J. Phycol. **46,** 534–540 (2010)

DOI: 10.1111/j.1529-8817.2010.00820.x

PHYLOGENETIC PLACEMENT, GENOME SIZE, AND GC CONTENT OF THE LIQUID-HYDROCARBON-PRODUCING GREEN MICROALGA *BOTRYOCOCCUS BRAUNII* STRAIN BERKELEY (SHOWA) (CHLOROPHYTA)[1]

Taylor L. Weiss

Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128, USA

J. Spencer Johnston

Department of Entomology, Texas A&M University, College Station, Texas 77843-2475, USA

Kazuhiro Fujisawa,[2] *Koremitsu Sumimoto, Shigeru Okada*

Laboratory of Aquatic Natural Products Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Joe Chappell

Department of Plant & Soil Sciences, Plant Biology Program, University of Kentucky, Lexington, Kentucky 40546, USA

and Timothy P. Devarenne[3]

Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128, USA

We report the genome size and the GC content, and perform a phylogenetic analysis on *Botryococcus braunii* Kütz., a green, colony-forming, hydrocarbon-rich alga that is an attractive source for biopetroleum. While the chemistry of the hydrocarbons produced by the B race of *B. braunii* has been studied for many years, there is a deficiency of information concerning the molecular biology of this alga. In addition, there has been some discrepancy as to the phylogenetic placement of the Berkeley (or Showa) strain of the B race. To clarify its classification, we isolated the Berkeley strain nuclear SSU (18S) rRNA gene and β-actin cDNA and used these sequences for phylogenetic analysis to determine that the Berkeley strain belongs to the Trebouxiophyceae class. This finding is in agreement with other B races of *B. braunii*, indicating the Berkeley strain is a true B race of *B. braunii*. To better understand molecular aspects of *B. braunii*, we obtained the Berkeley strain genome size as a first step in genome sequencing. Using flow cytometry, we determined the *B. braunii* Berkeley genome size to be 166.2 ± 2.2 Mb. We also estimated the GC content of the Berkeley strain as 54.4 ± 1.2% for expressed gene sequences.

Key index words: **18S rRNA sequences; *Botryococcus braunii*; Chlorophyta; GC content; genome size; hydrocarbons; phylogenetic analysis; Trebouxiophyceae**

Abbreviations: **18S rRNA, nuclear SSU (18S) ribosomal RNA; cDNA, complementary DNA; RT–PCR, reverse transcription–PCR**

B. braunii is a green, colonial microalga with unique liquid-hydrocarbon biosynthetic capabilities that have made this organism a promising source of renewable hydrocarbon fuels. The cells of a *B. braunii* colony are held together by an extracellular matrix composed of a polymer core of aldehydes derived from very long-chain fatty acids (Maxwell et al. 1968, Knights et al. 1970). Although some *B. braunii* liquid hydrocarbons can be found intracellularly, most liquid hydrocarbons are retained within the colony extracellular matrix.

Three distinct races of *B. braunii* (A, B, and L) are classified by the type of hydrocarbons occurring in the extracellular matrix. The A race accumulates the fatty acid–derived alkadienes and alkatrienes, while the L race accumulates the tetraterpene lycopadiene. The focus of this study, the B race, accumulates the triterpenoid hydrocarbons known as botryococcenes (Banerjee et al. 2002, Metzger and Largeau 2005).

B. braunii possesses many characteristics that make it attractive as a biofuel feedstock source. *B. braunii* has been observed forming massive, dense freshwater blooms, and it is suggested that *B. braunii* oils have contributed to oil deposits throughout the world (Traverse 1955, Gelpi et al. 1968, Wake

[1]Received 10 April 2009. Accepted 7 December 2009.
[2]Present address: Bristol-Myers Squibb, Shinjuku i-Land Tower 5-, Nishi-Shinjuku-6, Shinjuku, Tokyo 163-1328, Japan.
[3]Author for correspondence: e-mail tpd8@tamu.edu.

and Hillen 1980, 1981). Large quantities of botryococcene derivatives, originating from the B race, are found in current petroleum deposits (Moldowan and Seifert 1980). In addition, botryococcenes can be processed by the same methods used on petroleum (caustic hydrolysis) to yield petroleum equivalent products including gasoline, diesel, and kerosene (jet fuel) (Hillen et al. 1982, Kitazato et al. 1989). Finally, the *B. braunii* B-race strains typically accumulate hydrocarbons to 30%–40% of their dry weight and are capable of obtaining hydrocarbon contents up to 86% dry weight (Brown et al. 1969).

Some disagreement exists about the phylogenetic placement of the Berkeley strain of *B. braunii*. Cell morphology was the original standard used to establish *B. braunii* taxonomy (Komárek and Marvan 1992). Subsequent studies have been based on molecular phylogenetics using nuclear SSU (18S) rRNA genes (Sawayama et al. 1995, Senousy et al. 2004). The first molecular phylogenetic studies placed the Berkeley strain of *B. braunii* race B within the class Chlorophyceae (Sawayama et al. 1995). However, a follow-up study using the 18S rRNA gene sequences from one B race (Ayamé strain), two A races (CCAP 807/1 and Titicaca strains), and one L race (Songkla Nakarin strain) of *B. braunii* placed these races in the class Trebouxiophyceae (Senousy et al. 2004). This study by Senousy et al. (2004) included the original Berkeley strain 18S rRNA gene sequence from Sawayama et al. (1995) and left the Berkeley strain in the Chlorophyceae (Senousy et al. 2004). This placement of the Berkeley strain in the Chlorophyceae is possibly an artifact, since the original 18S rRNA gene isolated (Sawayama et al. 1995) was likely amplified from a contaminating alga belonging to the Chlorophyceae (Senousy et al. 2004).

Remarkably, of the estimated 200,000 algae species worldwide (Chapman 2005), only six species have had their genomes fully sequenced and annotated (Matsuzaki et al. 2004, Misumi et al. 2008). As a group, algae may be the only photosynthetic organisms capable of producing enough biofuel to meet transportation fuel demands (Chisti 2007, 2008a,b). However, algae are currently not able to meet oil production needs primarily due to a lack of a molecular-level understanding of algal cellular processes, which would allow genetic manipulation for increased oil production, growth rates, and/or photosynthetic capabilities (Chisti 2007). Genome sequencing of additional algal species will lay the foundation to better understand their molecular biology. As a beginning step in obtaining the *B. braunii* genome sequence, we have determined the genome size and GC content estimate of the Berkeley strain of *B. braunii* race B. We have also confirmed the phylogenetic placement of the Berkeley strain in the class Trebouxiophyceae with the other *B. braunii*.

MATERIALS AND METHODS

Culturing of algae. A pure culture of *B. braunii* Berkeley (Showa) B race (Nonomura 1988) was obtained through a combination of culture dilution with sterilized media and isolation of individual colonies under a microscope, which were transferred to fresh, sterile media for growth. Cultures were grown in modified Chu 13 media (Grung et al. 1989) using 13 W compact fluorescent 65 K lighting at a distance of 7.62 cm, which produced a light intensity of 280 μmol photons $\cdot$ m^{-2} $\cdot$ s^{-1}. Lighting was a cycle of 12:12 light:dark (L:D) at 22.5°C. The cultures were continuously aerated with filter-sterilized, enriched air containing 2.5% CO_2. Fifty milliliters of culture was used to inoculate 750 mL of subsequent subcultures every 4 weeks. The remaining culture volume was harvested by vacuum filtration using 35 μM nylon mesh (Aquatic Ecosystems Inc., Apopka, FL, USA). The accumulated colonies were rinsed with sterilized dH_2O, frozen in liquid nitrogen, and stored at −80°C for future analysis.

rDNA amplification. To reproduce the experiments carried out by Sawayama et al. (1995), the *B. braunii* Berkeley 18S rRNA gene was amplified by reverse transcription–PCR (RT–PCR) using total RNA. An algal sample from a pure Berkeley culture was ground in a mortar and pestle chilled with liquid nitrogen, and total RNA isolated by the SDS/LiCl method (López-Gómez and Gómez-Lim 1992, Okada et al. 2000). First-strand cDNA was synthesized from total RNA using Superscript II (Invitrogen, Carlsbad, CA, USA) and random hexamers. The first-strand cDNA was then used as a template to amplify the nuclear 18S rRNA gene by PCR in two overlapping sections using the same combinations of primers used by Sawayama et al. (1995)—primer combination 1,5′-TACCTGGTTGATCCTGCCAGTAG and 5′-CCAATCCCTAGTCGGCATCGT; and primer combination 2, 5′-AGATACCGTCGTAGTCTCAACCA and 5′-ACCTTGTTACGACTTCTCCTTCCTC. PCR was carried out with 1 unit of KOD plus DNA polymerase (Toyobo Co., Ltd., Osaka, Japan) in the supplied 1× PCR buffer, 1 mM $MgSO_4$, 0.2 mM dNTPs, 0.3 μM of each primer, and 1 μL of first-strand cDNA in a final volume of 50 μL. PCR cycles were as follows: initial incubation at 94°C for 2 min followed by 30 cycles of 94°C for 15 s, 61°C (for primer combination 1) or 59°C (for primer combination 2) for 30 s, and 68°C for 2 min. PCR products were isolated from agarose gels using MagExtractor (Toyobo), blunt ligated into the EcoRV site of pBluescript II KS+ (Stratagene, La Jolla, CA, USA), transformed into *Escherichia coli* XL-1 Blue MRF' strain (Stratagene) according to a standard $CaCl_2$ transformation procedure, and sequenced on both strands by the dideoxy-terminator method using a Thermo Sequenase Primer Cycle Sequencing kit (GE Healthcare, Piscataway, NJ, USA) in a Shimadzu DQ2000 DNA sequencer (Shimadzu, Kyoto, Japan). The isolated *B. braunii* Berkeley 18S rDNA sequence reported here was submitted to GenBank, accession number FJ807044.

Cloning of β-actin cDNA. The *B. braunii* Berkeley β-actin cDNA was isolated using standard PCR techniques as described above with 1 μL of a 1:100 dilution of a *B. braunii* Berkeley cDNA library (Okada et al. 2000) as a template, degenerate primers based on plant actin sequences, and Taq polymerase (Toyobo). Degenerate primer sequences were as follows: forward primer, 5′-ATGACGCARATHATGTTYGAR; reverse primer, 5′-CCACATYTGYTGRAANGT. PCR cycles were as follows: initial incubation at 94°C for 4 min followed by 30 cycles of 94°C for 30 s, 43°C for 30 s, and 72°C for 1 min, and one final cycle of 72°C for 7 min. The generated 699 bp PCR product was gel purified, blunt-end cloned, and sequenced as described above for the 18S rRNA gene. The resulting *B. braunii* Berkeley β-actin sequence was submitted to GenBank, accession number GU049780.

Phylogenetic analysis. The 18S rRNA gene sequences used for phylogenetic analysis were those produced here plus 58

obtained from GenBank based on accession numbers given in Senousy et al. (2004). The final 18S rRNA alignment used by Senousy et al. (2004) for phylogenetic analysis was obtained from TreeBASE (http://treebase.org/treebase/); a new alignment including the newly isolated *B. braunii* Berkeley 18S rRNA cDNA sequence (accession FJ807044) was produced using MEGA 4.1 (Tamura et al. 2007). The β-actin gene sequences used for phylogenetic analysis are that produced here plus the following sequences obtained from GenBank: AB080314.1 (*Trebouxia erici*), AB046453.1 (*Nannochloris* sp. SAG 251-2), AB046457.1 (*Parachlorella kessleri*), AB046451.1 (*Marvania coccoides*), AB292587.1 (*Marvania geminate*), 3107918 (*Nannochloris bacillaris*), AB080313.1 (*Chlorella vulgaris*, actin-2), AB080312.1 (*Chlorella vulgaris*, actin-1), AB292588.1 (*Lobosphaeropsis lobophora*), AB080311.1 (*Chlorella sorokiniana*), AB053215.1 (*Pseudochlorella* sp. CCAP 211/1A), 915193 (*Chlamydomonas reinhardtii*), 14719361 (*Dunaliella salina*), 170648 (*Volvox carteri*), 34581635 (*Chlamydomonas moewusii*), AF061018.1 (*Scherffelia dubia*), and 50355608 (*Ulva pertusa*). Organism synonyms reflect current species taxonomies as identified by DNA Data Bank of Japan (DDBJ)/European Molecular Biology Laboratory (EMBL)/GenBank accession number at the time of publishing.

Trees were constructed using MEGA 4.1 (Tamura et al. 2007), MrBayes (Hall 2001, Huelsenbeck and Ronquist 2001), and PhyMl 3.0 (Guindon and Gascuel 2003). Trees were drawn with MEGA 4.1 (Tamura et al. 2007) and TreeView (Page 1996). Data analysis methods included distance (neighbor joining [NJ] and minimum evolution [ME]), maximum parsimony (MP), and likelihood (maximum likelihood [ML] and Bayesian inference [BI]). For rRNA phylogeny, sequences of *Nephroselmis olivacea* and *Pseudoscourfieldia marina* remained the rooted outgroup (Senousy et al. 2004), and for β-actin phylogeny, sequences of *Scherffelia dubia* and *Ulva pertusa* provided the rooted outgroup. All bootstrap analyses were carried out with 500 replications each. All methods were applied with the default model of their respective programs, with the exception of ML, which was tested using both the HKY85 and generalized time-reversible (GTR) substitution models. The β-actin phylogeny was additionally assessed using translated amino acid sequences and nucleotide sequences excluding codon third positions. All remaining parameters were the same or equivalent to those found in Senousy et al. (2004). The *B. braunii* Berkeley rRNA and β-actin alignments have been submitted to TreeBASE (study accession no.: S2526; matrix accession no.: M4827, M4828).

Flow cytometry for genome size estimation. Genome size was estimated using a procedure modified from Johnston et al. (2004, 2005, 2007). Fifty milligrams of frozen *B. braunii* cells was resuspended in 500 μL of Galbraith buffer. This solution was treated with 1 mg of lyticase (Sigma, St. Louis, MO, USA) at room temperature with occasional gentle mixing for 20 min. The mixture was then spread across a glass slide, and the aggregated cells thoroughly chopped through with a new razor blade. The liquid was collected from the glass slide and was then passed through a 50 μm filter and diluted with Galbraith buffer (Galbraith et al. 1983) to 1 mL. The sample was then combined with the head from a female wildtype strain of *Drosophila virilis*, which served as a standard (2C = 333.3 Mb; Gregory and Johnston 2008). Clean, isolated nuclei from the algae and standard were released by grinding the chopped algae plus the *D. virilis* head in Galbraith buffer using 15 strokes of the A pestle in a Kontes 2 mL Dounce homogenizer. A *D. virilis*–only control was similarly prepared. The mixture was passed through a 50 μm filter, stained with 50 ppm propidium iodide (PI; Sigma), incubated for 30 min at 4°C in the dark, and run on a Beckman/Coulter Elete flow cytometer (Beckman/Coulter, Brea, CA, USA) with the laser emitting 25 mW of exciting light at 488 nm. Red fluorescence from PI (intercalated into the DNA of the 2C and 4C nuclei of *D. virilis* and sample) was detected using a high bandpass filter (615 nm). DNA content was determined by multiplying the ratio of the mean peak fluorescence of the 2C sample by that of the standard times 333.3 Mb. Nine Berkeley genome size estimations were carried out over a 3-week period, and a 1C average and standard error calculated based on the produced genome sizes.

GC-content analysis. A *B. braunii* cDNA library was previously created using a ZAP-cDNA Synthesis Kit (Stratagene; Okada et al. 2000). The library was plated, 15 random clones isolated, the contained cDNA sequenced using BigDye Terminator cycle sequencing (Applied Biosystems, Foster City, CA, USA), and gene identity determined using BLAST (Altschul et al. 1990) against the GenBank database. The 15 sequences showed high similarity with known or putative genes of green algae, most notably *Chlamydomonas reinhardtii*. These 15 genes were analyzed using the DNA sequence analysis program Seqool (http://www.biossc.de/seqool/index.html), and the bases were summed (>25 kb total) to obtain a percent GC content and standard error for the overall GC content.

RESULTS AND DISCUSSION

Phylogenetic placement of B. braunii *Berkeley.* Based on the DNA sequences of rRNAs (rDNA sequence), previous studies have shown that the Berkeley strain of the B race of *B. braunii* is phylogenetically distinct from the other *B. braunii* (Sawayama et al. 1995, Senousy et al. 2004). However, it had been suggested that the 18S rDNA sequence used for this analysis was isolated from an alga other than *B. braunii* Berkeley that was a contaminant in the culture (Senousy et al. 2004). Thus, we generated a monoculture of the Berkeley strain of *B. braunii* and used RT–PCR to isolate an 18S rDNA sequence from *B. braunii* Berkeley. DNA alignments indicate that the newly isolated Berkeley 18S rDNA sequence (accession FJ807044) is distinct from the original Berkeley 18S rDNA sequence (accession X78276; Sawayama et al. 1995; Fig. S1 in the supplementary material). This result supports the suggestion that the original Berkeley 18S rDNA sequence was from a contaminating algae species, probably from the *Dunaliella* clade in the Chlamydomonadales (Sawayama et al. 1995).

To further support that our newly isolated *B. braunii* Berkeley 18S rDNA sequence does indeed originate from a *B. braunii* strain, we performed a phylogenetic analysis similar to that of Senousy et al. (2004) using NJ, ME, ML, BI, and MP methods. All five methods gave nearly identical results overall (data not shown). The phylogenetic tree based on the ML method is shown in Figure 1 with bootstrap values from ML, BI, and MP shown. This analysis shows that the original Berkeley 18S rDNA sequence (accession X78276) clusters within the *Dunaliella* clade of the class Chlorophyceae as was reported earlier (Fig. 1; Senousy et al. 2004). Our new Berkeley 18S rDNA sequence (accession FJ807044) was placed in a cluster with the other *B. braunii* 18S rDNA sequences forming a monophyletic group within the class Trebouxiophyceae (Fig. 1). As reported previously (Senousy et al. 2004), *B. braunii*

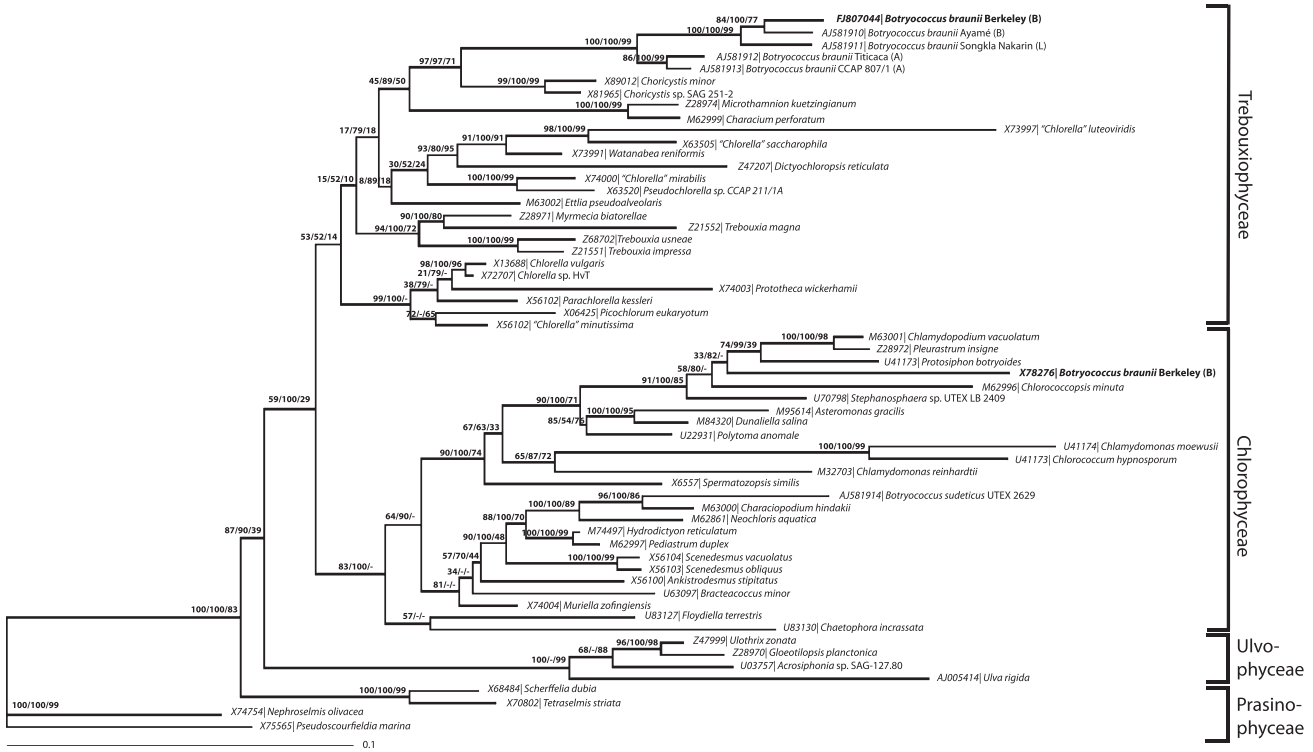

FIG. 1. Maximum-likelihood (ML) phylogenetic tree of 18S rRNA sequences. Sequences are identified by their corresponding GenBank, DNA Data Bank of Japan (DDJB), or European Molecular Biology Laboratory (EMBL) accession numbers and organism name. Branching credibility percentages from the ML method (left), Bayesian inference (middle), and bootstrap maximum parsimony (right) are listed. Where differential branching from the ML method occurs, confidence values for the dissenting method are represented by a dash. The focus of this study, *Botryococcus braunii* Berkeley (FJ807044 and X78276), is shown in bold. The scale bar is for 0.1 substitutions per site.

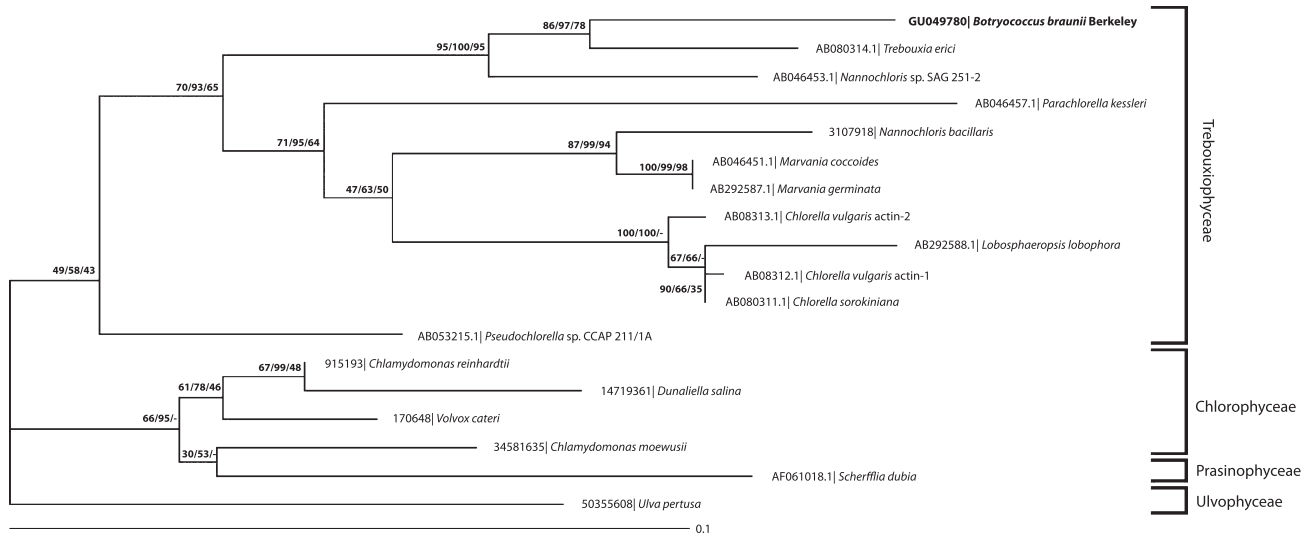

FIG. 2. Maximum-likelihood (ML) phylogenetic tree of actin nucleotide sequences (first two codon positions only). Sequences are identified by their corresponding GenBank, DNA Data Bank of Japan (DDJB), or European Molecular Biology Laboratory (EMBL) accession numbers and organism name. Branching credibility percentages from the ML method (left), Bayesian inference (middle), and bootstrap maximum parsimony (right) are listed. Where differential branching from the ML method occurs, confidence values for the dissenting method are represented by a dash. The focus of this study, *Botryococcus braunii*, Berkeley (GU049780), is shown in bold. The scale bar is for 0.1 substitutions per site.

forms two lineages within this group—one containing the A race, and one containing the B and L races (Fig. 1). This coincides with the nature of the hydrocarbons produced by these races (Senousy et al. 2004). The A race produces fatty acid–derived hydrocarbons, while the B and L races produce isoprenoid-based hydrocarbons (Banerjee et al. 2002, Metzger and Largeau 2005). Importantly, our new *B. braunii* Berkeley 18S rDNA sequence forms a lineage with the B race, Ayamé strain, separate from that of the L race, Songkla Nakarin strain, which was not given in the previous analysis (Fig. 1; Senousy et al. 2004).

To further support our contention that we are working with a monoculture, we isolated the β-actin cDNA from our *B. braunii* Berkeley strain cultures and used this sequence for phylogenetic analysis using ML, BI, and MP. The phylogenetic tree based on ML is shown in Figure 2 and indicates that the *B. braunii* Berkeley sequence groups with other algae from the Trebouxiophyceae confirming the placement of *B. braunii* Berkeley based on our 18S rDNA sequence. We were not able to speculate on the relationship of the Berkeley strain with other races of *B. braunii* as β-actin sequences from other races and strains of *B. braunii* were not available. Taken together, our phylogenetic analyses give strong evidence that the Berkeley strain we have used in our analysis is indeed a pure culture of *B. braunii* B-race strain and belongs in the class Trebouxiophyceae.

Genome size determination of B. braunii *Berkeley.* To make algae a viable source of oil for production of biofuels, more information is needed about the molecular biology of these organisms to understand basic cellular processes (Chisti 2007). Toward this end, we used flow cytometry (Johnston et al. 2004, 2005, 2007) to estimate the genome size of *B. braunii* Berkeley as a first step in determining the whole genome sequence of the Berkeley strain. We observed that *B. braunii* Berkeley has a genome size of 166.2 ± 2.2 Mb (mean ± SE; $n = 7$; Fig. 3). This translates to a 1C DNA content of 0.17 pg, based on 1 pg of DNA = 980 Mb (Bennett et al. 2000).

The estimated *B. braunii* Berkeley genome size is larger than any of the six completed and annotated algal genomes: *Chlamydomonas reinhardtii*, 121.0 Mb; *Osterococcus tauri*, 12.6 Mb; *Osterococcus lucimarinus*,

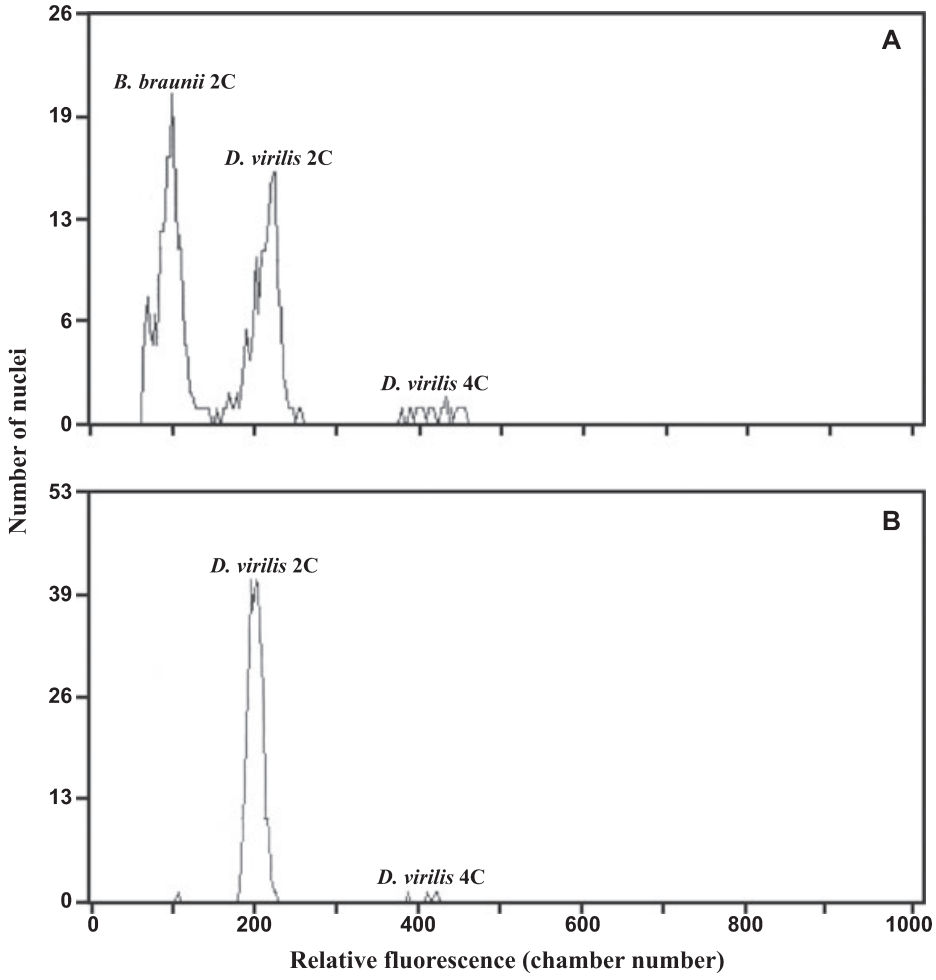

FIG. 3. Flow cytometry analysis of *Botryococcus braunii* Berkeley for genome size determination. Diagrams show the number of nuclei with differing levels of red fluorescence from propidium iodide binding to DNA of (A) 2C nuclei of *B. braunii* Berkeley, and 2C and 4C nuclei of *Drosophilia virilis*; and (B) 2C and 4C nuclei of *D. virilis* only.

13.2 Mb; *Cyanidioschyzan merolae*, 16.5 Mb; *Thalassiosira pseudonana*, 31.3 Mb (Misumi et al. 2008); and *Phaeodactylum tricornutum*, 26.5 Mb (Bowler et al. 2008). Of these, only *C. reinhardtii* and *O. tauri* are green algae (Chlorophyta), and none of them is from the Trebouxiophyceae. The *B. braunii* Berkeley genome size is closer to the green algae *Volvox carteri* (Chlorophyceae) genome (140 Mb), which has a draft sequence completed (http://genome.jgi-psf.org/Volca1/Volca1.home.html).

The genome size is not inconsistent with our phylogenetic placement of *B. braunii* Berkeley. On average, Trebouxiophyceae genome sizes (20–1,040 Mb) are larger than those of Chlorophyceae (9.8–588 Mb; Kapraun 2007). The *B. braunii* Berkeley genome size falls in the middle of each of these ranges. It has been proposed that relationships among algal genome sizes may be more related to ontogeny and reproductive modes rather than phylogenetic correlations (Kapraun 2005). Unfortunately, there is little to no information available on developmental processes or sexual cycles of *B. braunii*, which may be used to draw relationships among other known algal genome sizes. However, evolutionary relationships between very closely related organisms may be inferred from differences in genome sizes. For example, it is well established that there are genome size differences between species of closely related organisms such as fruit flies (Vieira et al. 2002, Bosco et al. 2007, Biemont 2008, Gregory and Johnston 2008). There is also increasing evidence that there are statistically significant differences in genome sizes between different strains of the same species (Cullis 2005, Bosco et al. 2007, Davison et al. 2007, Johnston et al. 2007, Biemont 2008, Gregory and Johnston 2008). Thus, the identification of the *B. braunii* Berkeley genome size lays the foundation for comparative genome size analysis between races and strains of *B. braunii*.

Genome GC content determination of B. braunii *Berkeley*. We estimated the GC content of *B. braunii* Berkeley using the coding sequences of 15 cDNAs randomly isolated from a cDNA library produced previously by us (Okada et al. 2000). The estimated GC content of these genes was determined to be 54.4 ± 1.2%. This estimate is likely to be slightly lower when introns are included in the analysis from genome sequencing since introns tend to be AT rich. For example, the whole genome of the green algae *C. reinhardtii* has a 64% GC content, while the expressed gene GC content of *C. reinhardtii* is 68% (Merchant et al. 2007).

Typical GC content for eukaryotic organisms is generally in the range of 30%–40%, with an upper limit for vertebrates and higher plants of 46% (Vinogradov 1994, 1998, Barow and Meister 2002, Meister and Barow 2007). Green algae tend to have genome GC content higher than most eukaryotes and rarely below 50%, but it can be as high as 71% (*Monoraphidium minutum;* Jarvis et al. 1992, Leon-Banares et al. 2004). All of the fully sequenced green algae genomes have a GC content of at least 58% (Misumi et al. 2008). Thus, the calculated GC content value of 54 ± 1.2% for *B. braunii* Berkeley is within the predicted range for green algae.

Our data presented here show that, based on genome size and GC content, the full sequence of the *B. braunii* Berkeley genome is an obtainable goal, especially since these parameters are within the range of other sequenced algal genomes. The information obtained from the Berkeley genome will aid in understanding the molecular mechanisms employed by this organism for such processes as hydrocarbon biosynthesis, the cell cycle, and photosynthesis.

We would like to thank Dr. Margaret E. Glasner from the Department of Biochemistry and Biophysics at Texas A&M University for help with the phylogenetic analysis.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–10.

Banerjee, A., Sharma, R., Chisti, Y. & Banerjee, U. C. 2002. *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals. *Crit. Rev. Biotechnol.* 22:245–79.

Barow, M. & Meister, A. 2002. Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry* 47:1–7.

Bennett, M. D., Bhandol, P. & Leitch, I. J. 2000. Nuclear DNA amounts in angiosperms and their modern uses-807 new estimates. *Ann. Bot.* 86:859–909.

Biemont, C. 2008. Genome size evolution: within-species variation in genome size. *Heredity* 101:297–8.

Bosco, G., Campbell, P., Leiva-Neto, J. T. & Markow, T. A. 2007. Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177:1277–90.

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., et al. 2008. The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–44.

Brown, A. C., Knights, B. A. & Conway, E. 1969. Hydrocarbon content and its relationship to physiological state in the green alga *Botryococcus braunii*. *Phytochemistry* 8:543–7.

Chapman, A. D. 2005. *Numbers of Living Species in Australia and the World*. Report for the Department of the Environment and Heritage, Canberra, Australia, pp. 1–64.

Chisti, Y. 2007. Biodiesel from microalgae. *Biotechnol. Adv.* 25:294–306.

Chisti, Y. 2008a. Biodiesel from microalgae beats bioethanol. *Trends Biotechnol.* 26:126–31.

Chisti, Y. 2008b. Response to Reijnders: do biofuels from microalgae beat biofuels from terrestrial plants? *Trends Biotechnol.* 26:351–2.

Cullis, C. A. 2005. Mechanisms and control of rapid genomic changes in flax. *Ann. Bot. (Lond.)* 95:201–6.

Davison, J., Tyagi, A. & Comai, L. 2007. Large-scale polymorphism of heterochromatic repeats in the DNA of *Arabidopsis thaliana*. *BMC Plant Biol.* 7:44.

Galbraith, D. W., Harkins, K. R., Maddox, J. M., Ayres, N. M., Sharma, D. P. & Firoozabady, E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220:1049–51.

Gelpi, E., Oro, J., Schneider, H. J. & Bennett, E. O. 1968. Olefins of high molecular weight in two microscopic algae. *Science* 161:700–2.

Gregory, T. R. & Johnston, J. S. 2008. Genome size diversity in the family *Drosophilidae*. *Heredity* 101:228–38.

Grung, M., Metzger, P. & Liaaen-Jensen, S. 1989. Primary and secondary carotenoids in two races of the green alga *Botryococcus braunii*. *Biochem. Syst. Ecol.* 17:263–9.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.

Hall, B. G. 2001. *Phylogenetic Trees Made Easy: A How-to Manual for Molecular Biologists.* Sinauer Associates, Sunderland, Massachusetts, 179 pp.

Hillen, L. W., Pollard, G., Wake, L. V. & White, N. 1982. Hydrocracking of the oils of *Botryococcus braunii* to transport fuels. *Biotechnol. Bioeng.* 24:193–205.

Huelsenbeck, J. P. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–5.

Jarvis, E. E., Dunahay, T. G. & Brown, L. M. 1992. DNA nucleoside composition and methylation in several species of microalgae. *J. Phycol.* 28:356–62.

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., Lopez, R. & Price, H. J. 2005. Evolution of genome size in *Brassicaceae*. *Ann. Bot. (Lond.)* 95:229–35.

Johnston, J. S., Ross, L. D., Beani, L., Hughes, D. P. & Kathirithamby, J. 2004. Tiny genomes and endoreduplication in *Strepsiptera*. *Insect Mol. Biol.* 13:581–5.

Johnston, J. S., Yoon, K. S., Strycharz, J. P., Pittendrigh, B. R. & Clark, J. M. 2007. Body lice and head lice (*Anoplura*: *Pediculidae*) have the smallest genomes of any hemimetabolous insect reported to date. *J. Med. Entomol.* 44:1009–12.

Kapraun, D. F. 2005. Nuclear DNA content estimates in multicellular green, red and brown algae: phylogenetic considerations. *Ann. Bot. (Lond.)* 95:7–44.

Kapraun, D. F. 2007. Nuclear DNA content estimates in green algal lineages: Chlorophyta and Streptophyta. *Ann. Bot. (Lond.)* 99:677–701.

Kitazato, H., Asaoka, S. & Iwamoto, H. 1989. Catalytic cracking of hydrocarbons from microalgae. *Sekiyu Gakkaishi* 32:28–34.

Knights, B. A., Brown, A. C., Conway, E. & Middleditch, B. S. 1970. Hydrocarbons from the green form of the freshwater alga *Botryococcus braunii*. *Phytochemistry* 9:1317–24.

Komárek, J. & Marvan, P. 1992. Morphological differences in natural populations of the genus *Botryococcus* (Chlorophyceae). *Arch. Protistenkd.* 141:65–100.

Leon-Banares, R., Gonzalez-Ballester, D., Galvan, A. & Fernandez, E. 2004. Transgenic microalgae as green cell-factories. *Trends Biotechnol.* 22:45–52.

López-Gómez, R. & Gómez-Lim, M. A. 1992. A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp. *Hortic. Sci.* 27:440–2.

Matsuzaki, M., Misumi, O., Shin, I. T., Maruyama, S., Takahara, M., Miyagishima, S. Y., Mori, T., et al. 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428:653–7.

Maxwell, J. R., Douglas, A. G., Eglinton, G. & McCornick, A. 1968. Botryococcenes-hydrocarbons of novel structure from the alga *Botryococcus braunii* Kützing. *Phytochemistry* 7:2157–71.

Meister, A. & Barow, M. 2007. DNA base composition of plant genomes. *In* Dolezel, J., Greilhuber, J. & Suda, J. [Eds.] *Flow Cytometry with Plant Cells.* Wiley-VCH Verlag GmBH & Co. KGaA, Weinheim, Germany, pp. 177–215.

Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., Terry, A., et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318:245–50.

Metzger, P. & Largeau, C. 2005. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. *Appl. Microbiol. Biotechnol.* 66:486–96.

Misumi, O., Yoshida, Y., Nishida, K., Fujiwara, T., Sakajiri, T., Hirooka, S., Nishimura, Y. & Kuroiwa, T. 2008. Genome analysis and its significance in four unicellular algae, *Cyanidioshyzon merolae*, *Ostreococcus tauri*, *Chlamydomonas reinhardtii*, and *Thalassiosira pseudonana*. *J. Plant. Res.* 121:3–17.

Moldowan, J. M. & Seifert, W. K. 1980. First discovery of botryococcane in petroleum. *J.Chem. Soc. Chem. Commun.* 19:912–4.

Nonomura, A. M. 1988. *Botryococcus braunii* var. *showa* (Chlorophyceae) from Berkeley, California, United States of America. *Jpn. J. Phycol.* 36:285–91.

Okada, S., Devarenne, T. P. & Chappell, J. 2000. Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, race B. *Arch. Biochem. Biophys.* 373:307–17.

Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* 12: 357–8.

Sawayama, S., Inoue, S. & Yokoyama, S. 1995. Phylogenetic position of *Botryococcus braunii* (Chlorophyceae) based on small subunit ribosomal RNA sequence data. *J. Phycol.* 31:419–20.

Senousy, H. H., Beakes, G. W. & Hack, E. 2004. Phylogenetic placement of *Botryococcus braunii* (Trebouxiophyceae) and *Botryococcus sudeticus* isolate UTEX 2629 (Chlorophyceae). *J. Phycol.* 40:412–23.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24:1596–9.

Traverse, A. 1955. Occurrence of the oil-forming alga *Botryococcus* in lignites and other Tertiary sediments. *Micropaleontology* 1:343–8.

Vieira, C., Nardon, C., Arpin, C., Lepetit, D. & Biemont, C. 2002. Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements? *Mol. Biol. Evol.* 19:1154–61.

Vinogradov, A. E. 1994. Measurement by flow cytometry of genomic AT/GC ratio and genome size. *Cytometry* 16:34–40.

Vinogradov, A. E. 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: the triangular relationship. *Cytometry* 31:100–9.

Wake, L. V. & Hillen, L. W. 1980. Study of a ''bloom'' of the oil-rich alga *Botryococcus braunii* in the Darwin River reservoir. *Biotechnol. Bioeng.* 22:1637–56.

Wake, L. V. & Hillen, L. W. 1981. Nature and hydrocarbon content of blooms of the alga *Botryococcus braunii* occurring in Australian freshwater lakes. *Aust. J. Mar. Freshw. Res.* 32: 353–67.

Supplementary Material

The following supplementary material is available for this article:

Figure S1. Alignment of 18S rDNA sequences from the original *B. braunii* Berkeley report (Sawayama et al. 1995) and the *B. braunii* Berkeley sequence reported here. Sequences were aligned using ClustalW. Accession numbers: original Berkeley sequence: X78276; Berkeley sequence reported here: FJ807044.

This material is available as part of the online article.

```
X78276   TACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAACTGTTTATACTGTGAA 85
FJ807044 TACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTCTAAGTATAAGCTGTTTATACTGTGAA 60
         ************************************************************************ ***************

X78276   ACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGGTACCTTACTACTCGGATAACCGTAGTAATTCTAGAGCTAATAC 170
FJ807044 ACTGCGAATGGCTCATTAAATCAGTTATAGTTTATTTGATGGTACCTTGCTACTCGGATAACCGTAGTAATTCTAGAGCTAATAC 170
         ********************************************** ********************************** 

X78276   GTGCGTAAATCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGGCCAGCCGGGCTTGCCCGACTCTTGGCGAATCATGATA 255
FJ807044 GTGCGTAAATCCCGACTTCTGGAAGGGACGTATTTATTAGATAAAAGGCTGACCGGGTTCGCCCGACTCTTGCTGACTCATGATA 255
         ************************************************   ***** * ************  ** ******** 

X78276   ACTTCACGAATCGCATGGCCTCGTGCCGGCGATGTTTCATTCAAATTTCTGCCCTATCAACTTTCGATGGTAGGATAGAGGCCTA 340
FJ807044 ACTCGACGGATCGCACGGGCTTGCCCCGGCGACGTTTCATTCAAATTTCTGCCCTATCAACTGTCGATGGTACGGTAGTGGCCTA 340
         ***  *** ****** ** ** *  ******* ***************************** ********* * *** ****** 

X78276   CCATGGTGGTGACGGGTGACGGAGGATTAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGCTACCACATCCAAGGAAGGCAG 425
FJ807044 CCATGGTGTTCACGGGTGACGGAGAATTAGGGTTCGATTCCGGAGAGGGCGCCTGAGAGACGGCGACCACATCCAAGGAAGGCAG 425
         ******** * ************* ********************** ******** ***** ******************** 

X78276   CAGGCGCGCAAATTACCCAATCCCGACACGGGGAGGTAGTGACAATAAATAACAATACTGGGCATTT--ATGTCTGGTAATTGGA 508
FJ807044 CAGGCGCGCAAATTACCCAATCCTGACACAGGGAGGTAGTGACAATAAATAACAATATCGGGGTTTCTAAACTCTGATAATTGGA 510
         ********************** ***** ***************************  ***  **   *  **** ******** 

X78276   ATGAGTACAATGTAAATATCTTAACGAGTATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC 593
FJ807044 ATGAGTACAATCTAAAATCCTTAACGAGGATCAATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCCAATAGC 595
         *********** ****   ********* *** ************************************************** 

X78276   GTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCGGGTGTGCCGTCGCGGTCTGCCT-CTGGTATGTACTGCGTT 677
FJ807044 GTATATTTAAGTTGTTGCAGTTAAAAAGCTCGTAGTCGGACTTCGGGTGGGGGCCGGCGGTCCGCCGACTGGTGTGCACTGC-CG 679
         ********************************** *** ******** *     ****** ***  ***** ** *****   

X78276   CGGTGCACCTTCCTGCTGGGGACGGGTTCCTGGGCTTCACTGTCTGGGACTCGGAGTCAGCAAAGTGACCTTGAGCAAACAAGAG 762
FJ807044 GGCCCCGCCTTGCTGCCGGAGATGGGAGCCTGGCGTTCGCTGTCCGGGCCCCGGACTCGGCGTGGTTACTTTGAGTAAATTAGAG 764
          *   * **** **** ** ** ***  *****  *** ***** *** * **** ** **   ** ** ***** ***  **** 

X78276   TGTTCAAAGCAAGCCTACGCTCTGAATTTTTTAGCATGGAATCACATGATAGGACTCTGGCCTATCTTGTTGGTCTGTAGGACCG 847
FJ807044 TGTTCAAAGCAGGCCTACGCTCTGAATATGTTAGCATGGAATAACGCGATAGGACTCTGGCCTATCTTGTTGGTCTGTGGGACCG 849
         *********** *************** * ************ **  ******************************** ****** 

X78276   GAGTAATGATTAAGAGGGACAGTCGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGAACATCT 932
FJ807044 GAGTAATGATTAAGAGGGACAGTCGGGGGCATTCGTATTTCATTGTCAGAGGTGAAATTCTTGGATTTATGAAAGACGGACTACT 934
         ********************************************************************************  **  ** 

X78276   GCGAAAGCATTTGCCAAGGATGTTTTCATTGATCAAGAACGAAAGTTGGGGGCTCGAAGACGATTAGATACCGTCGTAGTCTCAA 1017
FJ807044 GCGAAAGCATTTGCCAAGGATGTTTTCATTGATCAAGAACAAAAGTTGGGGGCTCGAAGACGATTAGATACCGTCCTAGTCTCAA 1019
         ************************************** ********************************** ********* 

X78276   CCATAAACGATGCCGACTAGGGATTGGAAGGTGTTCTTTTGACGACCCCTCCAGCACCTTATGAGAAATCAGAGTTTTTGGGTTC 1102
FJ807044 CCATAAACGATGCCGACTAGGGATTGGAGGGTGTTCTTTTGACGACCCCTCCAGCACCTTATGAGAAATCAAAGTTTTTGGGTTC 1104
         **************************** ****************************************** ************** 

X78276   CGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGAC 1187
FJ807044 CGGGGGGAGTATGGTCGCAAGGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAGGCGTGGAGCCTGCGGCTTAATTTGAC 1189
         ********************************************************************************** 

X78276   TCAACACGGGAAAACTTACCAGGTCCAGACATAGTGAGGATTGACAGATTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCA 1272
FJ807044 TCAACACGGGAAAACTTACCAGGTCCAGACATAGTGAGGATTGACAGATTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCA 1274
         ********************************************************************************** 

X78276   TGGCCGTTCTTAGTTGGTGGGTTGGCTTGTCAGGTTGATTCCGGTAACGAACGAGACCTCAGCCTACTAAATAGTCCGACCTGGT 1357
FJ807044 TGGCCGTTCTTAGTTGGTGGGTTGGCTTGTCAGGTTGATTCCGGTAACGAACGAGACCTCAGCCTGCTAAATAGTCCGACCTGGT 1359
         ************************************************************** ****************** 

X78276   CCTTCCAGACCGCCGACTTCTTAGAGGGACTCTCGGCGACTAGCCGGAGGAGGTGTGAGGCGATAACAGGTCTGTGATGCCCTTA 1442
FJ807044 TCTTCCAGGCCGCCGACTTCTTAGAGGGACTCTCGGCGACTAGCCGGAGGAAGTGTGAGGCGATAACAGGTCTGTGATGCCCTTA 1444
          ******* ****************************************** ******************************** 

X78276   GATGTTCTGGGCCGCACGCGCGCTACACTGATGCATGCAACGAGCCCAGCCTTGACCGAGAGCTCCGGGTAATCTAGGAAACTGC 1527
FJ807044 GATGTTCTGGGCCGCACGCGCGCTACACTGATGCATGCAACGAGCCTAGCCTTGACCGAGAGGTCCGGGTAATCTGGGAAACTGC 1529
         ********************************************** *************** ************* ********* 

X78276   ATCGTGATGGGGCTAAGTGATTGCAATTATTCATCTTCAACGAGGAATGCCTAGTAAGCGCCTGTCATCAGCAGGCGTTGATTAC 1612
FJ807044 ATCGTGATGGGGCTAGATGATTGCAATTATTCATCTTCAACGAGGAATGCCTAGTAAGCGCCTGTCATCAGCAGGCGTTGATTAC 1614
         ***************  ************************************************************************ 

X78276   GTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGGGTGTGCTGGTGAAGCGTTCGGATTGGTTTCAGTGGTTGGCA 1697
FJ807044 GTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGGGTGTGCTGGTGAAGCGTTCGGATCGGCGACTTCGGGCGGTC 1699
         ********************************************************************** **   *   **  ** 

X78276   ACTTCCGCTGTCGCCGAGAAGAACATTAAACCCTCCCACCTAGAGGAAGGAGAAGTCGTAACAAGGT                  1764
FJ807044 ACGCTTCGTGTCGCTGAGAAGTTCGTTAAACCCTCCCACCTAGAGGAAGGAGAAGTCGTAACAAGGT                  1766
         **      ****** ******  * ****************************************
```

Supplemental Fig. S1

Supplemental Figure S1. Alignment of 18S rDNA sequences from the original *B. braunii*, Berkeley report (Sawayama et al. 1995) and the *B. braunii*, Berkeley sequence reported here. Sequences were aligned using ClustalW. Accession numbers: original Berkeley sequence: X78276; Berkeley sequence reported here: FJ807044.