Complete Mitochondrial and Plastid Genomes of the Green Microalga *Trebouxiophyceae* sp. Strain MX-AZ01 Isolated from a Highly Acidic Geothermal Lake

Luis E. Servín-Garcidueñas and Esperanza Martínez-Romero

Center for Genomic Sciences, UNAM, Cuernavaca, Morelos, Mexico

We report the complete organelle genome sequences of *Trebouxiophyceae* sp. strain MX-AZ01, an acidophilic green microalga isolated from a geothermal field in Mexico. This eukaryote has the remarkable ability to thrive in a particular shallow lake with emerging hot springs at the bottom, extremely low pH, and toxic heavy metal concentrations. *Trebouxiophyceae* sp. MX-AZ01 represents one of few described photosynthetic eukaryotes living in such a hostile environment. The organelle genomes of *Trebouxiophyceae* sp. MX-AZ01 are remarkable. The plastid genome sequence currently presents the highest G+C content for a trebouxiophyte. The mitochondrial genome sequence is the largest reported to date for the *Trebouxiophyceae* class of green algae. The analysis of the genome sequences presented here provides insight into the evolution of organelle genomes of trebouxiophytes and green algae.

The *Trebouxiophyceae* are a class of phylum *Chlorophyta*, which comprises algae from marine and fresh waters (5). *Trebouxiophyceae* sp. strain MX-AZ01 was recently identified and isolated from a lake from the Los Azufres geothermal field in western Mexico, where it seems to be endemic. This strain thrives in a lake with pH 2.3 and may represent a novel species. Here we present the complete and annotated organelle genomes of *Trebouxiophyceae* sp. MX-AZ01. These sequences represent the first organelle genomes reported for an acidophilic trebouxiophyte.

Trebouxiophyceae sp. MX-AZ01 was maintained at the Center for Genomic Sciences in the culture collection of the Ecological Genomics Department of UNAM. DNA was sequenced with the Illumina GAIIx platform. Reads were assembled *de novo* using Velvet 1.2.07 (9). Some contigs with overrepresented coverage corresponded to organelle sequences. Reads were mapped to gap-surrounding sequences by using Maq 0.7.1 (6). Mapping reads and PCR amplifications were used to eliminate gaps. Coding sequences were predicted using GeneMark.hmm 2.0 (1). rRNA and tRNA were predicted using RNAmmer 1.2 (4) and tRNAscan-SE 1.21 (7). Further, gene predictions were manually verified. The organelle genomes from *Coccomyxa subellipsoidea* C-169 (8) and *Trebouxiophyceae* sp. MX-AZ01 were aligned using Mauve 2.3.1 (3).

The mitochondrial genome sequence (74.4 kb; 197-fold coverage) is the largest currently deposited in GenBank for the *Trebouxiophyceae* and has a G+C content of 53.4%. The mitochondrial genome comprises 42 putative coding genes, 24 tRNAs, and 3 rRNAs. Seven putative endonucleases were detected as part of intronic regions, with six of them located in the large-subunit rRNA gene and one present in the cytochrome *c* oxidase subunit 1. Group II introns were detected among four tRNA-coding genes ($trnH_{gug}$, $trnS_{gcu}$, $trnS_{uga}$, and $trnW_{cca}$). The gene order between *Trebouxiophyceae* sp. MX-AZ01 and *C. subellipsoidea* mitochondrial genome is conserved.

The plastid genome (149.7 kb; 312-fold coverage) has the highest G+C content, 57.7%, exceeding the recently described 50.7% G+C content of the *C. subellipsoidea* plastid (8). The *Trebouxiophyceae* sp. MX-AZ01 plastid genome comprises 81 putative coding genes, 32 tRNAs, and 3 rRNAs. All three rRNA genes clustered together, different from the ribosomal gene organization of the plastid genome of *C. subellipsoidea*. Three putative endonucleases were detected as part of intronic regions, with two of them located in the large-subunit rRNA gene and another one found in *psbA*. The group I intron found in *psbB* in the plastid genome of *C. subellipsoidea* was absent (8). Comparisons of the plastid genomes of *C. subellipsoidea* and *Trebouxiophyceae* sp. MX-AZ01 suggested the occurrence of several genome rearrangements.

The organelle genomes were sequenced as part of a research initiative aimed at describing the diversity of extremophiles from the Los Azufres geothermal field. These are the first complete genome sequences obtained from a microorganism isolated from Los Azufres. The nuclear genome from *C. subellipsoidea* has been reported previously (2). Future comparative genomics between nuclear genomes will allow better taxonomic classification of *Trebouxiophyceae* sp. MX-AZ01.

Nucleotide sequence accession numbers. The genome sequences determined in this study have been deposited in DDBJ/EMBL/GenBank under accession numbers JX315601 (mitochondrion) and JX402620 (plastid).

ACKNOWLEDGMENTS

We are thankful with the University Unit for Massive Sequencing (UUSM) from the National University of Mexico (UNAM) for Illumina-Solexa sample sequencing. The Comisión Federal de Electricidad (CFE) personnel gave us field assistance and permits for samplings at Los Azufres. We are also thankful to Michael Dunn for reviewing the manuscript and Jesus Campos Garcia for allowing us to process some samples at the Chemical-Biology Research Institute (IIQB) from the Universidad Michoacana de San Nicolás de Hidalgo (UMSNH). Samplings were carried out with the efforts of Hans Román, Jonathan Lopez, Cecilia Garci-

Received 4 September 2012 **Accepted** 4 September 2012

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

doi:10.1128/EC.00244-12

dueñas, and José Luis Servín. We are grateful to Paul Gaytán and Eugenio López for synthesis of oligonucleotides at the Synthesis Unit of the Biotechnology Institute of UNAM.

This work was supported by PAPIIT IN205412 from DGAPA, UNAM. L.E.S.-G. was supported with a scholarship from CONACyT as Ph.D. student at the Programa de Doctorado en Ciencias Biomédicas from UNAM.

REFERENCES

1. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27**:3911–3920.
2. **Blanc G, et al.** 2012. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. Genome Biol. **13**: R39.
3. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. **14**:1394–1403.
4. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
5. **Leliaert F, et al.** 2012. Phylogeny and molecular evolution of the green algae. Crit. Rev. Plant Sci. **31**:1–46.
6. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–1858.
7. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
8. **Smith DR, et al.** 2011. The GC-rich mitochondrial and plastid genomes of the green alga *Coccomyxa* give insight into the evolution of organelle DNA nucleotide landscape. PLoS One **6**:e23624. doi:10.1371/journal.pone.0023624.
9. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.