# Genetic Blueprint of Microalgae Carbon Productivity

*Igor Grigoriev (LBNL) and Shawn Starkenburg (LANL)*

**DOE Bioenergy Technologies Office (BETO) 2021 Project Peer Review**

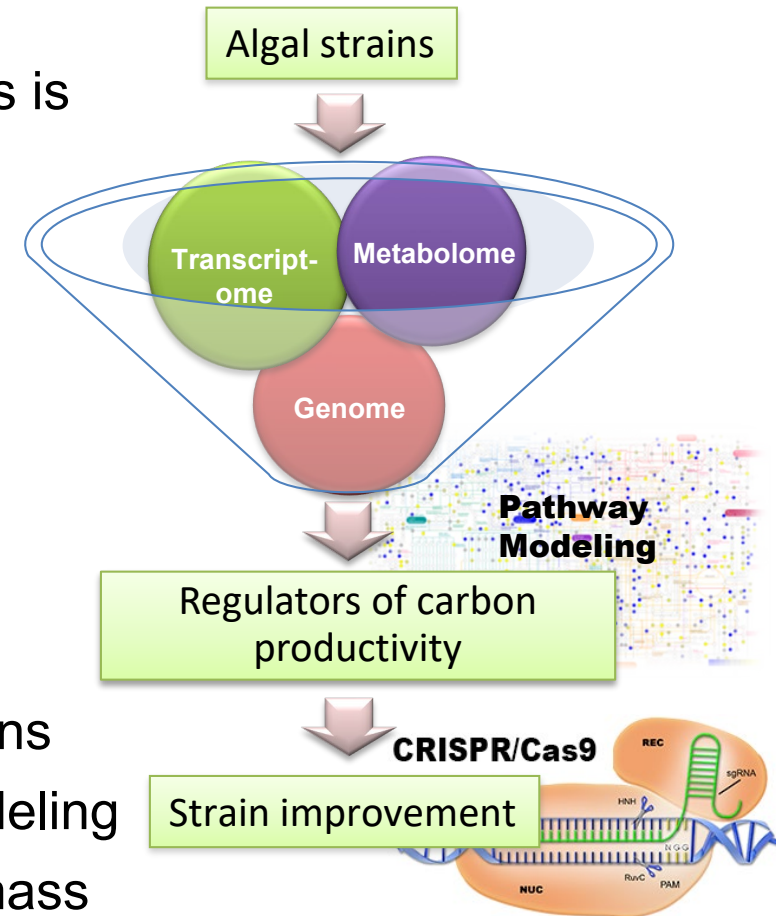# Project Overview

**Background:**

The potential of microalgae to emerge as major biofuel producers is limited by lack of extensive ecological, genetic, and biochemical information for most candidate production strains
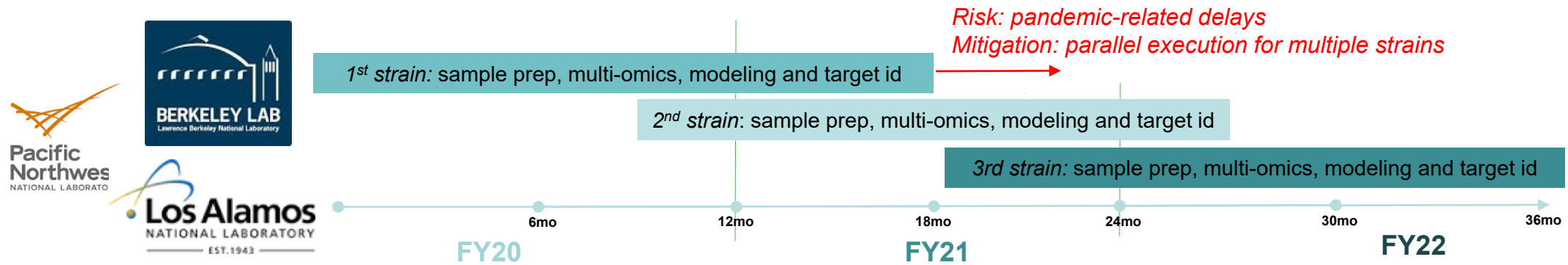
**Goal:**

Develop an algal functional genomics pipeline for production and interpretation of multi-omics measurements from multi-state perturbations experiments to identify gene targets for strain improvement and commercialization

**Objectives:**

– Produce high quality reference genomes for target algal strains

– Generate multi-omics data for system level analysis and modeling

– Identify gene targets for genetic modification to improve biomass production rates

# 1 – Management



Risk: pandemic-related delays
Mitigation: parallel execution for multiple strains

*1st strain*: sample prep, multi-omics, modeling and target id

*2nd strain*: sample prep, multi-omics, modeling and target id

*3rd strain*: sample prep, multi-omics, modeling and target id

| 6mo | 12mo | 18mo | 24mo | 30mo | 36mo |

FY20      FY21      FY22

**Participants:**
  *PNNL:* LEAPS experiments, sample collection and distribution
  *LANL:* transcriptomics+proteomics, integrative analysis
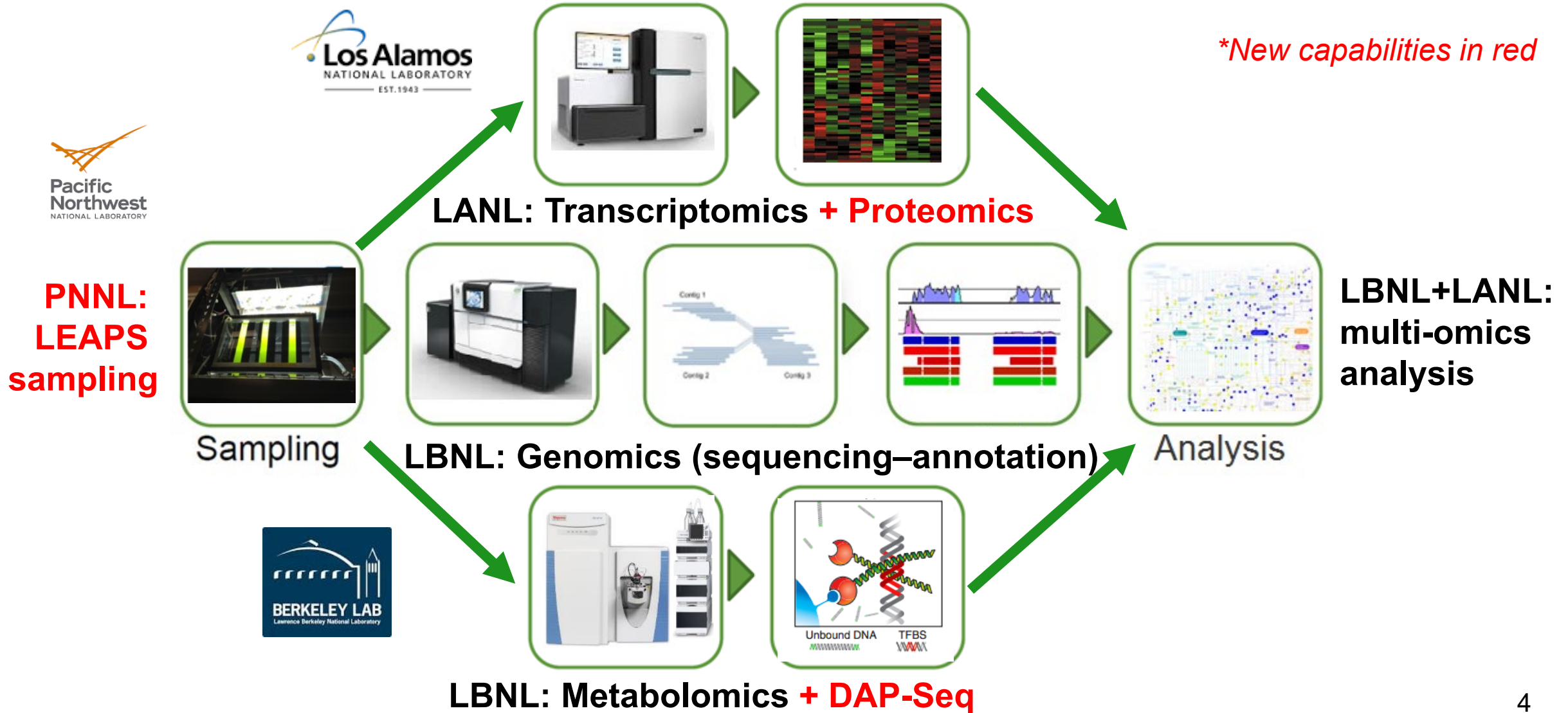  *LBNL:* genomics, metabolomics, DAP-Seq, integrative analysis

**Strains:**
  1. *Scenedesmus obliquus UTEX 393*
  2. *Picochlorum celeri (ExxonMobil)*
  3. *TBD strain from BETO DISCOVR*

**Communications:** bi-weekly postdoc calls, bi-weekly PI calls, monthly BETO calls

# 2 - Approach: Functional Genomics Pipeline

Built on the National Labs capabilities in algal growth, genome sequencing, multi-omics profiling, integrative analysis, gene network and metabolic modeling
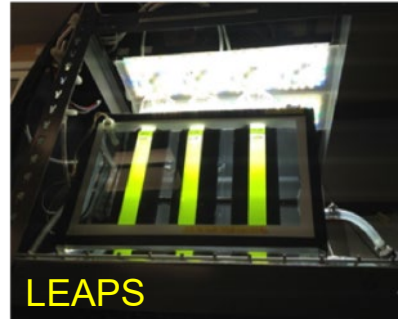


*New capabilities in red*

**PNNL: LEAPS sampling**

**LANL: Transcriptomics + Proteomics**

**LBNL: Genomics (sequencing–annotation)**

**LBNL+LANL: multi-omics analysis**

Sampling

Analysis

**LBNL: Metabolomics + DAP-Seq**

Unbound DNA    TFBS

4

# 2 - Approach: Multi-omics for Algal Strain Improvement

Platform for production and interpretation of multi-omics measurements (genome, transcriptome, metabolome, etc.) to identify gene targets for strain improvement
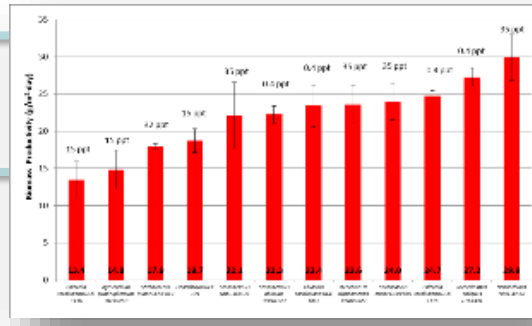
**Target strains**
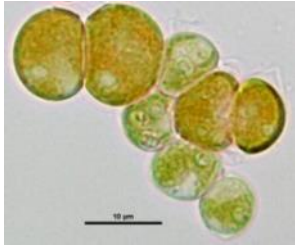


*Scenedesmus* sp. NREL 48B-D3

LEAPS

DISCOVR strain profiling

AzCATI outdoor testbeds in Mesa, AZ

**Genome sequencing and annotation**

**Gene co-expression network**

**Metabolomics**

**Validation of mutants**

Michael Guarnieri, NREL

**Target genes**

Calhoun et al., *Comm. Biology* 2021
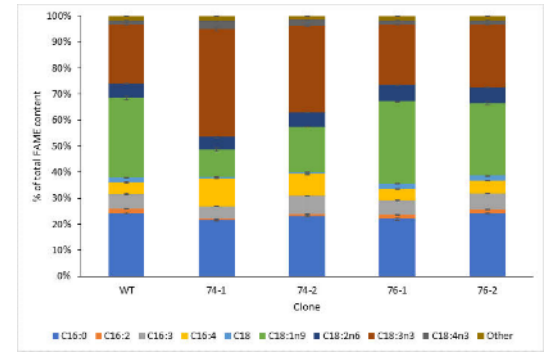
# 3 – Impact

- Algal strains with sequenced and annotated genomes and custom modeling tools will advance the state of technology and accelerate strain development and commercial viability of biomass and biofuels. This will enable improving productivity of algal strains by selection, screening, breeding, mixing cultures, and/or genetic engineering (Aft-C, Biomass Genetics & Development)

- *Strains selection* driven by DISCOVR screening and identification of high performing algae including industry-relevant strains (e.g., *Picochlorum celeri* by ExxonMobil) meeting/exceeding MYPP performance goals for the summer seasonal areal productivity (25 g/m$^2$/d in 2025).

- *Data disseminated* via JGI PhycoCosm (https://phycocosm.jgi.doe.gov) to support the broader science community and industrial stakeholders

- *Methodology shared* in publications (e.g., Calhoun et al, Comm Biology 2021) and conference presentations

- *Gene targets predicted* by the Blueprint can be *validated* through gene modulation in partner labs (e.g., NREL 46B-D3 validated at NREL)
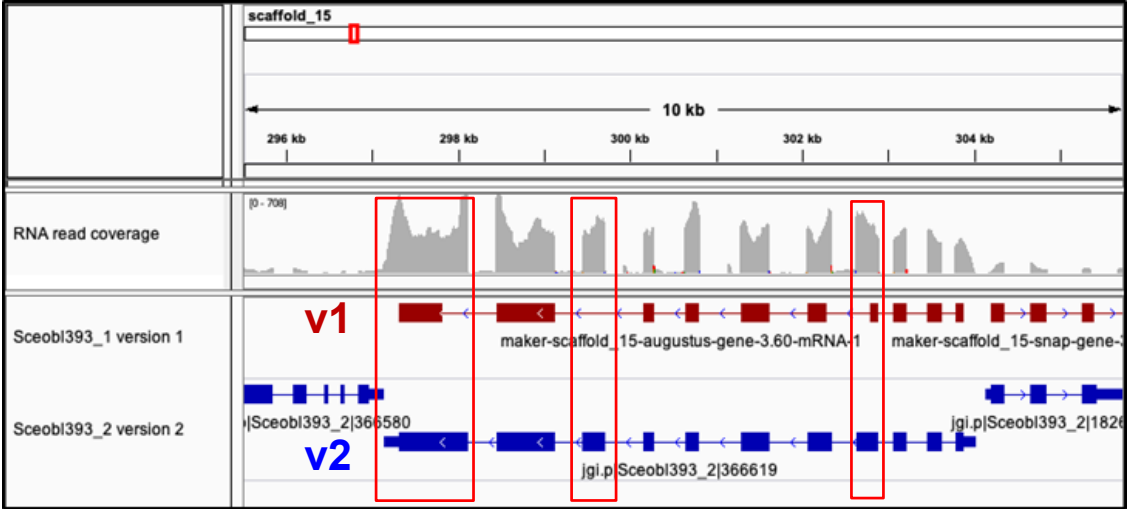
# 4 – Progress and Outcomes: Improved *S. obliquus UTEX* 393

Re-annotated UTEX 393: more genes, improved functional assignments, and curated gene models
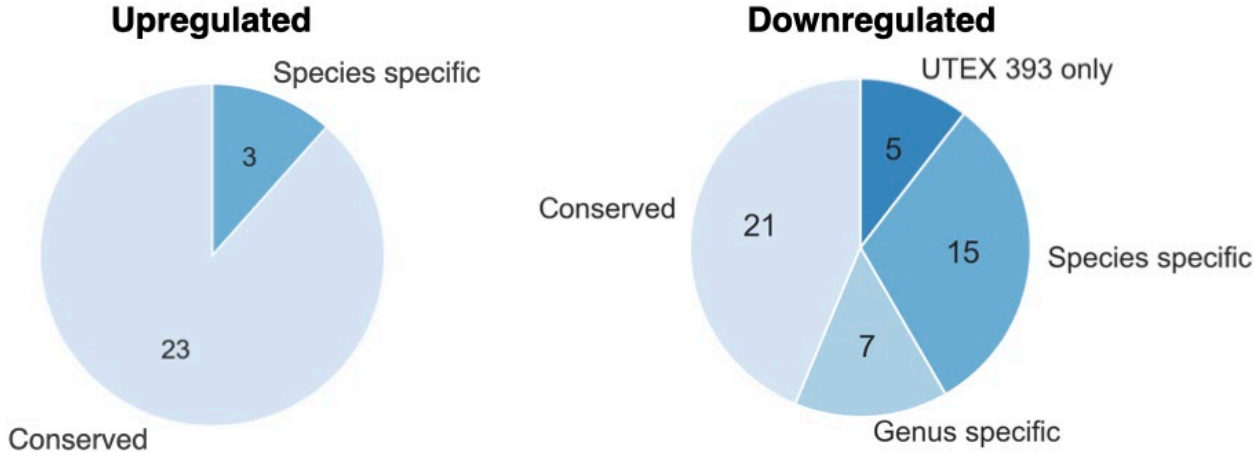


*Scenedesmus UTEX 393*
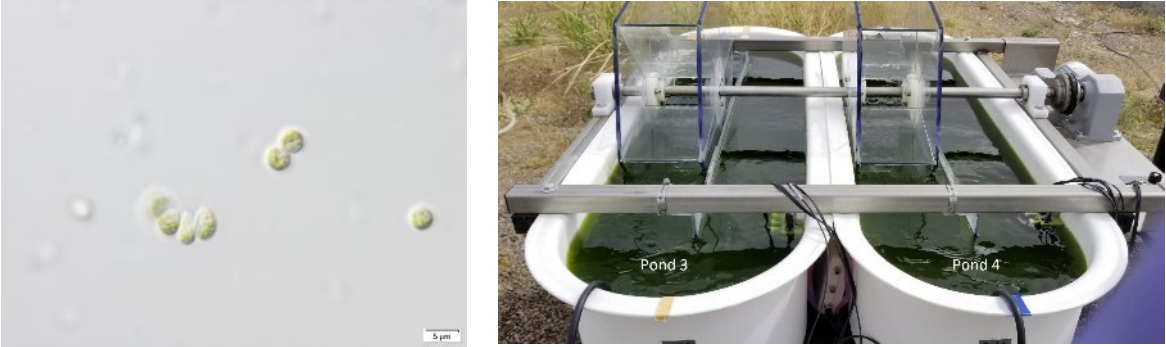
UTEX 393 in culture and at PNNL testbed site in AZ

| | **V1** (Carreres, 2017) | **v2** |
|---|---|---|
| Assembly size | 107.7Mbp | 107.9Mbp |
| #scaffolds/#contigs | 1,368 / 1,532 | 329 / 390 |
| #Genes | 16,779 (No RNA) | 19,017 |
| #Unique domains | 3,025 | 3,525 |
| Completeness | 94.5% / 68% | 97.8% / 88% |



Curated transcription factors (TF): **v2** vs. **v1** improvements



**Upregulated**

Species specific — 3
Conserved — 23

**Downregulated**

UTEX 393 only — 5
Species specific — 15
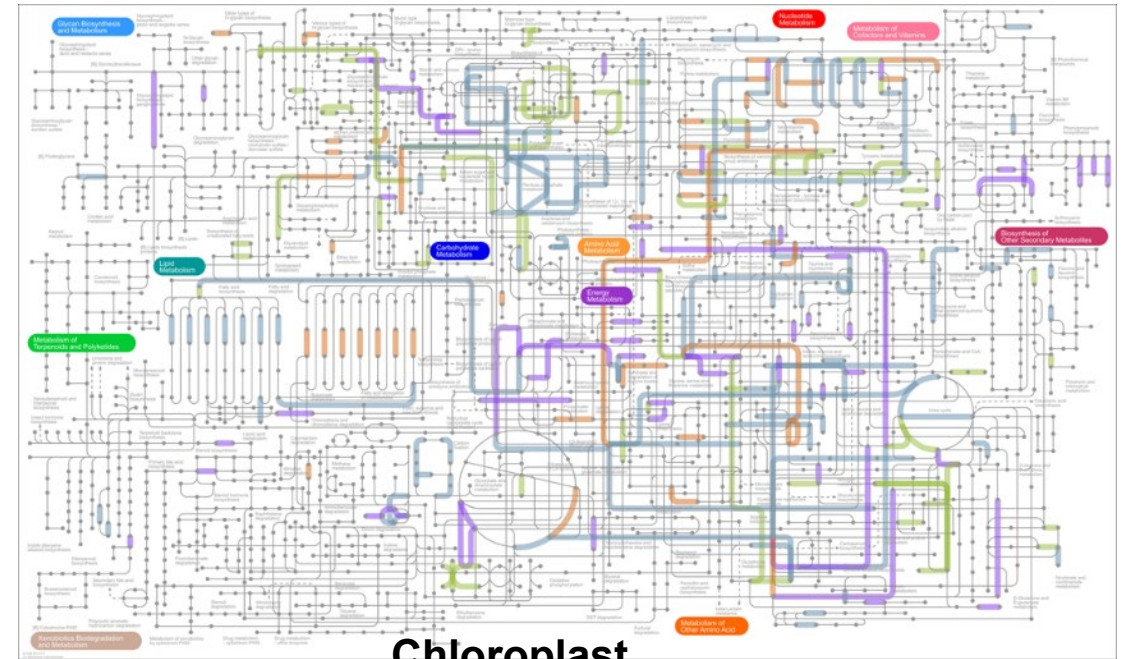Genus specific — 7
Conserved — 21

Differentially expressed and conserved TFs for DAP-Seq

# 4 – Progress and Outcomes: Metabolic Reconstruction

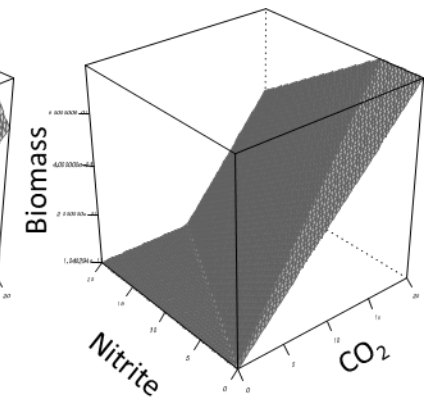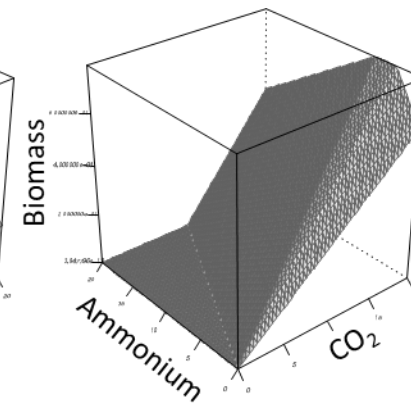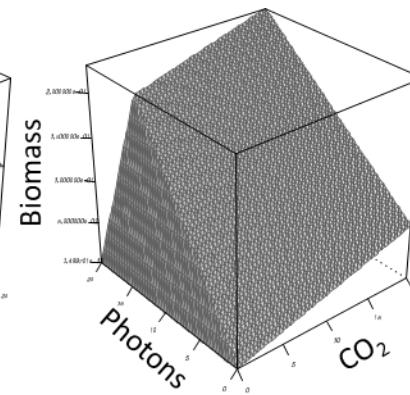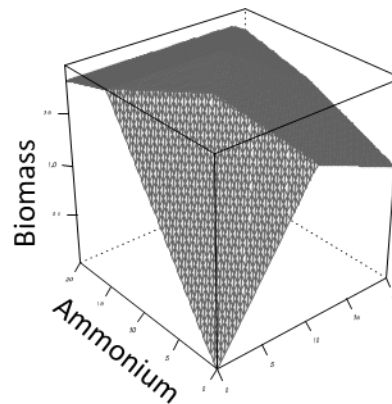## Comparative metabolic reconstruction of UTEX 393 (Go/No-Go) and other algal strains
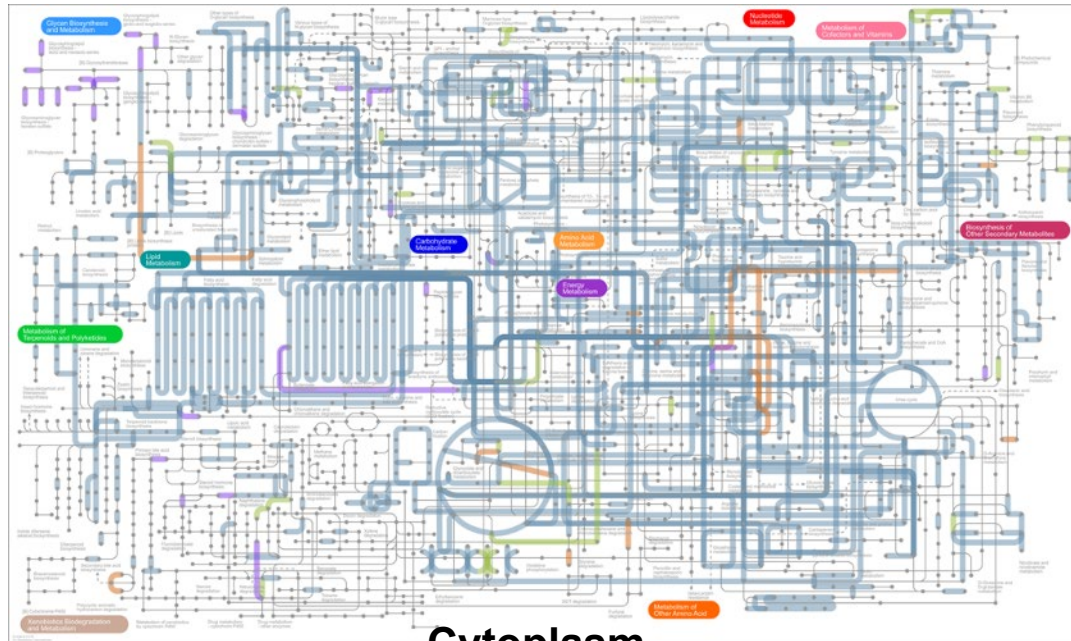


**Cytoplasm**



**Chloroplast**

Shared

Unique to *M. minutum*

Unique to *S. obliquus* UTEX393
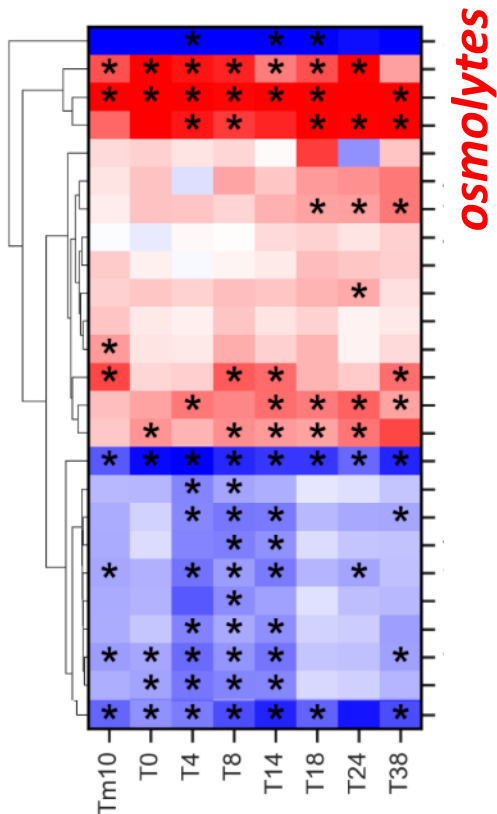
Unique to *C. reinhardtii*

*Flux Balance Analysis* model captures interactions of various nutrients and their impact on the growth function.
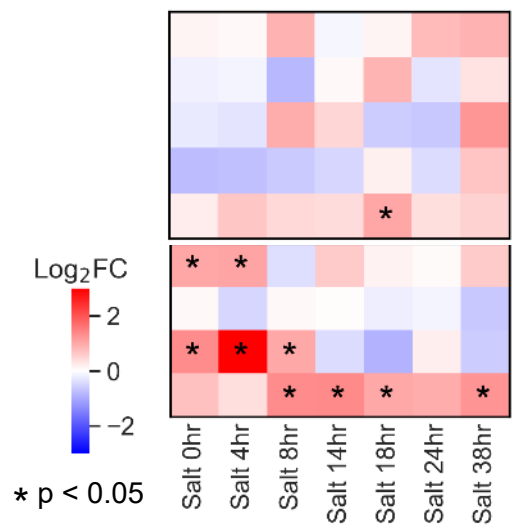


Phenotypic Phase Plane analysis graphs

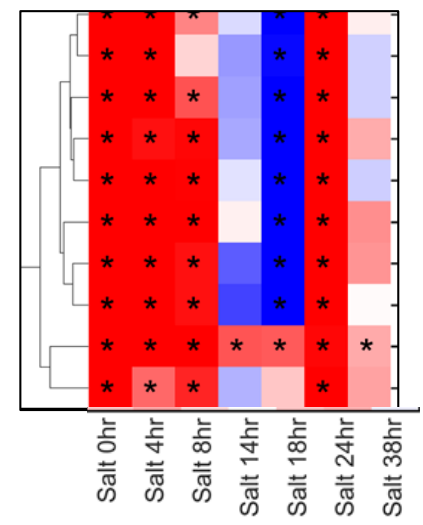# 4 – Progress and Outcomes: High Salinity Multi-omics

- Salt stress can improve lipid production
- Salinity may protect crops and inform engineering salt tolerant strains
- Genes responsive to salinity identified

**Polar metabolites**



*osmolytes*

Light/dark cycle

Time-course multiomics measurements in *S. obliquus* UTEX 393

T = 0 4 8 14 18 24    38hr    Sampling

15 ppt
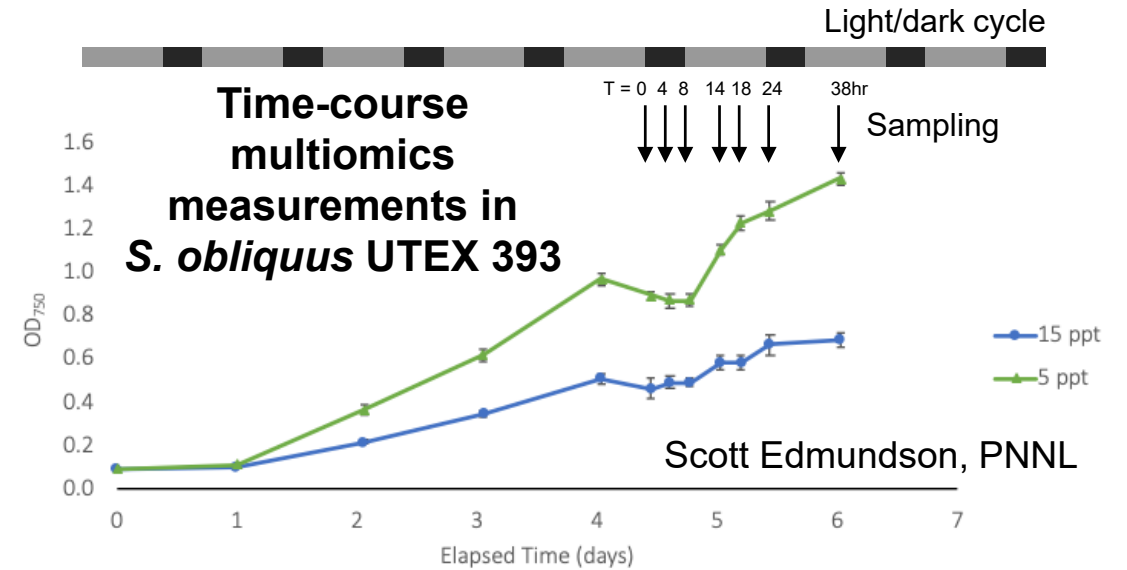5 ppt

Scott Edmundson, PNNL
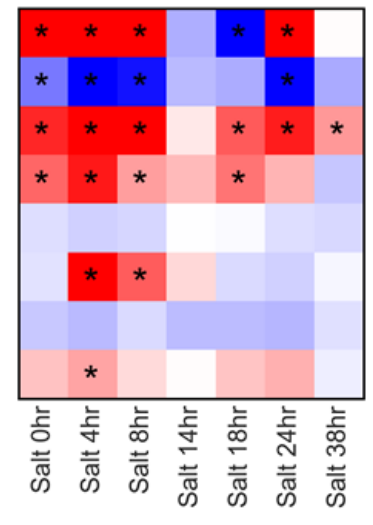
**Gene Expression Analysis**

Osmolyte biosynthesis        Transcription factors        Transporters

\* p < 0.05

9

# 4 – Progress and Outcomes: *Picochlorum celeri*

- The 27.4Mbp *P.celeri* genome was sequenced and published by ExxonMobil consortium
- We predicted ~7,340 genes per haplotype, >90% supported by transcriptomics or protein similarity
- Transcriptomics/metabolomics of samples collected from LEAPS runs are in progress
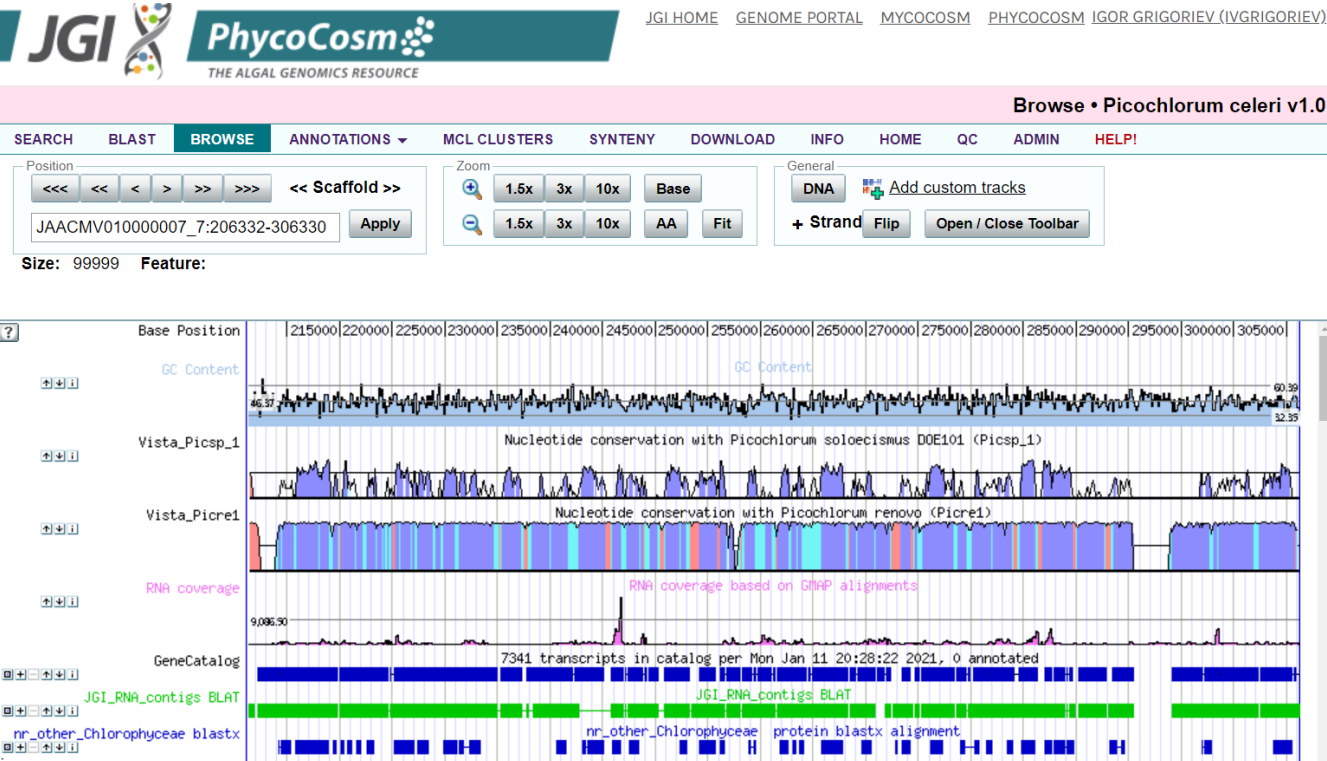




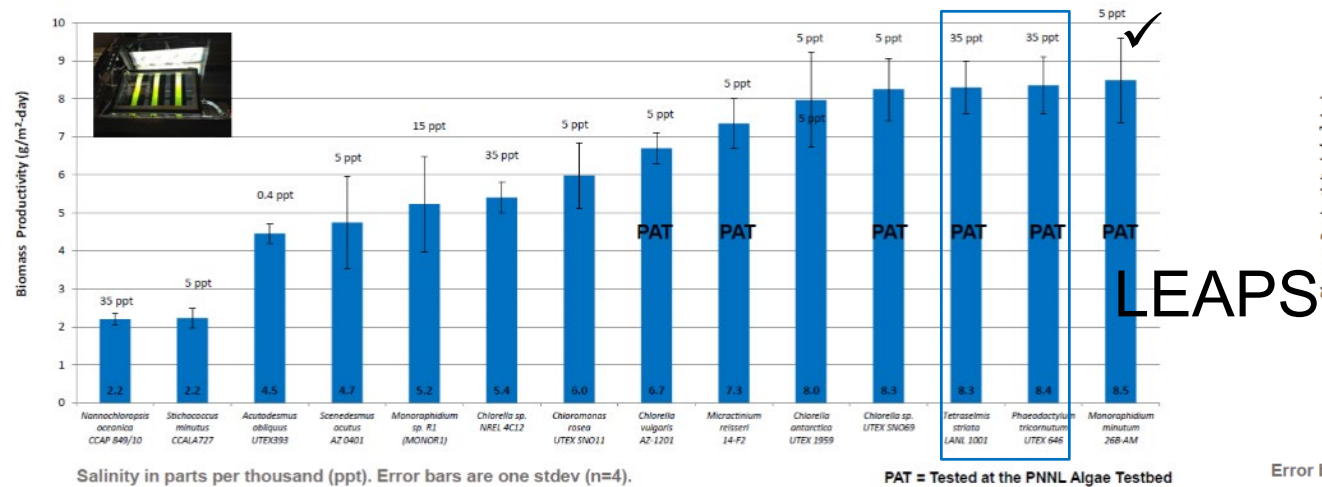Number of Annoated proteins across KEGG categories
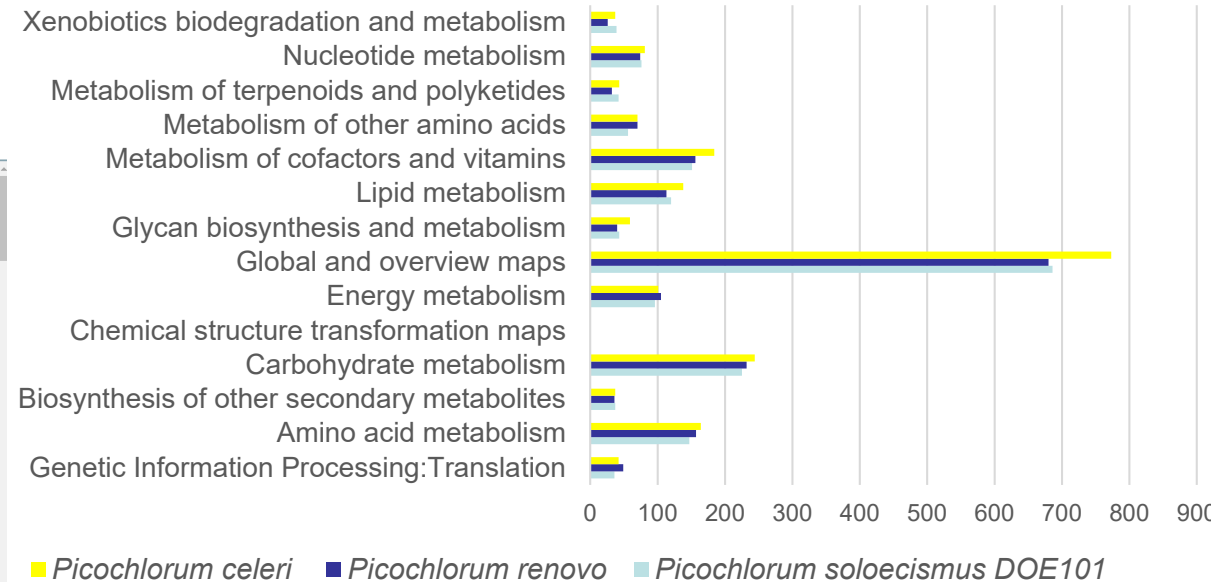
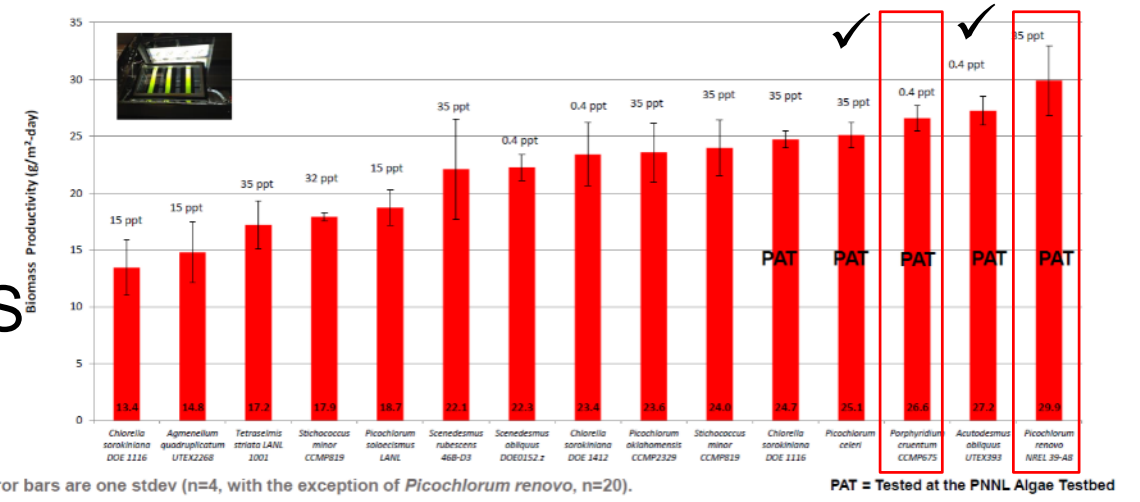# 4 – Progress and Outcomes: Future Work

Characterize multi-omics of DISCOVR highly productive strains and target genes for strain improvement



**The Blueprint Project 3rd year candidates from DISCOVR:**
- *Picochlorum renovo* NREL 39-A8
- *Tetraselmis striata* LANL1001
- *Phaeodactylum tricornutum* UTEX646
- *Porphyridium cruentum* CCMP675

# Summary

- **Improved the Blueprint pipeline** for multi-omics characterization of algal strains, both *in depth* (newly added proteomics and DAP-Seq capabilities) and *throughput* (3 strains in FY20-22 vs. 2 strains in FY17-19)

- **Disseminated** *data* (public JGI PhycoCosm), *methods* (manuscripts on temperature stress in *Scenedesmus* in press and on salt stress in *M. minutum* in preparation), and *gene targets for validation* (NREL 48B-D3 TFs validated at NREL)

- **Annotated genomes** of highly productive (> 25 g/m$^2$/d) salt-tolerant strains *S. obliquus* UTEX 393 and *P. celeri* to inform *metabolic modeling and regulatory network analysis*

- **Coordinated with BETO DISCOVR** to identify and rapidly characterize *industry-relevant strains* and *systems level responses* under same conditions (LEAPS system)

# Acknowledgements

# ADDITIONAL SUPPORTING SLIDES

# Quad Chart Overview

## Timeline
- 10/1/2019
- 9/30/2022

| | FY20 | Active Project |
|---|---|---|
| **DOE Funding** | (10/01/2019 – 9/30/2020) $350K LBNL $350K LANL $25K PNNL | (10/01/2020 – 9/30/2022) $600K LBNL $600K LANL $50K PNNL |

## Project Partners*
PNNL, Michael Huesemann (DISCOVR)

NREL, Micheel Guarnieri

## Barriers addressed
**Aft-C.** Biomass Genetics and Development: The productivity and robustness of algae strains against temperature and salinity could be improved by selection, screening, breeding, mixing cultures, and/or genetic engineering.

*Project Goal:*
Develop an algal functional genomics pipeline for production and interpretation of multi-omics measurements from multi-state perturbations experiments to identify gene targets for strain improvement and commercialization.

*End of Project Milestone:*
Using the improved and expanded milti-omics pipeline identify gene targets for strain improvement in *Scenedesmus obliquus* UTEX 393, *Picochlorum celeri*, and 3rd TBD target strain from DISCOVR

## Funding Mechanism

AOP

# FY21-22 Milestones

| | |
|---|---|
| Complete proteome, transcriptome, metabolome, and Dap-seq for *Scenedesmus obliquus* UTEX 393 @ LBNL and LANL | 12/31/2020 delayed due to SIP |
| Complete *Picochlorum celeri* LEAPS runs @ PNNL, collect and distribute samples | ✓ 12/31/2020 |
| Annotate *P. celeri* genome @ LBNL and complete RNA sequencing @ LANL | 3/30/2021 |
| Complete metabolomics (polars and lipids) for *P. celeri* @ LBNL | 6/30/2021 |
| Preliminary metabolic and regulatory models of *Scenedesmus obliquus* UTEX 393 constructed from an integrated multi-omics profile @ LBNL and LANL | 6/30/2021 (go/no-go) |
| Complete *P. celeri* proteomics, transcriptomics, metabolomics, and Dap-seq for integrative analysis and target gene identification @ LBNL and LANL | 09/30/2021 (year end) |
| Environmental simulations of a TBD strain @ PNNL | 12/31/2021 |
| Genome and transcriptome sequenced and assembled of TBD strain @ LBNL and LANL | 3/31/2022 |
| A preliminary metabolic and regulatory model constructed for a TBD strain from an integrated multi-omics profile @ LBNL and LANL | 6/30/2022 |
| Two target genes identified for a TBD strain @ LBNL and LANL. One+ improved strains demonstrating improved biomass production rates and/or biomass composition/quality. | 9/30/2022 (project end) |

# Overall Project Progress, per Strain

| Algal Strain | Project Phase | Genome sequencing | Annotation | Transcriptomics | Metabolomics | Proteomics* | DAP-Seq* | Integrative analysis | Target gene ID | Publication | Validation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Secenedesmus* sp.NREL 48B-D3 | FY17-FY18 | ✓ | ✓ | ✓ | ✓ | N/A | N/A | ✓ | ✓ | ✓ | ✓ |
| *Monoraphidium minutum* 26B-AM | FY19 | ✓ | ✓ | ✓ | ✓ | N/A | N/A | ✓ | ✓ | ⟳ | |
| *Scenedesmus obliquus* UTEX 393 | FY20 | + | ✓ | ✓ | ✓ | ⟳ | ⟳ | ⟳ | | | |
| *Picochlorum celeri* | FY21 | + | ✓ | ⟳ | ⟳ | | | | | | |

✓ -complete; ⟳ -in progress; * -new capabilities added in FY21; + -sequenced before this project;