Transcriptome analysis of an oil-rich race B strain of Botryococcus braunii (BOT-22) by de novo assembly of pyrosequencing cDNA reads

著者別名	馬場 将人, 白岩 善博, 渡邉 信
journal or publication title	Bioresource technology
volume	109
page range	292-296
year	2012-04
権利	(C) 2011 Elsevier Ltd. NOTICE: this is the author's version of a work that was accepted for publication in Bioresource technology. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Bioresource technology, Vol.109, Pages:292-296. doi:10.1016/j.biortech.2011.08.104
URL	http://hdl.handle.net/2241/117068

doi: 10.1016/j.biortech.2011.08.104

Transcriptome analysis of an oil-rich race B strain of *Botryococcus braunii* (BOT-22) by *de novo* assembly of pyrosequencing cDNA reads

Motohide Ioki [a, 1], Masato Baba [b, 1], Nobuyoshi Nakajima [a, b, *], Yoshihiro Shiraiwa [b], Makoto M. Watanabe [b]

[a] Environmental Biology Division, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki, 305-8506 Japan.

[b] Graduate School of Life and Environment Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan.

* Corresponding author: E-mail, naka-320@nies.go.jp; FAX, +81-29-850-2490

[1] These authors contributed equally to this work.

Abstract

To gain genetic insights into the biosynthesis of botryococcene oils in *Botryococcus braunii* race B, a transcriptome dataset of the BOT-22 strain containing 27,427 non-redundant sequences assembled from 209,429 complementary DNA reads was obtained via high-throughput 454 sequencing. Relatively reliable prediction of the gene product was feasible for 725 non-redundant sequences based on homology to previously characterized database sequences. Regarding the botryococcene oil biosynthesis, genes putatively associated with the mevalonate-independent isoprenoid biosynthesis pathway were retrieved, while no genes were found for the mevalonate pathway, suggesting that botrococcenes are biosynthesized through the mevalonate-independent pathway in *B.*

braunii. All transcriptome sequences have been deposited in the GenBank/EMBL/DDBJ database.

Keywords: *Botryococcus braunii*; Expressed sequence tag (EST); Hydrocarbon; pyrosequencing; Triterpenoid

1. Introduction

The photosynthetic alga *Botryococcus braunii* produces extraordinarily large amounts of hydrocarbon oils. *B. braunii* strains are grouped into different races, namely, race A, B, and L, depending on the types of liquid hydrocarbons they produce. The race A strains produce alkadienes and alkatrienes, while the race B strains produce triterpenoid hydrocarbons named botryococcenes and methylated squalenes. The race L strains synthesize tetraterpenes referred to as lycopadienes (reviewed in Banerjee et al., 2002; Metzger and Largeau, 2005).

Isopentenyl-diphosphate (IPP) is required for the race B hydrocarbon biosynthesis. IPP is known to be biosynthesized through the mevalonate-independent (MVI) pathway rather than the mevalonate (MVA) pathway in race B strains of *B. braunii* (Sato et al., 2003). Subsequently, triterpenes are formed from IPP and its isomer dimethylallyl-diphosphate (DMPP) through the triterpene synthesis pathway. Squalene synthase-like enzymes catalyze the C30 botryococcene formation (Niehaus et al., 2011). Then the C30 botryococcenes are often methylated using S-adenosyl methionine (SAM) as a methylating agent (Wolf et al., 1985). The genes encoding enzymes associated with the hydrocarbon oil biosynthesis are largely unknown with only few exceptions

including the squalene synthase gene (Okada et al., 2000; Niehaus et al., 2011).

In this study, expressed sequence tags (ESTs) of a representative race B strain of *B. braunii*, BOT-22, was anlyzed. The BOT-22 strain was used because of its excellence in both growth and oil productivity. This strain produces a triterpenoid-derived $C_{34}H_{58}$ hydrocarbon (personal communication with Takako Tanoi, Masanobu Kawachi, and Kunimitsu Kaya).

2. Materials and Methods

The original axenic culture of the BOT-22 strain of *B. braunii* was provided from Dr. Masanobu Kawachi (National Institute for Environmental Studies). The cells were grown as described in Ioki et al. (2011, this issue) with bubbling of filtered air. Cells were harvested using a 5-μm filter and immediately frozen in liquid nitrogen.

The Genome Sequencer-FLX System (Roche diagnostics, Basel, Switzerland) was used for the EST acquisition. Total RNA was extracted from cells harvested from a 250-ml culture using RNeasy Plant Mini Kit (Qiagen Inc., Chatsworth, CA, USA). Contaminating DNA was eliminated using RNase-Free DNase Set (Qiagen Inc.). Amplification of antisense RNA (aRNA) was performed as described by Vega-Arreguín et al. (2009) using 2 μg of total RNA. Reverse transcription was performed using 10 μg of aRNA and random 6mers. Then double-stranded cDNA was sytnthesized using biotinylated oligo(dT) primers with the GS-FLX adaptor. The aRNA-derived cDNA was fragmented by sonication and separated by electrophoresis on an agarose gel. Fragments of 300 to 800 bp were excised out of the gel. Then, 3'-terminal fragments were collected using streptavidin-coated magnetic beads (Dynal, Invitrogen, Carlsbad,

CA). On the 5'-ends of these fragments, the other GS-FLX adaptors were conjugated. sstDNA was collected using magnetic beads. The resulting cDNA library was quantified by fluorometry (Quant-iT Ribogreen, Invitrogen), clonally amplified by emulsion PCR and sequenced according to the standard procedures for 454 sequencing. The cDNA reads were filtered, clustered, and assembled into non-redundant sequences using the Paracel TranscriptAssembler^{TM} Version 2.6 software (Paracel, Pasadena, CA) under the auspices of the Dragon Genomics Center, Takara Bio Co. Ltd, Japan.

The non-redundant sequences were annotated and classified into different functional categories using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) program (Moriya et al., 2007) based on the BLASTX algorithm. The EST count for each hydrocarbon biosynthesis-related gene was determined according to the enzyme nomenclature (accepted and alternative names) of the BLASTX search hits from the nr database (ver. 2009.01.09) compiled by National Center for Biotechnology Information (NCBI) using the EST Viewer software (Dragon Genomics Center, Takara Bio Co. Ltd, Japan). All the BLASTX searches were performed with cutoff bit score of 50, which corresponds to the *e*-value of approximately 1×10^{-5}.

3. Results and Discussion

The pyrosequencing generated 209,429 cDNA reads. Majority of the cDNA reads were of 200 to 300 bases in length. The average length of reads was 202 bases (Fig. 1a). The cDNA reads are available in the DDBJ Sequence Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml) as DRR000584.

By assembling the cDNA reads, 27,427 non-redundant sequences of up to over 2 kb in length were retrieved. The average length of non-redundant sequences was 267 bases (Fig. 1b). Majority of the non-redundant sequences consisted of a single cDNA read, while some non-redundant sequences consisted of more than 2,000 cDNA reads (Fig. 1c). For 725 non-redundant sequences, the gene functions were successfully predicted based on homology to previously characterized genes of other organisms (Table S1). More than 50% of the 725 non-redundant sequences were predicted to be associated with metabolism, and many of the metabolic genes were predicted to be associated with terpenoid biosynthesis and steroid biosynthesis, reflecting the triterpene-producing nature of the BOT-22 strain (Table 1). The non-redundant sequences can be found in the DDBJ database (http://www.ddbj.nig.ac.jp/index-e.html) under accession numbers FX085123 through FX 112549.

The top 30 abundant ESTs were listed in Table 2. ESTs for genes associated with the photosynthetic apparatus were abundant (FX085489 [614 reads] for the light-harvesting complexII chlorophyll a/b binding protein 3 gene). A non-redundant sequence putatively encoding a carbonic anhydrase, which is involved in CO_2 concentrating mechanisms in most microalgae (Giordano et al., 2005), also consisted of a large number of pyrosequencing reads (FX085368, 793 reads) (Table 2).

Search for genes encoding enzymes potentially associated with botryococcene biosynthesis from CO_2 retrieved 17 non-redundant sequences (Table 3). The accession numbers of these non-redundant sequences are shown in Table S2. The BOT-22 transcriptome profile provided genetic evidence of the prevalence of botryococcene biosynthesis through the MVI pathway rather than the MVA pathway.

D-glyceraldehyde 3-phosphate, a starting metabolite of the MVI pathway, is

synthesized from 1, 3-bisphosphoglycerate by glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) [EC 1.2.1.12], glyceraldehyde-3-phosphate dehydrogenase (NADP(+)) (phosphorylating) [EC 1.2.1.13], or glyceraldehyde-3-phosphate dehydrogenase (NAD(P)(+)) (phosphorylating) [EC 1.2.1.59] (Martin and Cerff., 1986; Valverde et al., 1997; Rohmer, 1999). The BOT-22 transcriptome dataset contained a non-redundant sequence putatively encoding glyceraldehyde-3-phosphate dehydrogenase (NADP(+)) (phosphorylating) (FX085420, 344 reads) (Table 3, Table S2). This seemed to function in the reductive pentose phosphate cycle of photosynthesis because of its high similarity to the chloroplastic glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) (*e*-value was $4e^{-26}$) (Table S2). In addition, ESTs were found for glyceraldehyde-3-phosphate dehydrogenases (phosphorylating) (FX085422, 162 reads; FX085421, 3 reads), which seemed to participate in glycolysis because of its similarity to cytosolic glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) ($3e^{-47}$ and $3e^{-51}$, respectively).

Five non-redundant sequences were found for 4 glycolitic enzymes: phosphoglycerate kinase [2.7.2.3] (FX085563, 6 reads), phosphoglycerate mutase [5.4.2.1] (FX085137, 6 reads), phosphopyruvate hydratase [4.2.1.11] (FX085138, 2 reads), and pyruvate kinase [2.7.1.40] (FX085139, 4 reads; FX085140, 1 read). These transcripts seem to be associated with synthesis of pyruvate, which is the other starting substrate of the MVI pathway (Table 3, Table S2).

For the MVI pathway, the BOT-22 EST dataset contained non-redundant sequences putatively encoding 1-deoxy-D-xylulose-5-phosphate synthase [EC 2.2.1.7] (FX085274, 2 reads; FX085276, 5 reads; FX085275, 1 read; FX085277, 1 read)

catalyzing the formation of 1-deoxy-D-xylulose-5-phosphate from D-glyceraldehyde 3-phosphate and pyruvate in the first step in the MVI pathway and (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase [EC 1.17.7.1] (FX085293, 3 reads) catalyzing the reaction in the midst of the MVI pathway. A non-redundant sequence encoding a protein resembling squalene synthase [EC 2.5.1.21] (FX085141, 3 reads) was also found (Table 3, Table S2).

ESTs were found for 2 enzymes involved in the synthesis of acetyl-CoA, the starting substrate for the MVA pathway, from pyruvate: pyruvate dehydrogenase (acetyl-transferring) [EC 1.2.4.1] (FX085264, 4 reads; FX085136, 1 read; FX085135, 1 read) and dihydrolipoyllysine-residue acetyltransferase [EC 2.3.1.12] (FX085264, 4 reads) (Table 3, Table S2). However, no EST was found for enzymes in the MVA pathway (Table 3). These transcripts suggest that a considerable amount of photosynthetically fixed carbon is channeled through the reductive pentose phosphate cycle eventually forming acetyl-CoA from D-glyceraldehyde 3P providing substrates for fatty acid synthesis and the citric acid cycle. Absence of ESTs associated with the MVA pathway is consistent with the results of biochemical experiments showing that the MVA pathway has low or no activity in *B. braunii* (Sato et al., 2003). Retrieval of no ESTs may indicate weak gene expression. Alternatively, the genes may not exist in *B. braunii*. Many of the mevalonate pathway-related genes are missing in the sequenced genomes of green algal species such as *Chlamydomonas reinhadtii* (Merchant et al., 2007).

4. Conclusions

De novo pyrosequencing and assembly was performed on the oil-rich representative race B strain of *B. braunii*. The resulting transcriptome presented here contains genes encoding enzymes associated with the race B hydrocarbon biosynthesis. These transcripts strongly suggest the mevalonate-independent production of hydrocarbons in race B. The novel EST dataset acquired in this research is available in GenBank/EMBL/DDBJ database.

Acknowledgements

Takako Tanoi, Yurie Akutsu, and Haniyeh Bidadi (National Institute for Environmental Studies) provided technical assistance. This research was supported by the Core Research for Evolutionary Science and Technology program of Japan Science and Technology Agency.

References

Banerjee, A., Sharma, R., Chisti, Y., Banerjee, U.C., 2002. *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals. Crit Rev Biotechnol. 22(3), 245-279.

Giordano, M., Beardall, J., Raven, JA., 2005. CO_2 Concentrating Mechanisms in Algae: Mechanisms, Environmental Modulation, and Evolution. Annu. Rev. Plant Biol. 56, 99-131.

Ioki, M., Ohkoshi, M., Nakajima, N., Nakahira-Yanaka, Y., Watanabe, M.M., 2011. Isolation of herbicide-resistant mutants of *Botryococcus braunii*. Bioresour. Technol., this issue, in press, http://dx.doi.org/10.1016/j.biortech.2011.07.101.

Martin, W., Cerff, R., 1986. Prokaryotic features of a nucleus-encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (Sinapis alba). Eur. J. Biochem. 159(2), 323-331.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., Marshall, W.F., Qu, L.H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.P., Riaño-Pachón, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I.,

Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Martínez, D., Ngau, W.C., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., Grossman, A.R., 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318, 245-250.

Metzger, P., Largeau, C., 2005. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. Appl Microbiol Biotechnol. 66(5), 486-496.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A., Kanehisa, M., 2007. KAAS, an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182-185.

Niehaus, T.D., Okada, S., Devarenne, T.P., Watt, D.S., Sviripa, V., Chappell, J. 2011. Identification of unique mechanisms for triterpene biosynthesis in *Botryococcus braunii*. Proc Natl Acad Sci U S A. 108(30), 12260-12265.

Okada, S., Devarenne, T.P., Chappell, J., 2000. Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, race B. Arch. Biochem. Biophys. 373(2), 307-317.

Rohmer, M., 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. Nat Prod Rep. 16, 565–574.

Sato, Y., Ito, Y., Okada, S., Murakami, M., Abe, H., 2003. Biosynthesis of the

triterpenoids, botryococcenes and tetramethylsqualene in the B race of *Botryococcus braunii* via the non-mevalonate pathway. Tetrahedron Lett. 44, 7035-7037.

Valverde, F., Losada, M., Serrano, A., 1997. Functional complementation of an Escherichia coli gap mutant supports an amphibolic role for NAD(P)-dependent glyceraldehyde-3-phosphate dehydrogenase of Synechocystis sp. strain PCC 6803. J. Bacteriol. 179(14), 4513-4522.

Vega-Arreguín, J.C., Ibarra-Laclette, E., Jiménez-Moraila, B., Martínez, O., Vielle-Calzada, J.P., Herrera-Estrella, L., Herrera-Estrella, A., 2009. Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. BMC Genomics 10, 299.

Wolf, F.R., Nemethy, E.K., Blanding, J.H., and Bassham, J.A., 1985. Biosynthesis of unusual acyclic isoprenoids in the green alga *Botryococcus braunii*. Phytochemistry 24**,** 733.

Figure Captions

Figure 1 Features of the BOT-22 EST dataset generated by pyrosequencing. a, Distribution of cDNA read lengths. **b**, Histogram for length distribution of non-redundant sequences obtained after assembling of cDNA reads. **c**, Histogram representation of distribution of EST counts.

P.S. Please check Fig and tables at

http://dx.doi.org/10.1016/j.biortech.2011.08.10