GTP cyclohydrolase I activity and phosphomannomutase activity. The degree of difference between the control and treatment gene expression was important to note as well. The genes in this group showed a large amount down-regulation in *P. tricornutum* compared to negligible levels of up-regulation in *F. solaris*. It indicates that the difference in phenotype could be attributed to the down-regulation of these processes in *P. tricornutum* rather than the up-regulation in *F. solaris*. There were five *F. solaris* genes with the most pronounced difference in fold change between the two diatoms. These were fso:g2859, fso:g2860, fso:g12378, fso:g9753 and fso:g9752, which were homologous with the *P. tricornutum* gene estExt_Phatr1_ua_kg.C_chr_70081. Although there were no gene ontology annotations, the *P. tricornutum* homolog is noted for being up-regulated in response to iron deficiency and is thought to be part of a secretory pathway [9]. The gene, fso:g9752, also contains a domain found in the *Chlamydomonas reinhardtii* gene, FEA1. It seems to be required for growth in iron-deficient conditions as it was being up-regulated and secreted in *C. reinhardtii*, another oleaginous alga [19].

The genes in group 3 were in the largest group where the genes were down-regulated in *F. solaris* and up-regulated in *P. tricornutum*. There were many gene ontologies representing this group and they can be broadly divided into stress response in low nitrogen conditions, energy management, and intracellular transport and localization. While the up-regulation of these genes is expected in low nitrogen condition, it is important to note that these genes were down-regulated in *F. solaris*. The degree of down-regulation is comparatively small it is the largest for *F. solaris* in all 4 groups. The difference between the two diatoms is distinct due to the treatment means as the control means are quite similar. The large difference in gene expression between *F. solaris* and *P. tricornutum* can be separated into two patterns; the fold change is large in *P. tricornutum* compared to *F. solaris*, or small in *P. tricornutum* compared to *F. solaris*. Among the genes with large fold change differences, there were five *F. solaris* genes with small fold changes compared to the homologous three *P. tricornutum* genes. These were fso:g11687, fso:g15667, fso:g5409, fso:g17137 and fso:g20250 in *F. solaris*, and fgenesh1_pg.C_chr_21000194, estExt_fgenesh1_pg.C_chr_10568 and estExt_fgenesh1_pg.C_chr_30465 in *P. tricornutum*. The *P. tricornutum* gene, estExt fgenesh1 pg.C chr 30465, has been annotated with the calcium ion binding gene ontology due to the presence of a calcium-binding domain but the homologous *F. solaris* genes also contain a domain associated with 5'-Nucleotidase/apyrase. It is a family of enzymes that catalyses the hydrolysis of nucleotide molecules that is present in many species and whose metabolic importance depends on its cellular location. For the second pattern where there is a large fold change in *F. solaris* compared to *P. tricornutum*, we found two *F. solaris* genes that were homologous to one *P. tricornutum* gene. These genes were fso:g9841 and fso:g6241, and the homologous *P. tricornutum* gene was e_gw1.4.342.1. The *P. tricornutum* gene is described by two gene ontologies related to photosynthesis; oxidoreductase activity and phytochromobilin biosynthesis. Additionally, the *P. tricornutum* and *F. solaris* sequences contain domains associated with ferredoxin-dependent bilin reductase that synthesis phytobilin from heme.

The last two groups of genes represent genes whose fold changes were in the same direction but the degree of fold change is very different between *F. solaris* and *P. tricornutum.* In both groups, the degree of fold change is small in *F. solaris* and the degree of fold change is large in *P. tricornutum*. In Group 1 where the genes were up regulated in both diatoms, these genes were associated with metabolising amino acids to keto acids. However the small size of this group may have affected the selection of gene ontologies. The last group, Group 4, is of a better size and contains genes that were down-regulated in both diatoms. The gene ontologies for this group were mainly associated with response to oxidative stress as well as photosynthesis. Although the gene expression was down-regulated in both organisms, the fold change in *F. solaris* is much smaller than in *P. tricornutum*. In particular, the four genes with the largest difference in fold change that was annotated was attributed the response to oxidative stress gene ontology. The genes were fso:g10215, fso:g4650, fso:g7681 and fso:g19255, and their homolog was

SEEK DIGITAL LIBRARY

International Journal of Environmental Engineering– IJEE
Volume 2 : Issue 1 [ISSN : 2374-1724]
Publication Date: 30 April, 2015

estExt_fgenesh1_pg.C_chr_130213 in *P. tricornutum*. Interestingly, the gene with the greatest difference in fold change in this group was fso:g18615 or estExt_fgenesh1_pg.C_chr_140189 in *P. tricornutum* [9]. They contain a ferritin-related domain with similar descriptions to the domain in FEA1 in *C. reinhardtii* as described for fso:g9752 in Group 2. Upon closer inspection, the F. solaris fold change for fso:g18615 is so small, it is quite possible that it was assigned incorrectly due to biological variation.

In a further investigation, we looked at the KEGG pathway of the annotated genes in each group in an attempt to find underlying metabolic processes in the results. From the list of significant genes identified in our analysis, 12 *F. solaris* genes had a KEGG annotation belonging in a metabolic pathway. These genes matched five KEGG orthologs that are present in 8 pathways however we excluded the highest level pathway called metabolic pathway. Importantly, the remaining pathways were closely associated with lipid metabolism. There were two *F. solaris* genes from group 3 that matched the PPOX KEGG ortholog in the porphyrin and chlorophyll metabolism pathway. This gene is involved in the biosynthesis of cholorophyll and is present as two isoenzymes in plants [20]. Another KEGG ortholog, L-ascorbate peroxidase, was matched by four *F. solaris* genes in group 4. This ortholog is active in antioxidant activities in glutathione metabolism, and ascorbate and aldarate metabolism pathways. Specifically, it is in a reversible reaction that metabolises ascorbate into dehydroascorbate during glutathione metabolism and it undergoes the opposite reaction in ascorbate and aldarate metabolism [21]. The next KEGG ortholog was matched by two F. solaris genes in group 2. This ortholog is the GCH1 gene which is closely involved in THF tetrahydrofolate synthesis that is needed in the metabolism of amino acids and nucleic acids [22]. Similarly, the riboflavin metabolism pathway was also represented by the ACP KEGG ortholog that was matched by two F. solaris but these were in group 3. This gene takes part in a reaction that metabolises cardiolipin into a fatty acid [23]. The last pathway was glycerophospholipid metabolism which was represented by the CLD1 KEGG ortholog that matched two *F. solaris* genes in group 3. The CLD1 gene is part of a reaction that metabolises FMN into riboflavin [24]. Although it was unfortunate that there were only 12 annotated genes, they turned out to contain a lot of relevant information.

IV. Conclusion

We were able to find and identify the functions that are important in explaining the difference in gene expression between *F. solaris* and *P. tricornutum* while they were grown in low nitrogen and control conditions. The functions were identified by using gene ontologies and KEGG pathways so that we were able to characterise the differences more specifically. These functions and the differences in the degree of fold change indicate that some of the characteristics of *F. solaris* include being able to better adjust to low nitrogen conditions by keeping some genes expressed compared to *P. tricornutum* where they were down-regulated.

Genes of interest were identified without full annotations of *F. solaris* by relying on the similarity of *F. solaris* to the model organism, *P. tricornutum*. This is an important step in the analysis of new organisms where this situation is common. We were able to use the gene ontology annotations of *P. tricornutum* to describe the groups of *F. solaris* genes which led to better understanding of their differences. The use of absolute cut-offs for categorising may be improved upon by excluding or reassigning genes that were very close to the threshold. The interpretation of some of the groups was also difficult with the presence of unannotated genes or differences in annotations between *F. solaris* genes and the homologous *P. tricornutum* genes. However, the characterization of each group can still be used to select genes of interest for future study and annotation.

V. Method

A. Comparative Genomics

The sequences of *F. solaris* were searched against all the sequences of each species in the KEGG database using SSEARCH with the MIQS substitution matrix [25]. The species of the top matching sequences with an E-value of less than 0.0001 were taken and collected. The *F. solaris* sequences whose top match was from *P. tricornutum* were set aside and were assigned as homologs to the matching *P. tricornutum* sequence.

B. Expression Data

F. solaris was cultured in a control and treatment substrate. The treatment substrate was artificial sea water and the control substrate was a 10 fold dilution of the artificial sea water [13]. The RNA-Seq data from the cultures were obtained at time points 0, 24, 48 and 60 hours after being introduced to the substrates. Expression data taken at 60 hours were selected for analysis as determined from previous investigations [11]. The expression data for *P. tricornutum* were obtained from samples cultured in control and treatment substrates taken at 48 hours after being introduced to the substrates [26]. We compared the expression data of the two diatoms using species, condition, and homology as factors, and utilised the lsmeans package in the statistics program R [27] [28] in order to compute linear combinations of more than one mean.

The RNA-Seq data from *F. solaris* and *P. tricornutum* were normalized to decrease the effect of extreme RPKM values on fold change values. This is necessary as RNA-Seq is a competitive profiling method such that extreme values can distort the analysis. The RPKM values were corrected using the sRAP R package [27] [29] with the recommended threshold of 0.1 for minimizing the influence of the number of reads. The RPKM values are also log transformed during the normalization process.

The expression data was filtered to only include homologous sequences from each species such that there were 13,898 *F. solaris* entries and 6,589 *P. tricornutum* entries. Homologous *F. solaris* sequences that were best matched by the same *P. tricornutum* sequence had their RPKM values

SEEK DIGITAL LIBRARY

averaged as shown in (1). The total number of expression values totalled 13,178 genes, 6,589 from *F. solaris* and 6,589 from *P. tricornutum*.

$$F_x = \frac{\sum_{\forall i} s_i}{n} \quad (1)$$

where F_x is the average fold change for homologous *F. solaris* sequences that were best matched by *P. tricornutum* sequence x, i is the ith *F. solaris* sequence that is best matched by *P. tricornutum* sequence x and n is the number of *F. solaris* sequences that were best matched by *P. tricornutum* sequence x.

The fold change values of *F. solaris* and *P. tricornutum* were used calculate the difference in fold change for each sequence as detailed in (2).

$$D_x = P_x - F_x \quad (2)$$

where D_x is the difference in fold change between *P. tricornutum* sequence x and the *F. solaris* sequence homologous to x, P_x is the fold change of *P. tricornutum* sequence x and F_x is the fold change of the *F. solaris* sequence homologous to *P. tricornutum* sequence x.

C. Significance Test

The differences in fold change were transformed into *z*-scores which were calculated using (3).

$$z_x = \frac{D_x - E(D)}{SD(D)} \quad (3)$$

where z_x is the *z*-score of D_x, D_x is the difference in fold change between *P. tricornutum* sequence x and the *F. solaris* sequence homologous to x, E(D) is the expected value of D_x for all x and SD(D) is the standard deviation of D_x for all x.

The threshold of significance was set at 1% so sequences were selected if $-2{:}5758 < z_x < 2{:}5758$.

D. Gene Ontology

The gene ontologies of the significant sequences were tested for significance using the hypergeometric test in the GOstats R package [27] [30]. The GOstats package was used due to the inclusion of an option to perform a conditional hypergeometric test. It avoids an issue created when testing gene ontologies that are in a hierarchical structure of the gene ontology graph. The resulting p-values were corrected for multiple testing using Bonferroni's correction method. Gene ontologies with a p-value < 0.05 were selected to represent their respective group.

Acknowledgment

This work was supported by JST, CREST. We would also like to acknowledge Professor Wataru Fujibuchi for his assistance.

References

[1] J. W. Moody, C. M. McGinty, and J. C. Quinn, "Global evaluation of biofuel potential from microalgae," Proceedings of the National Academy of Sciences of the United States of America, vol. 111, no. 23, pp. 8691–6, Jun 2014.

[2] T. M. Mata, A. A. Martins, and N. S. Caetano, "Microalgae for biodiesel production and other applications: A review," Renewable and Sustainable Energy Reviews, vol. 14, no. 1, pp. 217–32, Jan-Feb 2010.

[3] P. D. Gressler, T. R. Bjerk, S. R. C., M. P. Souza, E. A. Lobo, A. L. Zappe, V. A. Corbellini, and M. S. Moraes, "Cultivation of Desmodesmus subspicastus in a tubular photobioreactor for bioremediation and microalgae oil production," Environmental Technology, vol. 35, no. 1-4, pp. 209–19, Jan-Feb 2014.

[4] C. Fuentes-Gr¨unewald, E. Garc´es, E. Alacid, N. Sampedro, S. Rossi, and J. Camp, "Improvement of lipid production in the marine strains Alexandrium minutum and Heterosigma akashiwo by utilizing abiotic parameters," Journal of Industrial Microbiology and Biotechnology, vol. 39, no. 1, pp. 207–16, Jan-Feb 2012.

[5] R. Radakovits, R. E. Jinkerson, S. I. Fuerstenberg, H. Tae, R. E. Settlage, J. L. Boore, and M. C. Posewitz, "Draft genome sequence and genetic transformation of the oleaginous alga Nannochloropsis gaditana," Nature Communications, vol. 3, Feb 2012.

[6] D. Shiran, I. Khozin, Y. M. Heimer, and Z. Cohen, "Biosynthesis of eicosapentaenoic acid in the microalga Porphyridium cruentum. I: The use of externally supplied fatty acids," Lipids, vol. 31, no. 12, pp. 1277–82, Dec 1996.

[7] M. Miyahara, M. Aoi, N. Inoue-Kashino, Y. Kashino, and K. Ifuku, " Highly efficient transformation of the diatom Phaeodactylum tricornutum by multi-pulse electroporation," Bioscience, Biotechnology, and Biochemistry, vol. 77, no. 4, pp. 874–876, 2013.

[8] K. E. Apt, P. G. Kroth-Pancic, and A. R. Grossman, "Stable nuclear transformation of the diatom Phaeodactylum tricornutum," Molecular Genetics and Genomics, vol. 252, no. 5, pp. 572–579, Oct 1996.

[9] C. Bowlder, A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, and R. P. Otillar, "The Phaeodactylum genome reveals the evolutionary history of diatom genomes," Nature, vol. 456, no. 7219, pp. 239–244, Nov 2008.

[10] M. Matsumoto, H. Sugiyama, Y. Maeda, R. Sato, T. Tanaka, and T. Matsunaga, "Marine diatom, Navicula sp. strain JPCC DA0580 and marine green alga, Chlorella sp. strain NKG400014 as potential sources for biodiesel production," Applied Biochemistry and Biotechnology, vol. 161, no. 1-8, pp. 483–90, May 2010.

[11] A. Satoh, K. Ichii, M. Matsumoto, C. Kubota, M. Nemoto, M. Tanaka, T. Yoshino, T. Matsunaga, and T. Tanaka, "A process design and productivity evaluation for oil production by indoor mass cultivation of a marine diatom, Fistulifera sp. JPCC DA0580," Bioresource Technology, vol. 137, pp. 132–8, Jun 2013.

[12] M. Muto, Y. Fukuda, M. Nemoto, T. Yoshino, T. Matsunaga, and T. Tanaka, "Establishment of a Genetic Transformation System for the Marine Pennate Diatom Fistulifera sp. Strain JPCC DA0580—A High Triglyceride Producer," Marine Biotechnology, vol. 15, no. 1, pp. 48–55, Feb 2013.

[13] D. Nojima, T. Yoshino, Y. Maeda, M. Tanaka, M. Nemoto, and T. Tanaka, "Proteomics Analysis of Oil Body-Associated Proteins in the Oleaginous Diatom," Journal of Proteome Research, vol. 12, no. 11, pp. 5293–301, Nov 2013.

[14] H. Rismani-Yazdi, B. Z. Haznedaroglu, C. Hsin, and J. Peccia, "Transcriptomic analysis of the oleaginous microalga Neochloris oleoabundans reveals metabolic insights into triacylglyceride

SEEK DIGITAL LIBRARY

accumulation," Biotechnology for Biofuels, vol. 5, no. 1, p. 74, Sep 2012.

[15] E. V. Armburst, J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brezezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kr¨oger, W. W. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson, and D. S. Rokhsar, "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution, and Metabolism," Science, vol. 306, no. 5693, pp. 79–86, Oct 2004.

[16] T. Tanaka, Y. Fukuda, T. Yoshino, Y. Maeda, M. Muto, M. Mitsufumi, S. Mayama, and T. Matsunaga, "High-throughput pyrosequencing of the chloroplast genome of a highly neutral-lipid-producing marine pennate diatom, Fistulifera sp. strain JPCC DA0580," Photosynthesis Research, vol. 109, no. 1-3, pp. 223–9, Sep 2011.

[17] I. Khozen-Goldberg and Z. Cohen, "Unraveling algal lipid metabolism: Recent advances in gene identification," Biochimie, vol. 93, no. 1, pp. 91–100, Jan-Feb 2011.

[18] Y. Liang, Y. Maeda, Y. Sunaga, M. Muto, M. Matsumoto, T. Yoshino, and T. Tanaka, "Biosynthesis of Polyunsaturated Fatty Acids in the Oleaginous Marine Diatom Fistulifera sp. Strain JPCC DA0580," Marine drugs, vol. 11, no. 12, pp. 5008–23, Dec 2013.

[19] M. D. Allen, J. A. Del Campo, J. Kropat, and S. S. Merchant, "FEA1, FEA2, and FRE1, encoding two homologous secreted proteins and a candidate ferrireductase, are expressed coordinately with FOX1 and FTR1 in iron-deficient Chlamydomonas reinhardtii," Eukaryotic Cell, vol. 6, no. 10, pp. 1841–1852, 2007.

[20] H. A. Dailey and T. A. Dailey, "Protoporphyrinogen oxidase of Myxococcus xanthus. Expression, purification, and characterization of the cloned enzyme," Journal of Biological Chemistry, vol. 271, no. 15, pp.8714–8, Apr 1996.

[21] T. Ishikawa and S. Shigeoka, "Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms," Bioscience, Biotechnology, and Biochemistry, vol. 72, no. 5, pp. 1143–54, May 2008.

[22] S. W. Lee, H. W. Lee, H. J. Chung, Y. A. Kim, Y. J. Kim, Y. Hahn, J. H. Chung, and Y. S. Park, "Identification of the genes encoding enzymes involved in the early biosynthetic pathway of pteridines in Synechocystis sp. PCC 6803," FEMS Microbiology Letters, vol. 176, no. 1, pp. 169–79, Jul 1999.

[23] S. Pohl, H. M. Mitchison, A. Kohlsch¨utter, O. Van Diggelen, T. Braulke, and O. Storch, "Increased expression of lysosomal acid phosphatase in CLN3-defective cells and mouse brain tissue," Journal of Neurochemistry, vol. 103, no. 6, pp. 2177–88, Dec 2007.

[24] A. Beranek, G. Rechberger, H. Knauer, H. Wolinski, S. D. Kohlwein, and R. Leber, "Identification of a cardiolipin-specific phospholipase encoded by the gene CLD1 (YGR110W) in yeast," Journal of Biological Chemistry, vol. 284, no. 17, pp. 11 572–8, Apr 2009.

[25] K. Yamada and K. Tomii, "Revisiting amino acid substitution matrices for identifying distantly related proteins," Bioinformatics, vol. 30, no. 3, pp. 317–25, Feb 2014.

[26] Z. K. Yang, Y. F. Niu, Y. H. Ma, J. Xue, M. H. Zhang, W. D. Yang, J. S. Liu, S. H. Lu, Y. Guan, and H. Y. Li, "Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation," Biotechnology for Biofuels, vol. 6, no. 1, p. 67, May 2013.

[27] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.r-project.org

[28] R. V. Lenth, lsmeans: Least-Squares Means, 2014.

[29] C. D. Warden, Y.-C. Yuan, and X. Wu, "Optimal Calculation of RNASeq Fold-Change Values ," International Journal of Computational Bioinformatics and In Silico Modeling, vol. 2, no. 6, pp. 285–292, 2013.

[30] S. Falcon and R. Gentleman, "Using GOstats to test gene lists for GO term association," Bioinformatics, vol. 23, no. 2, pp. 257–8, 2007.

About Author (s):

"… the difference in phenotype could be attributed to the down-regulation of these processes in *P. tricornutum* rather than the up-regulation in *F. solaris*."

SEEK DIGITAL LIBRARY