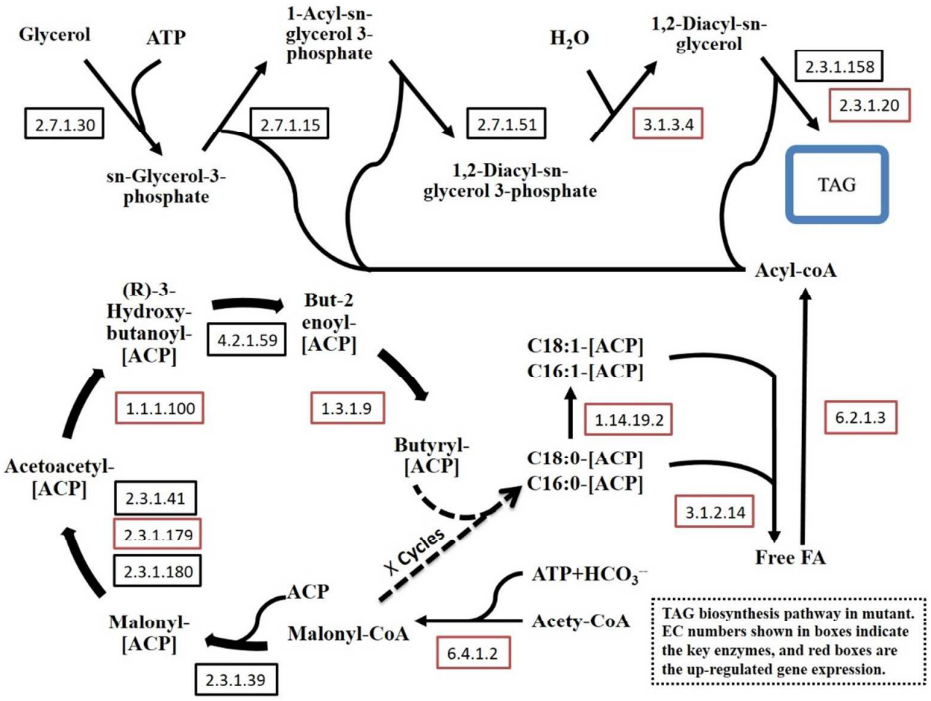
RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances



3,939 genes including cell growth and lipid synthesis key enzymes were up-regulated in mutant *Nitzschia* ZJU2 after two mutations by γ-rays which exhibited rapid growth and higher lipid productivity.

Gene expression and metabolic pathways related to cell growth and lipid synthesis in diatom *Nitzschia* ZJU2 after two-rounds of mutagenesis by γ-rays

Jun Cheng[a,*], Jia Feng[a], Ruolin Cheng[b], Chuanxi Zhang[b], Junhu Zhou[a], Kefa Cen[a]

[a] *State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China*

[b] *College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China*

Abstract:

The transcriptomes of original diatom strain (Wild-type, *Nitzschia sp.*) and a mutant strain (*Nitzschia* ZJU2), which exhibited rapid growth and high lipid productivity after two-rounds of mutagenesis by γ-rays, were sequenced using the Illumina sequencing platform. Genes in the metabolic pathway and those related to cell growth and lipid synthesis were compared between the two strains. Up to 25,804 and 33,198 transcripts were detected in *Nitzschia sp.* and *Nitzschia* ZJU2, respectively. A total of 3,939 genes were up-regulated in mutant *Nitzschia* ZJU2. Nine metabolic pathways involved in cell growth and carbohydrate and protein syntheses obviously changed. Genes involved in lipid synthesis, such as acetyl-CoA carboxylase and diacylgycerol *O*-acyltransferase, were obviously up-regulated. These phenomena promoted cell growth and lipid synthesis, so as to increase the lipid production of cells. Analysis of single nucleotide polymorphisms revealed the presence of 40,795 nonsynonymous mutation in *Nitzschia* ZJU2, which indicated that nuclear irradiation triggers algal mutation.

Keywords: diatom, gene expression, metabolic pathway, *Nitzschia sp*. transcriptome.

* Corresponding author: Prof. Dr. Jun Cheng. State Key Laboratory of Clean Energy Utilization, Zhejiang University, Hangzhou 310027, China. Tel.: +86 571 87952889; fax: +86 571 87951616. E-mail: juncheng@zju.edu.cn

1. Introduction

The gradual depletion of traditional fossil energy and global warming caused by greenhouse effect has placed utmost urgency on the search for renewable and, clean alternative energy [1]. Microalgal biomass that sequesters carbon dioxide can be converted into high-quality biodiesel, which can substitute fossil energy with zero carbon dioxide emission. Moreover, microalgae-derived biodiesel does not compete with food crops [2]. Diatoms are ubiquitous planktons with many species. Many diatoms have high potential to produce biodiesel after hardening culture [3]. Microalgal strains with high lipid productivity can be obtained by screening natural algal species [4], macro-control of growth factors [5, 6], and genetic engineering [7]. Single cell screening combined with γ-ray nuclear irradiation can also yield strains that exhibit rapid growth and high lipid productivity.

With the popularity and development of second-generation high-throughput sequencing, approaches, an RNA-Seq method based on the Illumina high-throughput sequencing platform has been widely used in recent years. Compared with previous methods, the RNA-Seq method has higher detection flux and accuracy, and can be used on any species. These qualities make the RNA-Seq method useful for transcriptome studies [8, 9]. Miller, et al. [10] used the RNA-Seq method to analyze changes in the metabolic pathways of *Chlamydomonas reinhardtii* under nitrogen deficiency. More than 27 % of the 130 reads of fatty acid-encoding genes were up-regulated, including diacylglycerol O-acyltransferase (DGAT) and other key enzymes that promote lipid

synthesis. However, they did not analyze changes in genes involved in photosynthetic and other metabolic pathways. Boyle, et al. [11] found that three acyltransferases aid in the accumulation of triglycerides (TAGs) in *Chlamydomonas* by RNA-Seq and gene analyses. However, changes in genes related to photosynthesis and other metabolic pathways were not analyzed. Rismani-Yazdi, et al. [12] identified genes encoding key enzymes in the green algae metabolic pathway that mediates biosynthesis in *Dunaliella tertiolecta*. In addition, the pathways mediating catabolism of lipid acid, TAG, and starch was also reconstructed. However, the effect of condition changes on the expression of genes involved in lipid synthesis was not analyzed as well. Cheng, et al. [13] also sequenced the *Nitzschia* metabolic pathway and found that salinity stress affected the expression of genes involved in lipid synthesis of *Nitzschia sp.* cells. However, changes in growth metabolism and gene expression related to lipid synthesis before and after the mutation were not compared. Therefore, a thorough study on the changes in the metabolic pathways and the expression of genes involved in lipid synthesis after nuclear irradiation in *Nitzschia* is warranted.

In this study, the transcriptomes of various *Nitzschia* strains were sequenced by an Illumina sequencing platform-based RNA-Seq approach. The metabolic pathway and expression of lipid synthesis-related genes of an original algal strain (*Nitzschia sp.*), and a mutant strain (*Nitzschia* ZJU2) that exhibits rapid growth and high lipid productivity after two-rounds of mutations, were analyzed. Results showed that expression levels of genes involved in key metabolic pathways, including photosynthesis and tricarboxylic acid cycle, were evidently changed. In addition,

expression levels of several genes involved in fatty acid synthesis, extension, and in TAG synthesis were significantly increased in the mutants.

2. Materials and methods

2.1. Microalgae strains and medium

Nitzschia sp. was purchased from the Institute of Hydrobiology, Chinese Academy of Science (FACHB-511). *Nitzschia* mutant cells were cultured by separating cells by the plating method after irradiation with ^{60}Co-γ-ray to ensure genetic uniformity of groups in sequencing samples. A genetically stable strain (i.e., *Nitzschia* ZJU1) that exhibited rapid growth and high lipid productivity was selected through fluorescence microscopy after Nile Red staining. *Nitzschia* ZJU1 was subjected to ^{137}Cs-γ ray irradiation and was cultured in a similar manner as *Nitzschia sp.*. Fluorescence microscopy was performed to select another genetically stable strain (i.e. *Nitzschia* ZJU2) that exhibited rapid growth and high lipid productivity. All cells used for RNA-Seq had been cultured in D1 medium in the presence of air until they reached logarithmic growth phase under the following conditions: constant temperature, 27±2 °C; light intensity, 6000 lux; and brightness ratio, 24 h: 0 h. The pH of the initial media was regulated at 8.5±0.1 (0.1 M HCl and 0.1 M NaCl). The cells were centrifuged using a Beckman centrifuge at 10,000×g for 10 min at 8 ℃. The D1 standard medium consisted of 0.12 g $NaNO_3$, 0.07 g $MgSO_4 \cdot 7H_2O$, 0.02 g $CaCl_2 \cdot 2H_2O$, 0.04 g KH_2PO_4, 0.08 g K_2HPO_4, 0.1 g Na_2SiO_3, 0.0002 g $MnSO_4$, 0.005 g ferric citrate, 20 mL soil extract liquid [14], 1 mL A_5 solution and 979 mL distilled water. The A_5 solution consisted of 286 mg H_3BO_3, 186 mg

$MnC1_2 \cdot 4H_2O$, 22 mg $ZnSO_4 \cdot 4H_2O$, 39 mg $Na_2MoO_4 \cdot 2H_2O$, 8 mg $CuSO_4 \cdot 5H_2O$ and 5 mg $Co(NO_3)_2 \cdot 6H_2O$ in 1000 mL of distilled water.

2.2. Identification of microalgae species

In order to confirm there was no contamination during the process of mutagenesis and screening of mutant cells, and to figure out whether the 18S rDNA had been altered by γ-ray irradiation during mutagenesis, the microalgae species of *Nitzschia* ZJU2 was identified before the comparison with the original strain *Nitzschia sp.* Identification of microalgae species was done by 18s rDNA analysis as described previously by Cheng's study [13]. The following primers were used: 18s-F, AACCTGGTTGATCCTGCCAGT, and 18s-R, TGATCCTTCTGCAGGTTCACCT. The gene was sub-cloned into the vector pMD19-T. Then, clones that were confirmed to be positive were selected for sequencing. The 18s rDNA genes of the remaining related 12 microalgae strains were downloaded from the NCBI database. The ClustalX [15] software was used to for multiple sequence alignment, and the MEGA5 [16] software was used to construct a phylogenetic tree using the Kimura-2-Parameter model. Bootstrap 1000 was used to determine the confidence probability of each branch. *Bolidomonas mediterranea* (Heterokontophyta) was used as the outside group.

2.3. cDNA library construction and Illumina sequencing

The original strain *Nitzschia sp.* and mutant *Nitzschia* ZJU2 were collected at their logarithmic growth phase by centrifugation. Total RNA was extracted using TRIzol Reagent (Invitrogen) reagent [13]. To obtain 5 μg of total RNA, mRNA was separated using the magnetic sand method, and

cleaved to synthesize double-chain cDNA. A was added at the 3' terminal, and index connection was linked (Truseq^{TM} RNA sample prep Kit). The target strip was enriched by PCR (15 cycles) and recycled using 2 % agarose gel. TBS380 (Picogreen) was used for definite quantitative determination, and bridge amplification was conducted for cBot cluster generation (TruSeq paired-end cluster kit v3-cBot-HS; Illumina) [13]. The Hiseq 2000 sequencing platform (100bp, TruSeq SBS kit v3-HS 200 cycles; Illumina) [13] was used for the 2*100 bp sequencing test.

2.4. Sequence assembly and annotation

As described previously by Cheng, et al. [13], the original sequencing data obtained using Illumina Hiseq 2000 includes sequencing linker sequence, low-quality read section, high-N rate sequence, and short-length sequence. These sequences negatively affect the quality of subsequent assemblies. Therefore, the original sequencing data were filtered. Low-quality sequences were removed to obtain high-quality sequencing data (clean data), which is important for accurate bioinformatic analysis. The obtained clean data were then transformed into sequencing assembly by using the Trinity de novo assembler software [17]. The BLAST program was used to compare unigenes obtained from the assembly with the GenBank non-redundant protein database nr (E value < 1 e-5), and the best annotation was selected. Blast2Go software (http://www.blast2go.org/) was used to obtain GO (Gene Ontology) information. The sequences were classified according to molecular function, cell component, and biological process. Gene function information annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG) was used as a basis to analyze relevant

metabolic pathways. All annotation information was organized to search for genes encoding the key enzymes involved in lipid synthesis, cell wall biogenesis, photosynthesis, and nitrogen metabolism. Transcription factors, which may participate in regulation of those genes, were also searched.

2.5. Analysis of differentially expressed genes before and after mutation

With the assembled transcriptome sequence (unigene set) as the reference database, RSEM (http://deweylab.biostat.wisc.edu/rsem/) was used to calculate the number of corresponding DNA segments, and gene expression in different samples was analyzed. EdgeR (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html) was used as basis to homogenize the number of segments counted by RSEM, and the expression difference between genes was calculated. The criteria used were: greater than two-fold change in gene expression, and a false discovery rate (FDR) of less than 1 %. GO function and pathway significant enrichment analyses were conducted by using goatools (https://github.com/tanghaibao/goatools) [18] and KOBAS software (http://kobas.cbi.pku.edu.cn/home.do) [19], respectively. The method was done by Fisher test. To control the false positive rate of calculation, four multiple detection methods (Bonferroni, Holm, Sidak, and FDR) were used to calibrate the p value. Enrichment in certain function or pathway was considered significant if p (p_FDR) ≤ 0.05.

2.6. Single Nucleotide Polymorphism (SNP) analysis

SNPs are DNA sequence polymorphisms caused by the mutation of a single nucleotide, and occur with allele frequency of ≥ 1 %. SNPs in animal and plant genomes with genetic stability are

conventionally secondary bit (individuals are third or fourth bit), which makes automatic analysis easy [20]. According to the SNP testing method of Ahmadian, et al. [21], sequences from two samples were compared for analysis of SNPs. SNPs were classified as either synonymous (amino acid sequence conserved) or non-synonymous (amino acid sequence changed) to predict whether the mutation was likely to affect protein function.

3. Results and Discussion

3.1. Growth comparison between *Nitzschia sp.* and *Nitzschia* ZJU2

γ- ray irradiation has been shown previously to penetrate microalgal cells and promote growth [14]. However, mutagenesis of algal cells by γ-ray is non-directional. Therefore, single-cell separation culture of algal strains was conducted after mutagenesis. Advantageous algal species with high density of biomass and high lipid productivity were screened by testing sample absorbance and by fluorescence microscopy after Nile Red staining [22]. Figure.1 shows the original microalgae strain *Nitzschia sp.* (Fig. 1a), the advantageous algal strain *Nitzschia* ZJU1 (Fig. 1b) screened out after the first mutation, and the advantageous algal strain *Nitzschia* ZJU2 (Fig. 1c) screened out after the second mutation. Cells were photographed under a fluorescence microscope after Nile Red staining (final content 1mg L^{-1}; staining time 7 min). *Nitzschia sp.* and *Nitzschia* ZJU2 had the lowest and the highest fluorescence intensity, respectively. NIS-ELEMENTS software was used to calculate the lipid content of microalgae:

$$LC_A = \frac{A_L}{A_C} \times 100\%,$$

where A_L is the total area of lipid droplets in the cells and A_C is the total area of cells. *Nitzschia sp.*, *Nitzschia* ZJU1, and *Nitzschia* ZJU2 had LC_A values of 10.5 %, 8.24 %, and 25.46 %, respectively, and cell densities of 0.5×10^6, 1.5×10^6, and 3.8×10^6 cells mL^{-1}, respectively. Since mutation of microalgae cells by nuclear irradiation is non-directional, the lipid content and biomass yield of mutants may either increase or decrease. A key technology is to efficiently screen out advantageous mutants through cell growth experiments. Accordingly, a advantageous mutant with the highest lipid yield, which was obtained by multiplying biomass dry weight with lipid content, was selected Although *Nitzschia* ZJU1 showed a lower lipid content compared to the original strain (precisely 9.88% vs. 11.87% based on lipid extraction and weighing, and roughly 8.24% vs. 10.50% based on fluorescence microscope images of microalgae cells dyed with Nile Red), the lipid yield of *Nitzschia* ZJU1 (39.52 mg/L=0.4g/L×1000mg/g×9.88%) was much higher than that of the original strain (15.43 mg/L=0.13g/L×1000 mg/g×11.87%). Therefore, *Nitzschia* ZJU1 was screened out as an advantageous strain after the first round of mutagenesis by γ-rays. Nuclear irradiation promoted the growth and increased the lipid yield production of the screened advantageous mutant strains. Moreover, the size of the cells slightly decreased as lipid enrichment of the cells improved. The average cell lengths of *Nitzschia sp.*, *Nitzschia* ZJU1, and *Nitzschia* ZJU2 were 12±1 μm, 11±1 μm, and 10±1 μm, respectively, with mean widths of 3±1 μm, 4±1 μm, and 4±1 μm, respectively. These results are consistent with those reported by Lynn, et al. [23].

The screened advantageous algal strains were maximized and cultured until the 10^{th} generation,

whose biomass dry weight and lipid yield production were shown in Figure.2. The growth curves of the 1st to 9th generations (the growth period of one generation of the strain was about two weeks) of these *Nitzschia* strains were not shown. The biomass density of *Nitzschia* ZJU2 significantly increased by 7.6 times compared with that of *Nitzschia sp.*, and its culture period was shortened from 15 d to 12 d. Meanwhile, the lipid enrichment capacity of mutant was also markedly improved. The lipid productivity of *Nitzschia* ZJU2 was 20 times greater than that of *Nitzschia sp.* After optimizing the medium for *Nitzschia* ZJU2 growth, the cell lipid content surpassed 55 % under nitrogen and silicon deficiency. Moreover, the total lipid yield production could reach up to twice that of *Nitzschia* ZJU2. To further investigate and explain the rapid growth and high lipid production of mutant, *Nitzschia sp.* and *Nitzschia* ZJU2 were selected for transcriptome sequencing and gene expression analysis.

3.2. Molecular identification of microalgal species

After PCR reaction, 18s rDNA sequence of *Nitzschia* ZJU2 was analyzed for identification. BLAST comparison revealed that the 18s rDNA sequence of *Nitzschia* strains shared 99 % similarity to that of Bacillariophyta, Bacillariophyceae, and *Nitzschia*. Maximum likelihood was used to perform phylogenetic analysis. Results showed that *Nitzschia* ZJU2 was in the same branch as *Nitzschia themalis* and shared 97 % with *Hantzschia amphioxys*. Comparison with two diatoms having complete full genome sequencing, namely, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*, it revealed 96 % and 91 % similarity, respectively. These results are consistent with

194 those reported by Cheng's study [13].

195 **3.3. Comparison of transcriptome sequencing, assembly, and annotation**

196 A total of 10 G data was generated by Illumina sequencing of *Nitzschia* ZJU2 cells. More than
197 96.24 % of these data were high-quality sequences. After impurities and redundancies were
198 removed from original data generated by the sequencing, clean data were obtained for assembly. A
199 total of 35,228 unigenes with lengths varying from 351 bp to 24,531 bp (average: 1,897.46 bp)
200 were obtained. The BLASTx algorithm was used to search and compare the non-redundant protein
201 library of NCBI. A total of 20,630 unigenes (58.6 %) were obtained after gene annotation.
202 Moreover, the Blast2GO software was used to obtain 9,563 classified unigene GO information (Fig.
203 3). The KEGG database was used as basis to analyze the metabolic pathway of gene products in
204 cells. A total of 8,592 sequences were annotated. This Transcriptome Shotgun Assembly project has
205 been deposited at DDBJ/EMBL/GenBank under the accession GBCF00000000. The version
206 described in this paper is the first version, GBCF01000000.

207 Up to 25,804 transcripts were detected in *Nitzschia sp.* cells. The distribution of GO terms
208 (Fig.3) showed that the transcripts number of Biological process, Cellular component, and
209 Molecular function were slightly increased in *Nitzschia* ZJU2 cells. This result indicated that gene
210 expression was more active in *Nitzschia* ZJU2 than in the other strains, which is consistent with the
211 observation of improved growth of *Nitzschia* ZJU2.

212 **3.4. Comparison of gene expression related to cell growth**

Transcript comparison between *Nitzschia sp.* and *Nitzschia* ZJU2 revealed the expression of 23,774 transcripts before and after mutagenesis (Fig. 4). After the microalgae were mutated, 3,939 transcripts were up-regulated and 691 genes were down-regulated. Enrichment analysis of GO function significance on genes with significant expression difference before and after mutagenesis showed that their GO functions were not significantly enriched. In addition, enrichment analysis of the KEGG pathway revealed that the gene expression in nine pathways had evidently changed after the mutagenesis (Table. 1).

These nine metabolic pathways were closely related to cell growth and carbohydrate and protein syntheses. Photosynthesis (including ko00195 and ko00710) was a biochemical pathway that transforms carbon dioxide and water into organics and oxygen. This pathway was the basis of the organism's survival. The carbohydrate formed in the reaction was the fundamental component of living cell structures and the main material for providing energy, and the principal function of this component was cell activity regulation. After nuclear mutagenesis, various genes participating in photosynthesis were evidently up-regulated, including the genes that encode for the key enzyme rubisco. This enzyme catalyzes the carboxylation of the carbonic acid receptor 1,5-ribulose diphosphate (RuBP) and carbon dioxide. The expression of phosphate ribulose kinase (EC2.7.1.19) significantly increased. This enzyme catalyzes 5-phosphate ribulose and ATP. Chlorophyll is the most important pigment related to photosynthesis. The activity of chlorophyll metabolism indirectly promotes photosynthesis. The increased expression of these enzymes enhanced

232 assimilation and promoted cell growth.

233 The tricarboxylic acid cycle (TAC) is the most effective method for the organism to gain
234 energy by oxidizing sugar and other substances. This cycle also acts as a hub for the metabolism
235 and transformation of sugar, lipid and protein. The intermediate products of the TAC, such as
236 oxaloacetate, α-ketoglutaric acid, pyruvate, and acetyl CoA, are the raw materials for synthesizing
237 sugar, amino acids, and fat. Oxidative phosphorylation after glycolysis and TAC is the main step in
238 producing ATP. The expression of genes involved in these metabolic pathways were found to have
239 increased in the mutant strain. This result indicate that cell metabolism was highly active and
240 materials for fatty acid synthesis were being produced in abundance in the mutant *Nitzschia* ZJU2
241 strain.

242 **3.5. Changes in lipid synthesis-related genes**

243 According to KEGG annotations, the genes involved in the pathways related to lipid synthesis
244 were identified and the TAG biosynthesis pathway was reconstructed speculatively (Fig. 5). In
245 several pathways related to lipid metabolism, the gene expression after nuclear mutagenesis
246 changed (Table. 2, Fig. 5). In microalgal cells, the synthetic pathway of lipids was similar to that of
247 the model organism *Chlamydomonas reinhardtii* and other higher plants [24-26]. Acetyl-CoA
248 carboxylase (acc) was the rate-limiting enzyme for the catalytic synthesis of fatty acid chains. Its
249 expression was increased by five times after mutagenesis. 3-oxoacyl-[acyl-carrier-protein] synthase
250 II (fabF) is involved in a condensation reaction in the first step of catalysis, and its expression was

up-regulated 2.7 times in *Nitzschia* ZJU2. 3-oxoacyl-[acyl-carrier protein] reductase (fabG) and enoyl-[acyl-carrier protein] reductase I (fabI) undergo a catalytic reduction reaction, and their expression levels had changed 2 and 4.3 times, respectively, after mutagenesis. The changes in the genes promoted the reaction progress.

The expression levels of several enzymes that catalyze the chain extension of fatty acid increased by more than two fold after mutagenesis in *Nitzschia* ZJU2. These enzymes include β-keto reductase, palmitic acid-protein thioesterase, 3-oxoacyl-[acyl-carrier-protein] synthase, and trans-2-enoyl-CoA reductase. After the synthesis of free fatty acids, acyl CoA was formed by the catalysis of long-chain acyl CoA synthase (ACSL), which was then added to a glycerine-3-phosphate backbone to synthesize TAG. This process required the participation of glycerine-3-phosphate-O-acyltransferase, 1-acylglycerol-3-phosphate-O-acyltransferase (AGPAT), phosphatidate phosphatase (PP), and DGAT, among other proteins. Particularly for DGAT, this process was the last reaction step for catalyzing TAG synthesis. DGAT was also the key rate-limiting enzyme in TAG synthesis. The expression of ACSL increased by 6.3 times after mutagenesis in *Nitzschia* ZJU2. In addition, the expression levels of AGPAT, PP, and DGAT increased by more than three times in *Nitzschia* ZJU2.

In Figure. 5, red square was used to indicate the *Nitzschia* ZJU2 genes that were up-regulated compared with the *Nitzschia sp*. genes. These genes were found to participate in nine metabolic pathways of lipid synthesis. The changes in gene expression may promote lipid synthesis, causing

the lipid content of the cells to increase. Recent studies have induced the overexpression of LPAT, DGAT, and other enzymes to enhance lipid productivity in plants [27]. However, these studies obtained poor results [28]. Therefore, analyzing the changes in the metabolic pathway for lipid production can be helpful in finding potential targets in microalgae suitable for effective enhancement of lipid production.

3.6. SNP analysis

SNPs are single nucleotide variations in the genome that may involve replacement, transversion, deletion, and insertion. SNPs were widespread under natural conditions, and nuclear irradiation increased the probability of base mutation. The comparison between *Nitzschia* sp. and *Nitzschia* ZJU2 samples revealed a total of 68,010 mutations. These mutations included 27,215 synonymous mutations and 40,795 non- synonymous mutations (Appendix A). The number of SNPs distributed in the third place of codon was the most abundant (i.e., 30,460). The number of SNPs at the second and first places were 17,156 and 20,394, respectively. These SNPs were distributed in 1,701 genes and participated in 250 metabolic pathways, among which, non-synonymous mutation occurred in 1,274 transcripts. That was the sequence of amino acids changed.

After mutation, the amino acid sequences of several carbon metabolism-related genes, including pyruvate orthophosphate dikinase (PPDK), pyruvate kinase, and phosphopyruvate carboxylase (PPC), were changed. These changes may affect the activity of proteins. These

289 enzymes not only function in carbon fixture, which was related to growth, but also indirectly
290 affected lipid synthesis. PPDK and pyruvate phosphokinase could catalyze pyruvate to produce
291 phosphoenolpyruvate (PEP). In oil seeds, PEP generated from glycolysis was conventionally
292 transported to the plastid to be used as a precursor of synthesized acetoacetyl coenzyme A [27, 29],
293 which served as a raw material for lipid synthesis. In addition, PPC catalyzed the conversion of
294 PEP into oxaloacetate, which synthesized the response competition with TAG [27]. Several non-
295 synonymous mutations were distributed in lipid synthesis-related genes, such as acetyl CoA
296 carboxylase, long-chain acyl CoA synthetase, and glycerine-3-phosphate-O-acyltransferase.
297 Whether or not changes in amino acid sequences caused by these mutations affected the secondary
298 structure of protein remained unknown.

299 Moreover, non-synonymous mutations were found in numerous protein kinases and
300 transcription factors. Protein kinases catalyzed protein phosphorylation and could transform
301 γ-phosphate in ATP into amino acid residues, which were incorporated into protein molecules.
302 Transcription factors could regulate gene expression upstream, respond to certain external stimuli,
303 and activate the expression of several proteins. Therefore, the sequence changes in protein kinase or
304 transcription factors may affect their structures and activities, thereby influencing cell metabolism
305 and growth.

306 Non- synonymous mutations, which alter genes at the amino acid level, can be classified as
307 harmful, neutral, or favorable. Irradiation-induced mutations occur randomly. In the selection

process after mutation, individuals with harmful mutation were mostly eliminated. Therefore, the surviving groups may be endowed with several favorable mutations. Overall, these SNP data may serve as a useful reference for studying the effects of nuclear mutation. However, whether or not such a mutation affects the structure, activity, and function of enzymes needs further analysis and verification.

4. Conclusion

The gene expression and metabolic pathways were analyzed to account for the more rapid growth and higher lipid productivity of mutant *Nitzschia* ZJU2, which was obtained after two-rounds of mutations by γ-rays compared to the original diatom *Nitzschia sp.* Genes expression in metabolic pathways related to cells growth, including photosynthesis and TAC, were significantly up-regulated in the mutant strain. In addition, the expression levels of several genes participate in fatty acid synthesis, extension, and TAG synthesis were significant higher in the mutant. Future studies should determine the effects of mutation at a certain site on the structure of gene-encoded proteins and elucidate the mechanism of action behind these effects.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (51176163), National High Technology R&D Program of China (2012AA050101), International Sci. & Tech. Cooperation Program of China (2012DFG61770 and 2010DFA72730), National Key Technology R&D Program of China (2011BAD14B02), Program for New Century Excellent Talents in

University (NCET-11-0446), Specialized Research Fund for the Doctoral Program of Higher Education (20110101110021), Science and Technology Project of Guangxi Province (1346011-1).

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the Electronic Supplementary Information.

References:

1. H. M. Amaro, Â. C. Macedo and F. X. Malcata, *Energy*, 2012, **44**, 158-166.

2. E. A. Ramos Tercero, E. Sforza and A. Bertucco, *Energy*, 2013, **60**, 373-379.

3. L. Brennan and P. Owende, *Renew Sust Energ Rev*, 2010, **14**, 557-577.

4. J. Sheehan, T. Dunahay, J. Benemann, P. Roessler and O. Danielo, *Colorado: TP-580-24190, National Renewable Energy Laboratory* 1998, **255**, 1-4.

5. Y. Chisti, *Biotechnol Adv*, 2007, **25**, 294-306.

6. Q. Hu, M. Sommerfeld, E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert and A. Darzins, *Plant J*, 2008, **54**, 621-639.

7. L. A. Zaslavskaia, J. C. Lippmeier, C. Shih, D. Ehrhardt, A. R. Grossman and K. E. Apt, *Science*, 2001, **292**, 2073-2075.

8. E. R. Mardis, *Annu Rev Genom Hum G*, 2008, **9**, 387-402.

9. Z. Wang, M. Gerstein and M. Snyder, *Nat Rev Genet*, 2009, **10**, 57-63.

10. R. Miller, G. Wu, R. R. Deshpande, A. Vieler, K. G A Rtner, X. Li, E. R. Moellering, S. Z A

346 Uner, A. J. Cornish, B. Liu and Others, *Plant Physiol*, 2010, **154**, 1737-1752.

347 11. N. R. Boyle, M. D. Page, B. Liu, I. K. Blaby, D. Casero, J. Kropat, S. J. Cokus, A.

348 Hong-Hermesdorf, J. Shaw and S. J. Karpowicz, *J Biol Chem*, 2012, **287**, 15811-15825.

349 12. H. Rismani-Yazdi, B. Z. Haznedaroglu, K. Bibby and J. Peccia, *BMC genomics*, 2011, **12**,

350 148.

351 13. R.-l. Cheng, J. Feng, B.-X. Zhang, Y. Huang, J. Cheng and C.-X. Zhang, *Bioenerg. Res.*,

352 2013, 1-14.

353 14. J. Cheng, Y. Huang, J. Feng, J. Sun, J. Zhou and K. Cen, *Bioresource Technol*, 2013, **136**,

354 496-501.

355 15. M. Larkin, G. Blackshields, N. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F.

356 Valentin, I. M. Wallace, A. Wilm and R. Lopez, *Bioinformatics*, 2007, **23**, 2947-2948.

357 16. K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar, *Mol Biol Evol*, 2011,

358 **28**, 2731-2739.

359 17. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis,

360 L. Fan, R. Raychowdhury and Q. Zeng, *Nat Biotechnol*, 2011, **29**, 644-652.

361 18. H. Tang, X. Wang, J. E. Bowers, R. Ming, M. Alam and A. H. Paterson, *Genome Res*, 2008,

362 **18**, 1944-1954.

363 19. C. Xie, X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C.-Y. Li and L. Wei,

364 *Nucleic Acids Res*, 2011, **39**, W316-W322.

20. A. J. Brookes, *Gene*, 1999, **234**, 177-186.

21. A. Ahmadian, B. Gharizadeh, A. C. Gustafsson, F. Sterky, P. Nyrén, M. Uhlén and J. Lundeberg, *Anal Biochem*, 2000, **280**, 103-110.

22. P. Greenspan, E. P. Mayer and S. D. Fowler, *J Cell Biol*, 1985, **100**, 965-973.

23. S. G. Lynn, S. S. Kilham, D. A. Kreeger and S. J. Interlandi, *J Phycol*, 2000, **36**, 510-522.

24. J. Ohlrogge and Others, *The Plant Cell*, 1995, **7**, 957

25. J. B. Ohlrogge, C. R. Somerville and Others, *Biochim Biophys Acta Lipids Lipid Metabol*, 1991, **1082**, 1-26.

26. W. R. Riekhof, B. B. Sears, C. Benning, R. Miller, G. Wu, R. R. Deshpande, A. Vieler, K. G A Rtner, X. Li, E. R. Moellering, S. Z A Uner, A. J. Cornish, B. Liu and Others, *Eukaryotic Cell*, 2005, **4**, 242-252.

27. N. E. M. M. Courchesne, A. Parisien, B. Wang and C. Q. Lan, *J Biotechnol*, 2009, **141**, 31-41

28. R. Radakovits, R. E. Jinkerson, A. Darzins and M. C. Posewitz, *Eukaryotic Cell*, 2010, **9**, 486-501.

29. S. Baud, S. Wuillème, B. Dubreucq, A. De Almeida, C. Vuagnat, L. Lepiniec, M. Miquel and C. Rochat, *Plant J*, 2007, **52**, 405-419.

List of Figures and Tables:

Fig.1 Fluorescence microscope images of microalgae cells dyed with Nile Red

384 Fig.2 Biomass dry weight and Lipid yield of microalgae strains.

385 Fig.3 Distribution of level 2 GO terms of *Nitzschia sp.*and *Nitzschia* ZJU2 transcripts.

386 Fig.4 Transcriptional expression of *Nitzschia sp.*and *Nitzschia* ZJU2 cells.

387 Fig.5 TAG biosynthesis pathway reconstructed based on *de novo* assembly and annotation of

388 *Nitzschia* ZJU2 transcriptome.

389 Table 1. Different gene expression in metabolic pathways of *Nitzschia sp.*and *Nitzschia* ZJU2 cells

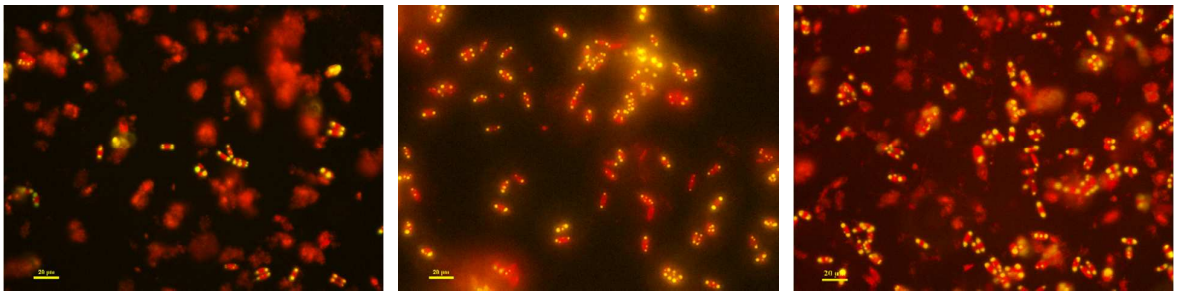
390 Table.2 Changes in transcript expression of genes related to lipid biosynthesis in *Nitzschia* ZJU2.

391

392

393

394



395

396 (a) *Nitzschia sp.* (b) *Nitzschia* ZJU1 (c) *Nitzschia* ZJU2

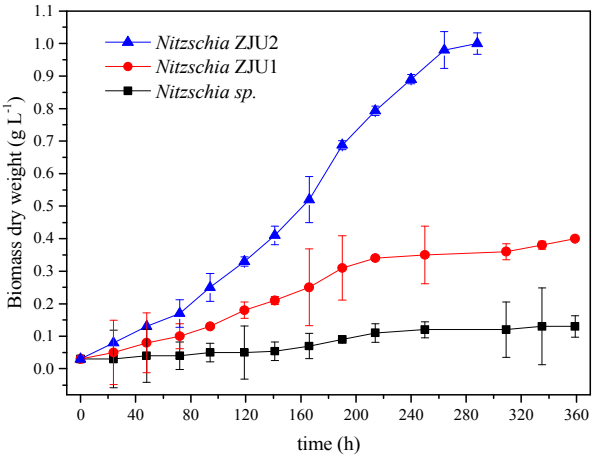
397 Fig.1 Fluorescence microscope images of microalgae cells dyed with Nile Red.

398 (Note: *Nitzschia sp.* was the original algal strain. *Nitzschia* ZJU1 was the mutant that exhibited

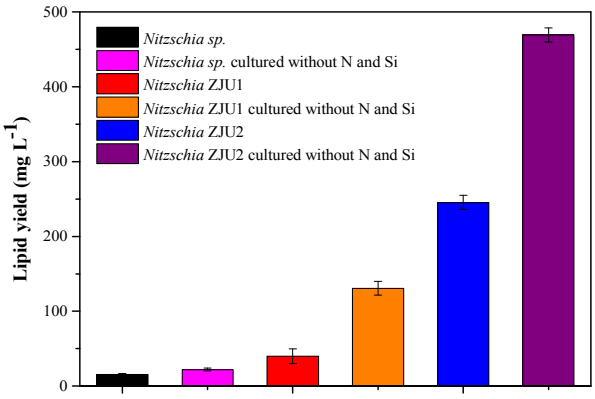
399 rapid growth and high lipid productivity after the first mutation. *Nitzschia* ZJU2 was the mutant

400 that exhibited rapid growth and high lipid productivity after two mutations.)

401

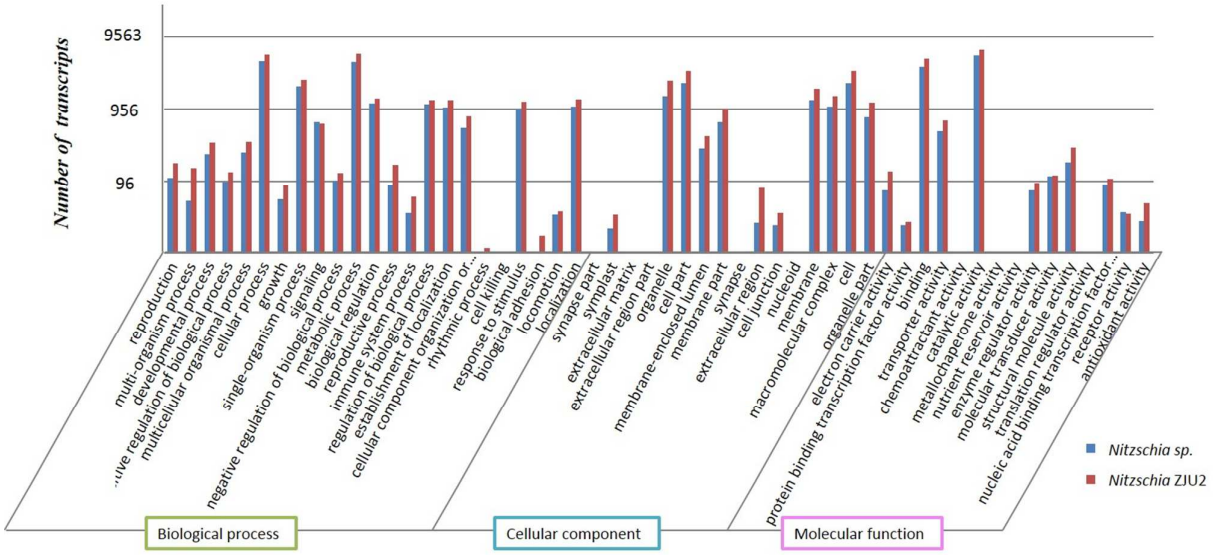


(a) Growth curves of different *Nitzschia* strains.



(b) Lipid yields of different *Nitzschia* strains cultured in different media

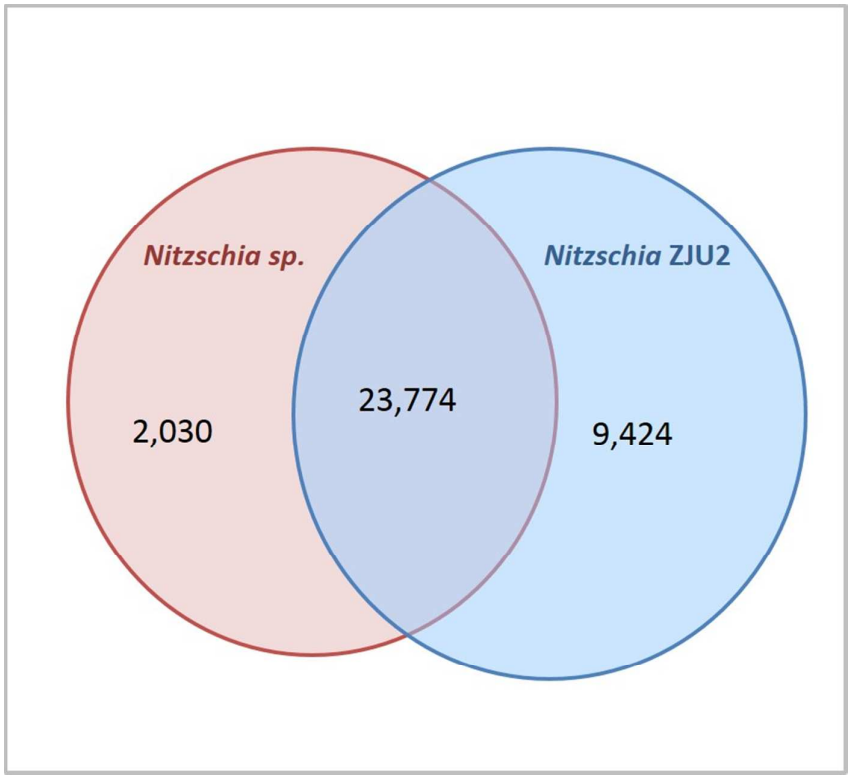
Fig.2 Biomass dry weight and Lipid yield of microalgae strains.



407

408 Fig.3 Distribution of level 2 GO terms of *Nitzschia sp.*and *Nitzschia* ZJU2 transcripts.

409

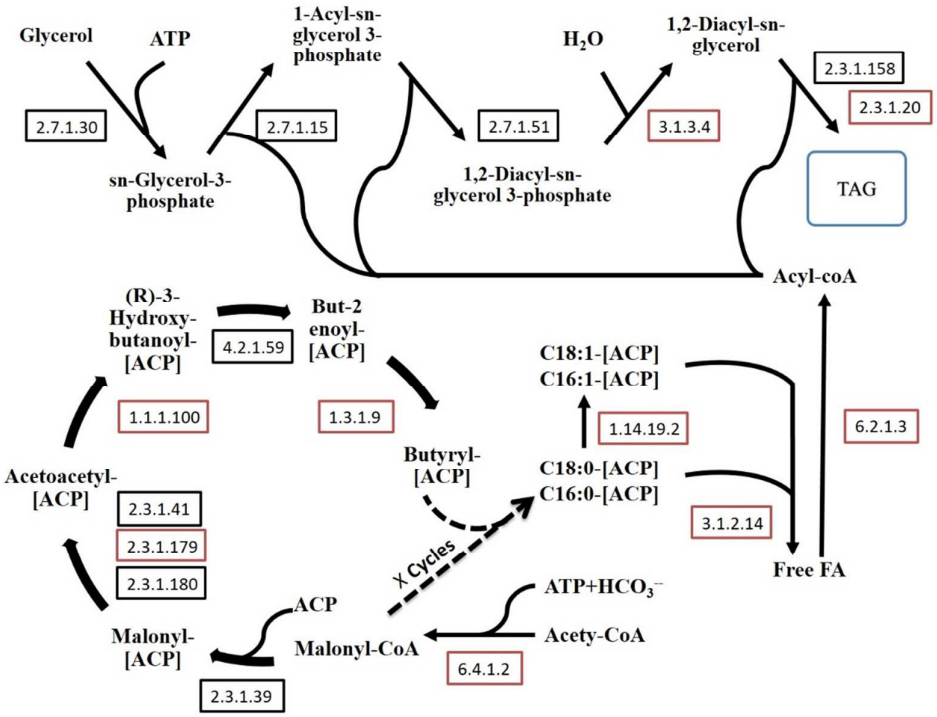


410

411 Fig.4 Transcriptional expression of *Nitzschia sp.*and *Nitzschia* ZJU2 cells.

412

413



414

415 Fig.5 TAG biosynthesis pathway reconstructed based on *de novo* assembly and annotation of

416 *Nitzschia* ZJU2 transcriptome. EC numbers shown in boxes indicate the key enzymes, and red

417 boxes are the up-regulated gene expression.

418

419 Table 1. Different gene expression in metabolic pathways of *Nitzschia sp.*and *Nitzschia* ZJU2 cells

Pathway	Id	Sample number/ Background number	P-Value
Ribosome	ko03010	92/319	3.92E-08
Oxidative phosphorylation	ko00190	62/217	2.50E-05
Starch and sucrose metabolism	ko00500	38/127	0.001096
Porphyrin and chlorophyll metabolism	ko00860	32/103	0.001847
Photosynthesis	ko00195	26/77	0.001847
Spliceosome	ko03040	92/410	0.001912
Citrate cycle (TCA cycle)	ko00020	40/144	0.002638
Cell cycle - Caulobacter	ko04112	11/22	0.004239
Carbon fixation in photosynthetic organisms	ko00710	43/175	0.017272

420

421

422

423 Table.2 Changes in transcript expression of genes related to lipid biosynthesis in *Nitzschia* ZJU2.

Description	KO/Gene_ID	Name	EC No.	Fold change	FDR
acetyl-CoA carboxylase carboxyltransferase subunit alpha	K01962	accA	EC:6.4.1.2	5.5	1.53E-05
acetyl-CoA carboxylase biotin carboxyl carrier protein	K02160	accB, bccP		4.0	6.08E-05
acetyl-CoA carboxylase / biotin carboxylase	K01961	accC	EC:6.4.1.2 6.3.4.14	5.1	2.06E-05
3-oxoacyl-[acyl-carrier-protein] synthase II	K09458	fabF	EC:2.3.1.179	2.7	0.0002
3-oxoacyl-[acyl-carrier protein] reductase	K00059	fabG	EC:1.1.1.100	2.0	0.0005
enoyl-[acyl-carrier protein] reductase I	K00208	fabI	EC:1.3.1.9 1.3.1.10	4.3	4.69E-05
acyl-[acyl-carrier-protein] desaturase	K03921	DESA1	EC:1.14.19.2	4.9	2.48E-05
acy-ACP thioesterase	K10782	Fata	EC:3.1.2.14 3.1.2.-	2.0	0.0006
beta-keto reductase	K10251	KAR	EC:1.1.1.-	2.0	0.0005
palmitoyl-protein thioesterase	K01074	PPT	EC:3.1.2.22	4.2	0.0008
3-ketoacyl-CoA synthase	K15397	KCS	EC:2.3.1.-	2.1	0.0005
mitochondrial trans-2-enoyl-CoA reductase	K07512	MECR, NRBF1	EC:1.3.1.38	2.8	0.0002
long-chain acyl CoA synthase	K01897	ACSL	EC:6.2.1.3	6.3	9.40E-06
1-acyl-sn-glycerol-3-phosphate acyltransferase	K00655 K14674	AGPAT	EC:2.3.1.51	3.3	0.0001
phosphatidate phosphatase	K15728	PP	EC:3.1.3.4	4.7	3.00E-05
diacylglycerol O-acyltransferase	K00635 K11155	DGAT	EC:2.3.1.20	3.4	0.0003

424