Genomics Data 6 (2015) 231–233

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Nitrogen deprivation-induced de novo transcriptomic profiling of the oleaginous green alga *Botryococcus braunii* 779

Zhenyu Xu [a], Jing He [b], Shuyuan Qi [a], Jianhua Liu [a,b,*]

[a] Ocean College, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Ocean Research Centre of Zhoushan, Zhoushan, Zhejiang 316021, China

ARTICLE INFO

Article history:
Received 15 September 2015
Accepted 15 September 2015
Available online 24 October 2015

Keywords:
Botryococcus braunii
Hydrocarbons
Illumina paired-end sequencing
Nitrogen deprivation
Transcriptomic profiling

ABSTRACT

To assess the effect of nitrogen deprivation (ND), a moderately growing A-race *Botryococcus braunii* subisolate 779 was subjected to nitrogen deprivation for 3 days. De novo transcriptome was assembled and annotated by using Trinity software and Basic Local Alignment Search Tools (BLAST), respectively. Comparative analysis indicates that transcriptomes of A-races differ from those of B-races. Furthermore, majority of the homologous ESTs in A-race but not B-race transcriptomes were unknown sequences. Upon ND, level of photosynthetic transcripts, but not photosynthetic efficiency was downregulated. Unlike hydrocarbon contents, ESTs involved in hydrocarbon biosynthesis were not upregulated. Taken together, our results imply that A- and B-races belong to different *B. braunii* subspecies. Upon ND, excess photosynthetic transcripts are recycled for nitrogen; and hydrocarbon accumulation is not via de novo biosynthesis. Here we describe in details the data contents and analytic methodologies associated with the data uploaded to Gene Expression Omnibus (accession number GSE71296).

Specifications	
Organism/cell line/tissue	*Botryococcus braunii*
Sex	Not applicable
Sequencer or array type	Illumina Hiseq 2000, PE90
Data format	Raw data: FQ format; processed data: TXT or XLS format
Experimental factors	N-replete vs. N-depleted
Experimental features	EST levels are based on the de novo assembled transcriptome
Consent	All protocols were provided
Sample source location	Zhoushan, Zhejiang Province, China

1. Direct link to the deposited data

Deposited data can be found at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE71296.

2. Experimental design, materials and methods

2.1. Culture manipulation

The green microalga *Botryococcus braunii* 779 [1] was cultivated using BB (or 2× BB) liquid medium in a 2.5-L low-form flask with shaking at 100 rpm at 25 °C under continuous illumination flux density of ~250 μmol photon m^{-2} s^{-1}. Culture was shaken at 100 rpm at 25 °C (or RT) supplied with 2% CO_2 through bubbling. For analysis of the subisolate in response to ND, the log-phase cells in nitrogen-replete 2× BB medium (or 2BB + N) were harvested by centrifugation at 2000 rcf at 8 °C for 5 min and resuspended in nitrogen-depleted 2× BB medium (or 2BB − N) isovolumetrically. After growing in 2BB − N medium for 3 days, cells were harvested by centrifugation for transcriptomic analysis.

2.2. RNA-seq data acquisition

Total RNA was extracted from *B. braunii* 779 cells prior to and after ND using TRIzol Plus RNA Purification System (Invitrogen-Life Technologies Co., Carlsbad, CA, USA) according to manufacturer's protocol. Approximately 4 μg of the resulting total RNA was used for synthesis of cDNA using the TruSeq RNA Sample Prep Kit (Invitrogen-Life Technologies Co.) according to manufacturer's instruction including synthesis of first- and second-strand cDNA, end repair, 3′-end adenylation, adapter ligation, fragment enrichment (e.g., ~260 bps in length), and library validation, quantification, and quality assessment with a bioanalyzer (Agilent Technologies; Santa Clara, CA, USA). The libraries are sequenced using the Illumina HiSEQ 2000 Sequencer (BGI, Shenzhen, China).

2.3. De novo transcriptome assembly and annotation

Short-read sequences derived from samples prior to (i.e., 779_12-19_I1_1/2.fq) and after (i.e., 779N12-10_I1_1/2.fq) were pooled for

* Corresponding author at: Ocean Research Centre of Zhoushan, Zhejiang University, 10 Tiyu Road, Room 502, Lin-Cheng New District, Zhoushan, Zhejiang 316021, China.

http://dx.doi.org/10.1016/j.gdata.2015.09.019

assembly of transcriptome using Trinity software [2]. Pooled sequences were first subjected to quality control analysis using FastQC (ver 0.11.2). As a result, 12 nucleotides at the 5′-end in all reads and 5 nucleotides at the 3′-end in some reads were trimmed. The minimum length of the output reads was set to 36 nucleotides using Trimomatic (ver 0.32). The resulting reads were subsequently used in de novo transcriptome assembly using Trinity (ver 20140717). A total of 138,295 sequence contigs were generated with the minimal of 300 nucleotides in length. All contigs were subject to redundancy test using CD-Hit. After removal of redundant (i.e., identity >90%) and low counts (i.e., counts per contig <40) contigs, a total number of 61,220 non-redundant high quality contigs or ESTs (i.e., ESTs were used hereafter) were obtained and designated as non-redundant transcriptome.

Subsequently, non-redundant ESTs were subjected to sequence homologous comparison analysis against 6 comprehensively annotated algal genomes (i.e., *Coccomyxa subellipsoidea* C-169 v2, *Chlorella variabilis* NC64A v1, *Chlamydomonas reinhardtii* v4, *Micromonas pusilla* RCC299 v3, *Ostreococcus lucimarinus* v2 and *Thalassiosira pseudonana* CCMP 1335) (http://genome.jgi-psf.org) using BLASTX method. Majority of the best-hits were derived from the *C. subellipsoidea* genome. For simplicity, only 4 well-separated species were shown in Venn diagram (Fig. 1). Based on the cutoff for e-value < 1E-07 and homology percentage > 40%, 12,292 ESTs were found to have a best-hit in the annotated genomes, of which, 8888 (or 72.3%) and 3386 (or 27.5%) were associated with at least one GO function and KEGG ortholog function, respectively.

2.4. Comparative analysis between transcriptomes

Transcriptomes of A-race *B. braunii* Bot-88 and B-race Showa were first processed by removal of redundant ESTs using CD-Hit (i.e., identity >90%). The resulting ESTs were subjected to BLASTN analysis against the transcriptome of *B. braunii* 779. At a threshold of e-value < 1E-05, we found that 1140 shared homology to those in both Bot-88 and Showa, 5860 were unique to Bot-88 [3], and 7489 unique to Showa [4] (Fig. 2).

2.5. Mapping and normalization of EST levels

To assess the effect of nitrogen deprivation, read counts per individual ESTs were normalized to the FPKM (i.e., Fragments Per Kilobase of EST per Million fragments mapped). For this reason, RNA-seq results derived from duplicate samples prior to and after ND were mapped to the *B. braunii* 779 non-redundant transcriptome using Bowtie software. Based on the raw counts, differentially expressed ESTs were obtained by using RSEM and EdgeR software (i.e., level change >2-fold, p-value <0.05).

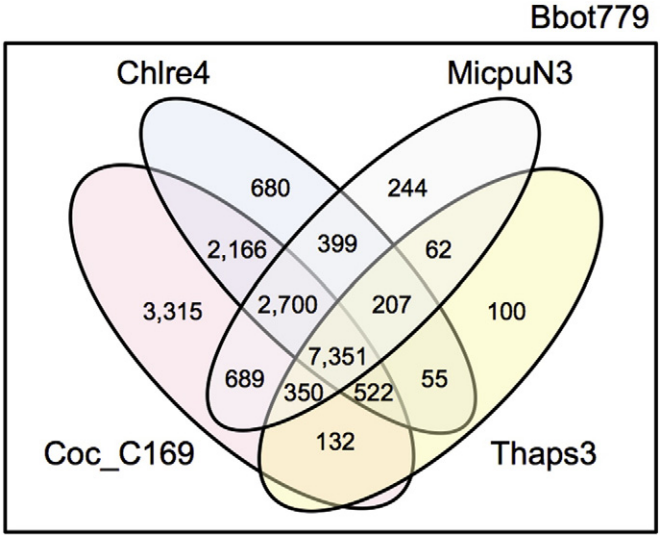

Fig. 1. Venn diagram of *B. braunii* 779 ESTs sharing homology to 4 annotated genomes.

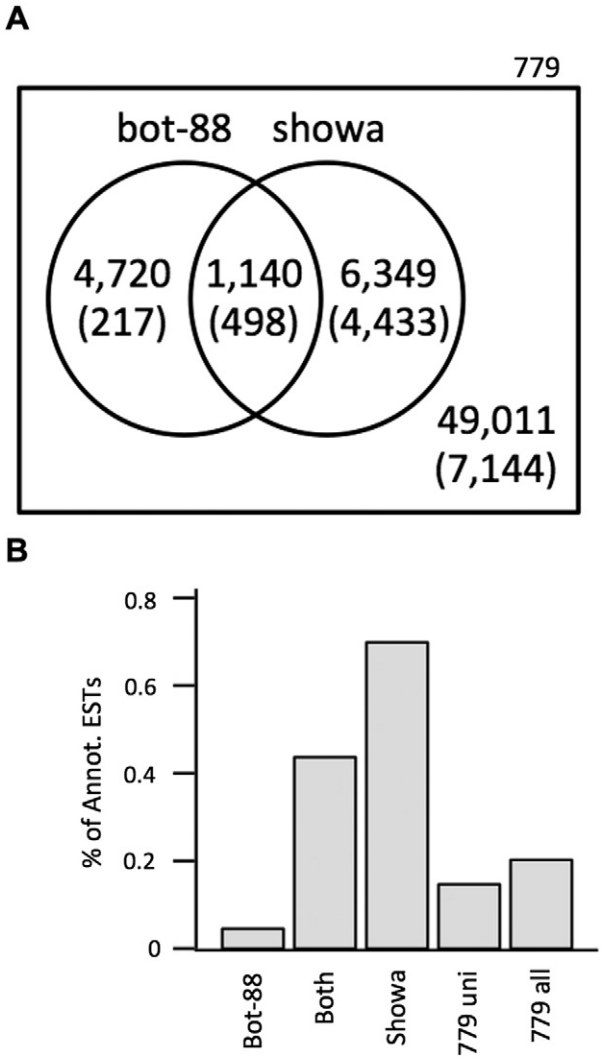

Fig. 2. Comparative analysis of *B. braunii* 779 transcriptome with that of *B. braunii* Bot-88 and Showa. (A) Venn diagram of *B. braunii* 779 ESTs sharing homology to those in *B. braunii* Bot-88 and Showa. Numbers in parentheses indicate the number of annotated ESTs. (B) Percentage of annotated sequences in various groups of ESTs. Percentage of annotated ESTs that share homology to Bot-88 only (or Bot-88), to both Bot-88 and Showa (or Both), to Showa only (or Showa), share no homology to either Bot-88 or Showa (or 779 uni or 779 unique), and 779 background level (or 779 all) are indicated.

2.6. Sliding window analysis

We hypothesized that a subset of co-regulated ESTs might display a coherent level and ratio in actively growing cells and cells in response to environmental stress factors, respectively. In this study, we applied a sliding window approach for identification of potentially co-regulated subsets of ESTs associated with various biological processes or metabolic pathways. In brief, a moving window of 1024 ESTs in size and a sliding step of 512 ESTs in width were applied along a set of 12,288 ESTs sorted based on rank by level or ratio. Significant subsets of ESTs were based on the threshold of fold-change >2 and p-value <0.05 after Bonferroni correction.

3. Discussion

We described here a study of de novo transcriptomic profiling of a non-model microalga *B. braunii* 779 in response to ND. This dataset is

composed of RNA-seq sequencing data derived from cells prior to and after ND using Illumina Hiseq 2000 PE90 methodology. By comparative analysis between transcriptomes of different *B. braunii* strains, we found that many novel genes were present in the A-race *B. braunii* subspecies. Through a sliding window approach, we showed that a subset of highly transcribed ESTs associated with photosynthesis was downregulated upon ND.

References

[1] F. Lei, D. Sun, Z. Xu, J. He, S. Qi, X. Chen, W. Chew, J. Liu, Transcriptomic analysis of a moderately growing subisolate *Botryococcus braunii* 779 (Chlorophyta) in response to nitrogen deprivation. Biotechnol. Biofuels 8 (2015) 130.

[2] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (2011) 644–652.

[3] M. Baba, M. Ioki, N. Nakajima, Y. Shiraiwa, M.M. Watanabe, Transcriptome analysis of an oil-rich race A strain of *Botryococcus braunii* (BOT-88-2) by de novo assembly of pyrosequencing cDNA reads. Bioresour. Technol. 109 (2012) 282–286.

[4] I. Molnar, D. Lopez, J.H. Wisecaver, T.P. Devarenne, T.L. Weiss, M. Pellegrini, et al., Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa). BMC Genomics 13 (2012) 576.