The chloroplast genome of the marine microalga *Tisochrysis lutea*

Alejandra B. Méndez-Leyva, Jingya Guo, Elisabeth A. Mudd, Jerry Wong, Jean-Marc Schwartz and Anil Day

School of Biological Sciences, The University of Manchester, Manchester, UK

**ABSTRACT**

*Tisochrysis lutea* is a haptophyte microalga important in aquaculture. The 105,837-bp chloroplast genome encodes 113 proteins, 27 different transfer-RNAs and ribosomal-RNAs. Phylogenetic analysis on currently available chloroplast sequences showed *T. lutea* was most closely related to *Emiliania huxleyi*. Unlike *E. huxleyi* chloroplast DNA, *T. lutea* contains directly-repeated ribosomal repeats following multiple genome rearrangements during evolution. The non-identical 4.9-kb ribosomal direct-repeats are distinguished by 42 nucleotide polymorphisms and three single- or double-nucleotide INDELs. This is unlike land plant and green algal chloroplast genomes in which efficient copy correction by homologous recombination ensures identical long inverted repeats containing the rRNA operons. Imperfect rDNA repeats are also found in the chloroplast genomes of cryptophytes, rhodophytes and other haptophytes. Reduced copy correction rates in these algal lineages raise questions on the mechanism(s) by which genetic uniformity of multi-copy chloroplast genomes is maintained and has implications for developing chloroplast transformation in these species.

*Tisochrysis lutea* CCAP-927/14, previously classified as *Isochrysis sp.* (Bendif et al. 2013), is a marine microalga widely used to feed juvenile fish and shellfish larvae (Duerr et al. 1998). It is a rich source of carotenoids and DHA (Carrier et al. 2004). We report the *T. lutea* 105,837 bp chloroplast genome (cp genome) sequence which has applications for phylogenetic studies (Edvardsen et al. 2000) and developing chloroplast transformation (Day and Goldschmidt-Clermont 2011) for industrial applications.

CCAP-927/14 was from the stored culture-collection-of-algae-and-protozoa (CCAP), OBAN (56°27′05.2″N, 5°26′27.0″W). Total DNA was sequenced on the Illumina MiSeq platform (San Diego, USA) with 300-bp paired-end reads (Nextera DNA kit libraries). De novo assembly with SPAdes (Bankevich et al. 2012; Arkin et al. 2018) produced three overlapping contigs of cpDNA comprised of two single-copy regions and a rDNA repeat, which assembled into a single circular contig. Ambiguities were resolved by Sanger sequencing of PCR products including the rDNA repeats, which were amplified with primers in adjacent single-copy sequences. Average coverage was >1,300 reads (Geneious-R11, Biomatters Ltd, Auckland). Coding regions and tRNA genes located with Vector NTI advance v9 (Invitrogen, Carlsbad) and ARAGORN (Laslett and Canback 2004) were annotated using BLAST searches against GenBank sequences and *E. huxleyi* cpDNA.

The *T. lutea* cp genome contains two single-copy regions of 54,842 and 41,289 bp separated by direct rDNA repeats of 4852 and 4854 bp. Multiple rearrangements distinguish the distant cp genomes in Figure 1. The MAFFT (Katoh and Standley 2013) sequence alignment of the co-linear ribosomal operons was used in TreeBuilder (Geneious-R11) with neighbour-joining (Tamura and Nei 1993) and Pseudo-nitzschia as the outgroup. The rDNA sequence varies in number and orientation in these taxa (Figure 1). *T. lutea* groups with *E. huxleyi* consistent with previous data (Edvardsen et al. 2000). Mauve alignment (Darling et al. 2004) of *T. lutea* and *E. huxleyi* cpDNA showed rearrangement of ten sequence blocks and recombination between the single-copy regions that switched the orientation of rDNA repeats.

Features shared with *E. huxleyi* (Puerta et al. 2005) include no introns, 111 protein-coding genes, two 16S-trnI-trnA-23S-5S ribosomal RNA operons, 27 different transfer RNAs and overlapping psbD/psbC genes. Two *T. lutea* ORF products of 57 and 76 amino acids with high similarity to hypothetical bacterial proteins (GenBank EAZ93707 and EAL14248) were truncated or absent in *E. huxleyi*. Of nine unidentified ORFs (>30 codons) in *E. huxleyi* (Puerta et al. 2005), only ORF35 (psbY) was present in *T. lutea*. The imperfect *T. lutea* rDNA repeats include 19 nucleotide substitutions in the 23S rRNA genes. Long PCR using primers bordering the rDNA repeats in all combinations indicated a lack of recombination between the rDNA repeats.

In Chlamydomonas and plant chloroplasts, RecA-mediated homologous recombination appears to underpin the
following: i) copy correction between the large rDNA-inverted repeats, ii) the uniformity of cp genomes and iii) integration of foreign transgenes (Day and Madesis 2007). Apparent lack of copy-correction-recombination between rDNA repeats in the chloroplasts of haptophytes (Puerta et al. 2005; Hovd et al. 2014, this work), cryptophytes (Douglas and Penny 1999) and rhodophytes (Reith and Munholland 1993) raises questions on the cpDNA-replication-recombination-repair pathways in these algae, and the design of chloroplast transformation vectors.

Acknowledgements

We thank the CCAP, SAMS Ltd, Scottish Marine Institute, Oban, UK, for T. lutea CCAP 927/14, GenBank for sequences used in the Phylogenetic Tree, the Genomics Technologies Next Generation Sequence Facility and Richard Leong in research IT for their support.

Disclosure statement

The authors alone are responsible for the content and writing of the paper. None of the authors report any conflict of interest.

Funding

The work was supported by The University of Manchester. Alejandra B. Méndez-Leyva was a Conacyt-Mexico PhD scholar (no. 399942).

ORCID

Anil Day https://orcid.org/0000-0001-7835-9572

References

Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnol. 36:566–569.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Computational Biol. 19:455–477.

Bendif E, Probert I, Schroeder DC, de Vargas C. 2013. On the description of Tisochrysis lutea gen. nov. sp. nov. and Isochrysis nuda sp. nov. in the Isochrysidales, and the transfer of Dicrateria to the Prymnesiales (Haptophyta). J Appl Phycol. 25:1763–1776.

Carrier G, Baroukh C, Rouxel C, Duboscq-Bidot L, Schreiber N, Bougaran G. 2018. Draft genomes and phenotypic characterization of Tisochrysis lutea strains. Toward the production of domesticated strains with high added value. Algal Res. 29:1–11.

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14:1394–1403.

Day A, Goldschmidt-Clermont M. 2011. The chloroplast transformation toolbox: selectable markers and marker removal. Plant Biotechnol J. 9: 540–553.

Day A, Madesis P. 2007. DNA replication, recombination, and repair in plastids. In: Bock R, editor. Cell and molecular biology of plastids. Vol 19. Topics in current genetics. Berlin, Germany: Springer-Verlag Berlin, Heidelberg; p. 65–119.

Duerer EO, Molnar A, Sato V. 1998. Cultured microalgae as aquaculture feeds. J Mar Biotechnol. 6:65–70.

Douglas SE, Penny SL. 1999. The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved syntenic groups confirm its common ancestry with red algae. J Mol Evol. 48: 236–244.

Edvardsen B, Eikrem W, Green JC, Andersen RA, Moon-van der Staay SY, Medlin LK. 2000. Phylogenetic reconstructions of the Haptophyta

Figure 1. Neighbour-joining consensus tree of the chloroplast ribosomal RNA operon (rDNA) from the indicated species. Bootstrap values (1000 replicates) are indicated at nodes. The number and orientation of the rDNA operon sequence are indicated: SR, single rDNA sequence; IR, two inverted rDNA sequences; DR, two directly repeated rDNA sequences. The position of T. lutea MF795089 (this work) is arrowed. Sequences retrieved from GenBank. Scale bar: substitutions per site.
inferred from 18S ribosomal DNA sequences and available morphological data. Phycologia. 39:19–35.

Hovde BT, Starkenburg SR, Hunsperger HM, Mercer LD, Deodato CR, Jha RK, Chertkov O, Monnat RJ, Cattolico RA. 2014. The mitochondrial and chloroplast genomes of the haptophyte *Chrysochromulina tobin* contain unique repeat structures and gene profiles. BMC Genomics. 15:604.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol and Evol. 30:772–780.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucl Acids Res. 32:11–16.

Puerta MVS, Bachvaroff TR, Delwiche CF. 2005. The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. DNA Res. 12:151–156.

Reith M, Munholland J. 1993. The ribosomal RNA repeats are non-identical and directly oriented in the chloroplast genome of the red alga *Porphyra purpurea*. Current Genet. 24:443–450.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512–526.