

NOTE

16S AND 23S PLASTID RDNA PHYLOGENIES OF *PROTOTHECA* SPECIES AND THEIR AUXANOGRAPHIC PHENOTYPES¹

Aren Ewing,² Shane Brubaker, Aravind Somanchi, Esther Yu, George Rudenko, Nina Reyes, Karen Espina

Solazyme Inc., 225 Gateway Boulevard, South San Francisco, California 94080, USA

Arthur Grossman

Solazyme Inc., 225 Gateway Boulevard, South San Francisco, California 94080, USA

Carnegie Institution for Science, Stanford, California 94305, USA

and Scott Franklin

Solazyme Inc., 225 Gateway Boulevard, South San Francisco, California 94080, USA

Because algae have become more accepted as sources of human nutrition, phylogenetic analysis can help resolve the taxonomy of taxa that have not been well studied. This can help establish algal evolutionary relationships. Here, we compare *Auxenochlorella protothecoides* and 23 strains of *Prototheca* based on their complete 16S and partial 23S plastid rDNA sequences along with nutrient utilization (auxanographic) profiles. These data demonstrate that some of the species groupings are not in agreement with the molecular phylogenetic analyses and that auxanographic profiles are poor predictors of phylogenetic relationships.

Key index words: 16s; 23s; auxanographic; phylogeny; plastid; *Prototheca*; species

Members of the genus *Prototheca* are nonphotosynthetic, achlorophyllous, plastid bearing, eukaryotic algae (Butler 1954, Ciferri et al. 1957, Pringsheim 1963, Turner and Lloyd 1966, Cooke 1968a,b, Nadakavukaren and McCracken 1973 and 1977). These round or ovoid (3–40 μm) single-celled organisms reside in the Chlorophyta or green algal lineage, which is composed of a number of algal classes including the Ulvophyceae, Chlorophyceae, and the Trebouxiophyceae, with the last of these containing the genus *Prototheca* (Graham and Wilcox 2000). Despite the fact that Trebouxiophycean algae have gained acceptance as a source of human nutrition, including *Chlorella* sp., *Auxenochlorella* (*Chlorella*) *protothecoides* (Kay 1991, Ravishankar et al. 2006, Day et al. 2009, Szabo et al. 2012, 2013) and *Prototheca*, both for its oils (Szabo et al. 2014), and as a poten-

tial source of vitamin C (Running et al. 2003), these organisms remain understudied, with the taxonomy of many Trebouxiophytes being poorly resolved; this is particularly true of the genus *Prototheca*.

Historically, taxonomy of the genus *Prototheca* has been based primarily on morphological and auxanographic features (Pore 1972, 1985). More recently, molecular taxonomic approaches have been used in an attempt to classify members of the genus *Prototheca* (Ueno et al. 2003, Roesler et al. 2006), but the utility of this approach was limited owing to the complex nature of the gene families used for the analysis (Ueno et al. 2007). Indeed, nuclear 18S rDNA sequences for *Prototheca* strains have been shown to exist in multiple copies many of which are in fact sequence variants, thereby complicating their utility in phylogenetic analysis, especially in closely related species. Due to this uncertainty, 16S and 23S plastid rDNA sequences were selected for this analysis.

In the work described here, we have sequenced the complete 16S plastid rDNA and a phylogenetically informative region of the 23S plastid rDNAs from 23 different *Prototheca* species, as well as from an isolate of *A. protothecoides*, which is the only photoautotroph in this group. These strains were obtained from various culture collections around the world. Here, we report phylogenetic relationships among the various *Prototheca* species, based on 16S and 23S plastid rDNA sequence information, and assess whether the phylogenetic relationships correlate with the ability of the organisms to use specific fixed carbon compounds.

Strains, growth, and genomic DNA extraction. The strains that were the subject of this study and their source collections are listed in Table 1. Cells were grown axenically and maintained essentially as

¹Received 24 January 2014. Accepted 30 April 2014.

²Author for correspondence: e-mail aewing@solazyme.com.

Editorial Responsibility: S. Lin (Associate Editor)

TABLE 1. Nutrient utilization of *Prototheca* species and *Auxenochlorella protothecoides*.

| Clade | Collection ID | Species | Authority | Arginine ^b | Fructose ^b | Glycerol ⁶ | Lysine ³ | PHC ^a | Proline ³ | Trehalose ¹¹ |
|-------|---------------|----------------------------|---------------------------|-----------------------|-----------------------|-----------------------|---------------------|------------------|----------------------|-------------------------|
| A | UTEX 250 | <i>A. protothecoides</i> * | (Krüger) Kalina & Puncová | + | + | + | - | - | - | + |
| B | SAG 263-11 | <i>P. wickerhamii</i> | Tubaki & Soneda | + | - | - | - | - | + | - |
| C | CBS 157.74 | <i>P. wickerhamii</i> | Tubaki & Soneda | + | + | + | + | + | + | + |
| | UTEX1439 | <i>P. moriformis</i> | W. Krüger | + | + | + | + | + | + | + |
| | UTEX1441 | <i>P. moriformis</i> | W. Krüger | + | + | (+) | + | + | + | + |
| | UTEX1435 | <i>P. moriformis</i> | W. Krüger | + | + | + | + | + | + | + |
| | UTEX1437 | <i>P. moriformis</i> | W. Krüger | + | + | + | + | + | + | + |
| | ATCC 16529 | <i>P. wickerhamii</i> | Tubaki & Soneda | + | + | + | + | + | + | + |
| D | ATCC 16528 | <i>P. stagnora</i> | W.B. Cooke | - | - | - | - | - | - | - |
| | CBS 605.66 | <i>P. stagnora</i> | W.B. Cooke | + | - | - | - | - | - | - |
| | ATCC 50112 | <i>P. ulmea</i> | R.S. Pore | + | - | - | - | - | - | - |
| | SAG 263-2.2 | <i>P. sp.</i> | W. Krüger # | + | - | - | - | - | - | - |
| E | SAG 2064 | <i>P. blaschkaea</i> | U. Roesler et al. | + | - | - | - | - | (+) | - |
| | ATCC 30253 | <i>P. zopfii</i> | W. Krüger | + | + | (+) | + | - | + | - |
| | UTEX 288 | <i>P. moriformis</i> | W. Krüger | + | + | + | + | - | - | - |
| | SAG 263-4 | <i>P. zopfii</i> | W. Krüger | + | + | + | + | - | - | - |
| | UTEX1442 | <i>P. stagnora</i> | W.B. Cooke | + | - | - | - | - | - | - |
| | ATCC 50081 | <i>P. moriformis</i> | W. Krüger | + | - | - | - | - | - | - |
| | SAG 43-80 | <i>P. zopfii</i> | W. Krüger | + | - | - | - | - | - | - |
| | UTEX329 | <i>P. kruegeri</i> | Pringsheim | - | - | - | - | - | - | - |
| | SAG 263.7 | <i>P. zopfii</i> | W. Krüger | - | - | - | - | - | + | - |
| | SAG 263-2.1 | <i>P. sp.</i> | W. Krüger # | + | + | + | - | - | - | - |
| | SAG 2021 | <i>P. zopfii</i> | W. Krüger | + | + | + | - | - | - | - |
| | SAG 2063 | <i>P. zopfii</i> | W. Krüger | + | + | + | - | - | - | - |

The nutrient utilization data presented in Table 1 are sorted in the same order as the 16S plastid rDNA tree. Nutrient utilization was determined by growth as described in the Materials and Methods. The names in the authority column conform to the convention given in algaebase.org (Guiry and Guiry 2014). The phenotypic responses were scored as - for no utilization, (+) for doubtful utilization and + for positive utilization. The nutrient source is labeled a) Enteric/Nonfermenter and b) Gram Positive. PHC is the abbreviation for p-n-p-phosphorylcholine. For *A. protothecoides* (marked with an * in Table 1), incubations were for 48 h at 28°C. SAG 263-2 (marked with a # in Table 1) was determined not to be uni-algal. The two resulting isolates were therefore isolated and split into two unique strains, 263-2.1 and 263-2.2.

described in Pore (1985). Genomic DNA was prepared as described in Rudenko and Walbot (2001).

Amplification and Sequencing of 16S and 23S rDNA. Approximately 1 ng of extracted genomic DNA was used as template for the PCR reactions. Primers were designed to anneal to regions of the 16S rRNA and 23S rRNA genes conserved across a wide variety of organisms. Oligonucleotides used for the 16S rDNAs were designed to amplify a region of ~1,431 bp of plastidic DNA and included a forward primer with the sequence 5'-AGAGTTTGATCCTG GCTCAG-3' and a reverse primer with the sequence 5'-GGCTACCTTGTTACGACTTC-3', annealing to nucleotides 122583 to 122602 and 124033 to 124014, respectively in Genbank Accession No. HQ914635. Oligonucleotides used for the 23S rDNAs were designed to amplify a region of ~530 bp of plastidic DNA and included a forward primer with the sequence 5'-TGTTGAAGAATGAGCCGCGAC-3' and a reverse primer with the sequence 5'-CAGTGAGCTATTACGC ACTC-3', annealing to nucleotides 637 to 658 and 1182 to 1163, respectively, in Genbank Accession No. HQ914635. Negative controls contained no template DNA.

Nutrient utilization. Algal cultures were incubated aerobically for 5 d at 28°C, essentially as described by Pore (1985), on medium containing 2% glucose, or 2% glucose supplemented with 4 g · L⁻¹ yeast extract. Nutrient utilization for each strain was assessed using

the microbial BBL Crystal Identification Systems (Becton Dickinson, Franklin Lakes, NJ, USA) as described previously (Roesler et al. 2006). Both the Enteric/Nonfermenter ID kit and the Gram Positive ID kit from Crystal ID Systems were employed to establish utilization profiles. *Prototheca* sample panels were incubated for 48 h at 37°C in a 70% humidified chamber. *Auxenochlorella* sample panels were incubated for 48 h at 28°C in a 70% humidified chamber. Panels were scored using the BBL Crystal Panel Viewer (Becton Dickinson).

Sequence alignment and tree construction. Both 16S and 23S plastid rDNA sequence data were used to build phylogenetic trees by the maximum-likelihood method. The multiple alignment algorithm used was Clustal W while tree generation was performed with the program PHYML, with *A. protothecoides* selected as an outgroup to root the tree. Parameters were GTR+G+I mode of sequence evolution search with 1,000 bootstraps; other tree generation methods tested included Bayesian, Parsimony and Neighbor Joining, all of which yielded essentially identical clading patterns.

The phylogenetic trees generated for both 16S (Fig. 1) and 23S (Fig. 2) rDNA share a very similar structure, with only slight differences within Clade E. The bootstrap value for each branch length demonstrates the strength of each branch point; the



FIG. 1. A 16S plastid rDNA phylogenetic tree generated by the maximum-likelihood method. The 16S rDNA plastid sequences from the 24 algal strains form five distinct clades, regardless of the tree generation algorithm used; the maximum-likelihood tree is displayed in Figure 1 along with associated bootstrap values. The five clades (A–E), are composed of: (A) *Auxenochlorella protothecoides*; (B) A single *Prototheca wickerhamii* isolate (C) A mixture of *P. wickerhamii* and *P. moriformis* isolates; (D) *P. ulmea* and *P. stagnora*; (E) primarily *P. zopfii* isolates with two *P. moriformis*, as well as *P. kruegeri* and *P. blaschkaea*.

branches separating each clade have support of 97 or higher. These bootstrap values demonstrate the robustness of this tree, which had the same clading structure regardless of the tree building method used; very similar trees were generated using the Bayesian, Parsimony and Neighbor Joining methods.

Using pairwise comparisons between individuals in each clade, the 16S rDNA data show that Clade A and B share 93.9% sequence similarity (UTEX 250 and SAG 263-11); Clades B and C (SAG 263-11 and ATCC 16529) share 86.8% sequence similarity; Clades C and D (ATCC 16529 and CBS 605.66) share 67.9% sequence similarity; Clades C and E (ATCC 16529 and SAG 43-80) share 71.6% sequence similarity; Clades D and E (CBS 605.66 and SAG 43-80) share 76.1% sequence similarity.

The auxanographic data (Table 1) for these algal strains do not entirely conform to either the current taxonomic groupings or the molecular data presented here. All organisms listed grew on glucose (data not shown). The clades described in the phylogenetic trees group the majority of the trehalose consuming algae together (Clade C), with the exception of *P. wickerhamii* SAG 263-11, which was unable to use trehalose. Furthermore, *A. protothecoides* UTEX 250 (Clade A) also was able to use trehalose, and was only able to grow at the lower temperature of 28°C. *P. kruegeri* UTEX 329 and

P. stagnora ATCC 16528 showed no growth on any of the alternative carbon sources, and *P. zopfii* SAG 263.7 only grew when proline was used as a carbon source. Furthermore, a number of organisms in Clades D (*P. stagnora* CBS 605.66, *P. ulmea* ATCC 50112, *P. sp.* SAG 263-2.2) and E (*P. stagnora* UTEX 1442, *P. moriformis* ATCC 50081, *P. zopfii* 43-80) only grew on arginine as an alternative carbon source, while others in Clade E could use a number of different carbon sources (fructose, glycerol, lysine and proline). The finding that the specificity of the strains with respect to the utilization of different fixed carbon sources is not strictly correlated with the clading results, is interesting since some of these strains show close phylogenetic relations (see for example *P. zopfii* SAG 263.7 and *P. kruegeri* UTEX 329).

Evolutionary relationships based on phylogenies relative to nutritional information. The molecular genetic data provide a robust phylogeny of the different *Prototheca* strains based on both 16S and 23S plastid rDNA sequences. This finding reinforces the conclusions of previous work (Ueno et al. 2003) using 18S nuclear rDNA and 16S plastid rDNA sequences. However, the data presented here are more extensive than previous studies because a larger number of taxa were examined and we included 23S plastid rDNA information. The same strains that we used

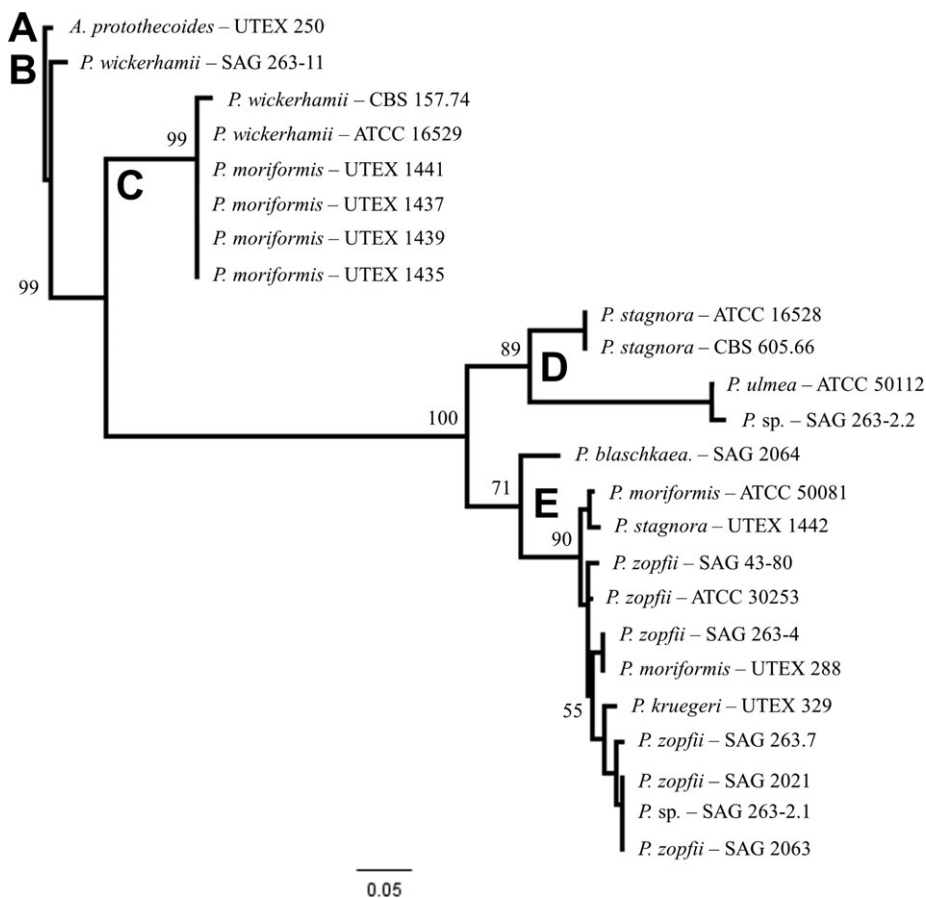


FIG. 2. 23S plastid rDNA tree generated by the maximum-likelihood method. Like the 16S rDNA plastid sequences the 23S rDNA plastid sequences form five distinct clades regardless of the tree generation algorithm used. The five clades are identical in makeup to those described for the 16S clades.

for the phylogenetic analyses were analyzed for nutrient utilization, and there were a number of cases in which the nutrient utilization profile was not consistent with the phylogenetic grouping of the organism. While nutrient utilization studies are relatively easy to perform and have been widely used to classify *Prototheca* isolates in the past, our work clearly shows that inferences concerning phylogenetic relationships based on nutrient utilization are problematic. As an example, *P. stagnora* ATCC 16528 (Clade D; Figs. 1 and 2) shows the exact same auxanographic fingerprint as *P. kruegeri* UTEX 329 (Clade E; Figs. 1 and 2) and yet molecular genetic characterization shows these strains to be quite disparate.

Trehalose utilization was previously used to distinguish *P. wickerhamii* from all other *Prototheca* species. While others have successfully grown *P. wickerhamii* SAG 263-11, one of the most well-characterized strains, on trehalose (Pore 1985, Ueno et al. 2005), in this study the growth of *P. wickerhamii* SAG 263-11 was not supported by trehalose (under either of the temperature conditions used). This is surprising since the strain previously tested (Ueno et al. 2005) is the same as ours based on identical 16S and 23S rDNA sequences (Genbank: AJ245645 and Knauf and Hachtel 2002). It is possible that the strain lost its ability to utilize this sugar during long years of culturing in the absence of trehalose; this could be

a feature associated with many nutrient uptake and assimilation systems. Furthermore, some of the features associated with nutrient utilization can be acquired as a consequence of lateral gene transfer and would not reflect the phylogeny of the organism. Overall, our results are cautionary and suggest that basing strain classification solely on auxanographic features, which has been done in the past, can be misleading, while sequence-based phylogenies, particularly when utilizing multiple gene sequences, provide stronger evidence for evolutionary relationships.

Naming inconsistencies between strains. The naming of the different species within the genus *Prototheca* has been driven largely by morphological and auxanographic characteristics over the past 120 years. As observed, based on the 16S and 23S plastid rDNA phylogenetic trees that we generated, the current taxonomy forms paraphyletic groups. Among two of the five clades, *P. moriformis* isolates share near identical 16S and 23S plastid rDNA sequences with *P. wickerhamii* isolates in Clade C and *P. zopfii* isolates in Clade E. Similarly *P. wickerhamii* isolates are spread among Clades B and C. Furthermore, *P. wickerhamii* SAG 263-11 appears to be more closely related to *A. protothecoides* UTEX 250 than to the *P. wickerhamii* isolates in Clade C, which includes strains designated as both *P. wickerhamii* and *P. moriformis*. Based on the molecu-

lar genetic data presented here, the strains named *P. moriformis*, *P. wickerhamii* and *P. stagnora* are part of paraphyletic groups that should be resolved in future studies.

- Butler, E. 1954. Radiation-induced chlorophyll-less mutants of *Chlorella*. *Science* 120:274–5.
- Ciferri, R., Montemartini, A. & Ciferri, O. 1957. Morphological and assimilative characteristics and speciology of *Protothecae*. *Nuovi. Ann. Ig. Microbiol.* 8:554–63.
- Cooke, W. B. 1968a. Studies in the genus *Prototheca*. I. Literature review. *J. Elisha Mitchell Sci. Soc.* 84:213–6.
- Cooke, W. B. 1968b. Studies in the genus *Prototheca*. II. Taxonomy. *J. Elisha Mitchell Sci. Soc.* 84:217–20.
- Day, A. G., Brinkmann, D., Franklin, S., Espina, K., Rudenko, G., Roberts, A. & Howse, K. S. 2009. Safety evaluation of a high-lipid algal biomass from *Chlorella protothecoides*. *Regul. Toxicol. Pharmacol.* 55:166–80.
- Graham, L. E. & Wilcox, L. W. 2000. *Algae*. Prentice Hall, Inc., Upper Saddle River, New Jersey.
- Guiry, M. D. & Guiry, G. M. 2014. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. Available at <http://www.algaebase.org> (searched on April 7, 2014).
- Kay, R. A. 1991. Microalgae as food and supplement. *Crit. Rev. Food Sci. Nutr.* 30:555–73.
- Knauf, U. & Hachtel, W. 2002. The genes encoding subunits of ATP synthase are conserved in the reduced plastid genome of the heterotrophic alga *Prototheca wickerhamii*. *Mol. Genet. Genomics* 267:492–7.
- Nadakavukaren, M. J. & McCracken, D. 1973. *Prototheca*: an alga or fungus? *J. Phycol.* 9:113–6.
- Nadakavukaren, M. J. & McCracken, D. A. 1977. An ultrastructural survey of the genus *Prototheca* with special reference to plastids. *Mycopathologia* 61:117–9.
- Pore, R. S. 1972. Nutritional basis for relating *Prototheca* and *Chlorella*. *Can. J. Microbiol.* 18:1175–7.
- Pore, R. S. 1985. *Prototheca* taxonomy. *Mycopathology* 90:129–39.
- Pringsheim, E. G. 1963. *Farblose Algen. Ein Beitrag zur Evolutionsforschung*. Gustav Fischer Verlag, Stuttgart, pp. i–xi, 1–471.
- Ravishankar, G. A., Sarada, R., Sandesh Kamath, B. & Namitha, K. K. 2006. Food applications of algae. In Shetty, K., Paliyath, G., Pometto, A. & Levin, R. E. [Eds.] *Food Biotechnology*, 2nd edn. CRC Press, Boca Raton, Florida, pp. 491–524.
- Roesler, U., Moller, A., Hensel, A., Baumann, D. & Truyen, U. 2006. Diversity within the current algal species *Prototheca zopfii*: a proposal for two *Prototheca zopfii* genotypes and description of a novel species, *Prototheca blaschkeae* sp. nov. *Int. J. Syst. Evol. Microbiol.* 56:1419–25.
- Rudenko, G. N. & Walbot, V. 2001. Expression and post-transcriptional regulation of maize transposable element MuDR and its derivatives. *Plant Cell.* 13:553–70.
- Running, J. A., Burlingame, R. P. & Berry, A. 2003. The pathway of L-ascorbic acid biosynthesis in the colourless microalga *Prototheca moriformis*. *J. Exp. Bot.* 54:1841–9.
- Szabo, N. J., Matulka, R. A., Kiss, L. & Licari, P. 2012. Safety evaluation of a high lipid Whole Algalin Flour (WAF) from *Chlorella protothecoides*. *Regul. Toxicol. Pharm.* 63:155–65.
- Szabo, N. J., Matulka, R. A., Kiss, L. & Licari, P. 2013. Safety evaluation of Whole Algalin Protein (WAP) from *Chlorella protothecoides*. *Food Chem. Toxicol.* 59:34–9.
- Szabo, N. J., Matulka, R. A., Marone, P. A., Bauter, M. R., Chan, T., Franklin, S., Carney, J. R., McQuaid, S. L., Rakitsky, W., Green, R. & Licari, P. 2014. Safety evaluation of oleic-rich triglyceride oil produced by a heterotrophic microalgal fermentation process. *Food Chem. Toxicol.* 65:301–11.
- Turner, G. & Lloyd, D. 1966. The chemical composition and structure of the cell wall of *Prototheca zopfii*. *Biochem. J.* 99:55.
- Ueno, R., Hanagata, N., Urano, N. & Suzuki, M. 2005. Molecular phylogeny and phenotypic variation in the heterotrophic green algal genus *Prototheca*. *J. Phycol.* 41:1268–80.
- Ueno, R., Huss, V., Urano, N. & Watabe, S. 2007. Direct evidence for redundant segmental replacement between multiple 18S rRNA genes in a single *Prototheca* strain. *Microbiology* 153:3879–93.
- Ueno, R., Urano, N. & Suzuki, M. 2003. Phylogeny of the non-photosynthetic green microalgal genus *Prototheca* (Trebouxio-phyceae, Chlorophyta) and related taxa inferred from SSU and LSU ribosomal DNA partial sequence data. *FEMS Microbiol. Lett.* 223:275–80.