# Structural Gene Organization in *Chromochloris zofingiensis* Can Drive Advancements in Synthetic Biology

Jeffery M. Moseley,[1*] (jlmoseley@berkeley.edu), Sean D. Gallaher,[2*] Sean R. McCorkle,[3] Crysten E. Blaby-Haas,[3] and **Sabeeha S. Merchant**[1]

[1]Quantitative Biosciences Institute, Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, [2]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90035, [3]Biology Department, Brookhaven National Laboratory, Upton, NY 11973

**Project Goals: Our overarching research goal is to design and engineer high-level production of biofuel precursors in photoautotrophic cells of the unicellular green alga *Chromochloris zofingiensis*. Our strategy involves large-scale multi-'omics systems analysis to understand the genomic basis for energy metabolism partitioning as a consequence of carbon source. We are integrating the systems data in a predictive model that will guide the redesign and engineering of metabolism in *C. zofingiensis.* Toward our goal, we are also leveraging fundamental discoveries in green algal genomics and transcriptomics to build new synthetic biology tools for the design and synthesis of engineered bioproduction pathways.**

In contrast to prokaryotes, where functionally cooperating proteins are often encoded by operons and co-transcribed, such structural organization had been thought to be rare in most eukaryotes. However, as the number of sequenced eukaryotic genomes and transcriptomes has increased, and the function of those encoded proteins has been revealed, non-random gene organization (i.e. physical clustering of pathway members and co-regulated genes) is emerging as a characteristic of eukaryotic genomes. Current methods for identifying functionally cooperative gene neighborhoods in eukaryotes depend on the availability of functional annotations, which limits our ability to identify clustered functional gene units in algal genomes. Over half of algal proteins are of unknown function, and many of the remaining genes may be mis-annotated, or have vague functional assignments. A related challenge is the quality of structural annotations that are needed to predict coding regions and serve as the input for downstream comparative genomic analyses.

To address these concerns in the oleaginous green alga, *Chromochloris zofingiensis,* we recently embarked on an effort to re-annotate the genome using single-molecule, long-read sequencing of whole transcripts on the PacBio platform (Iso-Seq)[1]. Unexpectedly, this effort revealed hundreds of examples where two, three, or more genes were exclusively co-transcribed on polycistronic transcripts. In a survey of seven green algal species representing the breadth of the chlorophyte lineage, we observed that many of these polycistronic operons were evolutionarily conserved among species that had diverged hundreds of millions of years ago. Informed by these discoveries,

we designed synthetic bicistronic constructs that were capable of co-expressing pairs of proteins, including reporter and selectable proteins, in vitro. These findings provide an opportunity to build synthetic biology tools for designing optimized algal strains. Here, we present our progress on using highly conserved polycistronic mRNAs as scaffolding for algal engineering.

*C. zofingiensis* has a robust cell wall that complicates efforts at engineering the species to express transgenes. Given that many of the polycistronic loci we identified in *C. zofingiensis* were widely conserved in Chlorophytes, we used known polycistronic loci from that species to identify candidates in a more genetically tractable species: the oleaginous Trebouxiophyte alga, *Auxenochlorella protothecoides* (strain Cp0710). In a preliminary survey we found four potential polycistronic genes that are conserved between *A. protothecoides*, *Chlamydomonas reinhardtii* and *C. zofingiensis*. Iso-Seq analysis of *A. protothecoides* transcription has not been reported, but mapping of RNA-seq reads to these four loci was consistent with the expression of polycistronic transcripts. We have demonstrated transformation and gene targeting by homologous recombination in another *A. protothecoides* strain, UTEX 250, and we employed this platform to test whether inter-ORF sequences from the putative *A. protothecoides* polycistronic genes could be used for in vivo transgene expression. An artificial polycistronic gene using the *A. protothecoides* TOM22-SDHAF3 homolog inter-ORF sequence to link synthetic ORFs encoding Arabidopsis THIC (hydroxymethyl pyrimidine synthase) and Chlamydomonas BKT1 (beta-carotene ketolase) was targeted to the lycopene cyclase epsilon (LCYE) locus in *A. protothecoides* UTEX 250. Transformants were selected on the basis of gain-of-function thiamine prototrophy, requiring THIC activity. The orange color of the transformants, resulting from synthesis of red keto-carotenoids, indicated that BKT1 was functional as well. Future experiments will explore the use of artificial polycistronic genes to alter fatty acid and lipid composition.

[1] Gallaher, S.D. *et al.*, Widespread polycistronic gene expression in green algae. *Proc. Natl. Acad. Sci. U.S.A.,* in press (2021).

*This research is supported by the DOE Office of Science, Office of Biological and Environmental Research (BER), award number DE-SC0018301 and a FICUS JGI/EMSL project, award number 503551.*