



# Algal Translational Genomics (ATG)

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

*March 2021*

*Advanced Algal Systems*

Shawn Starkenburg, Los Alamos National Laboratory (LANL)



# Project Overview

**Goal:** Use genomics to accurately characterize the metabolic potential of leading production strains to enable performance improvements

## **Motivation:**

### Lack of Curated Metabolic Models

- Un-curated Annotations/genes misannotated, plus gaps in key metabolic pathways
- IMPACT: impedes strain improvement and tailoring of cultivation conditions to maximize yield

### Incomplete Genomes for Genetic Engineering

- High degree of fragmentation- causes incomplete gene models
- long repetitive DNA elements and allelic/structural variation in polyploid genomes confound assembly algorithms (only 1 euk. genome completed from telomere to telomere!)
- IMPACT: hinders strain improvement via genetic modification, mating/breeding studies, and/or engineering of chromosomes to improve strain performance

### Lack of Knowledge of Carbon Utilization from Waste Streams



# Management



## THE TEAM:

Personnel	Role
Shawn Starkenburg	Science Direction, Administration
Joseph Msanne (NMC)	Annotation Curation and Phenotyping Studies
Justin Zhang, Hajni Daligault, Thomas Biondi	Genome Assembly, Bioinformatics, System admin.
Yuliya Kunde	Sequencing Technician

**ADMINISTRATION: Weekly Team Meetings, One-on-One progress checks/consults as needed, Performance / Merit tied to Milestone Completion, and publications. Quarterly Written Communication with BETO with 6-week pulse checks with BETO TPM via Webex.**

## RISKS & MITIGATION:

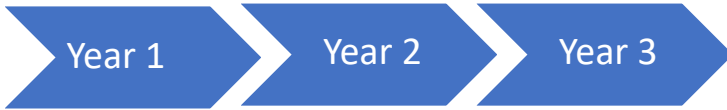
False Positives from HTP Phenotyping: Validate with Flask Studies

HMW DNA Extraction: Evaluate different Protocols, augment cell lysis protocol

Poor Scaffolding: Built In Mitigation by comparing 3 Technologies



# Approach



Task 1: Annotation  
Curation & Metabolic  
Pathway Construction

↑  
March  
2021

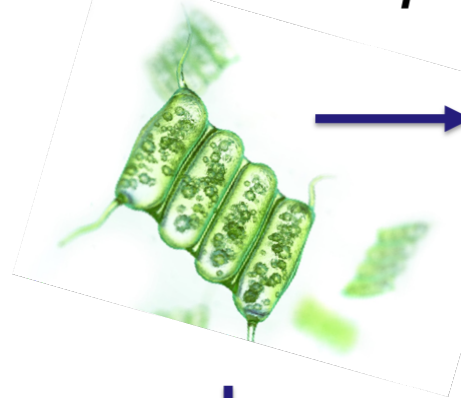
Task 2: Phenotyping and  
Validation Studies

Task 3: Advancement of Genome  
Assembly Methods

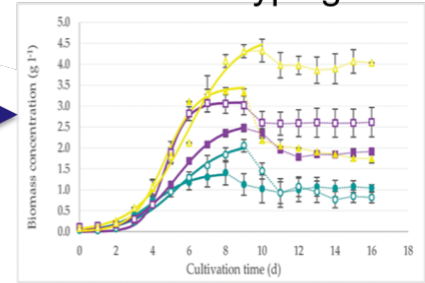
## Key Milestones:

- 1) Curated pathways for organic and inorganic carbon uptake
- 2) 100% Complete Assembly of *Scenedesmus* Genome

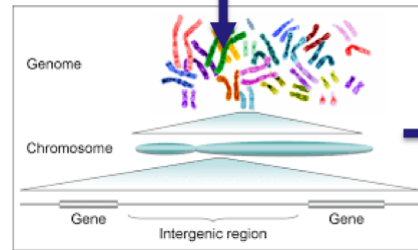
## *Scenedesmus* sp.



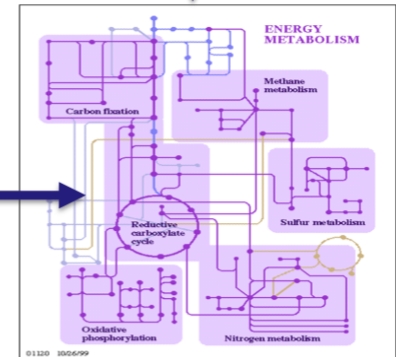
## Organic C Phenotyping



Validate



DEVELOP NEW ASSEMBLY  
METHODS

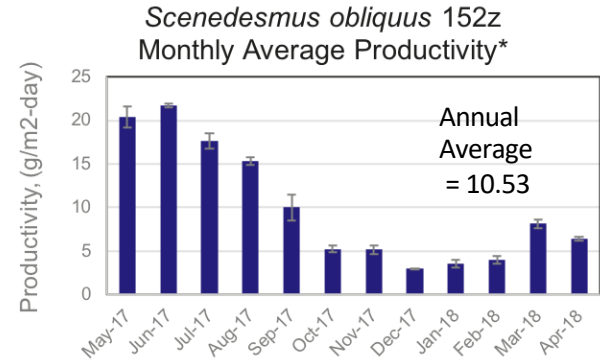


Curated C Metabolism

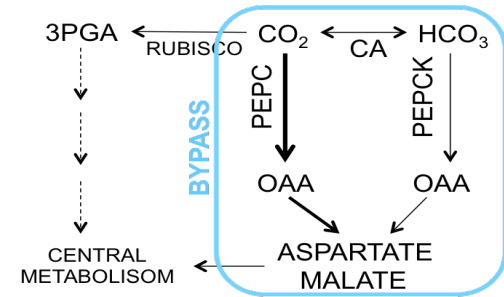


# Why *Scenedesmus*?

- **Established Baseline Productivity; contributes to SOT (among the best performers)**
- **Waste to Energy Potential; high productivity and nutrient removal rates in waste water treatment systems**
- **Genomic resources for multiple strains to assess conserved and unique metabolic properties**
  - Strain 46DB3 (Guarnieri, NREL)
  - *S. obliquus* 152Z, UTEX393
  - *S. quadricaudus*
- **Perfect Genome Size; sufficiently large to challenge scaffolding methods**
- **Life cycle is not known – (Needed for breeding strategies to improve traits)**



\*Courtesy of MicroBio Engineering



Proposed Model of Carbon Fixation  
In *Scenedesmus* (J. Polle, CUNY)



# IMPACT



## OUTCOMES

- Curation of genome annotations to develop comprehensive models of canonical (and alternative) autotrophic and heterotrophic carbon assimilation
- Experimental validation of both conserved and strain specific pathways for enhancements.
- Lays the groundwork for improving biomass production rates under carbon- and light-limited conditions in large-scale ponds, and improve carbon input cost sustainability through discovery and utilization of efficient carbon metabolism pathways.

## INDUSTRIAL RELEVANCE

- Optimizing cultivation in organic rich wastewater sources to improve productivity
- Helps achieve carbon input cost sustainability through discovery and utilization of efficient carbon metabolism pathways
- Enable academic and industrial entities to both contribute to the work and reap the benefits of BETO funded national laboratory programs
- Enables genetic manipulation and breeding of production strains to improve biomass yields/traits



# Progress



## Task 1: Annotation Curation & Metabolic Pathway Construction



Literature Search  
What is known about C metabolism in *Scenedesmus*?

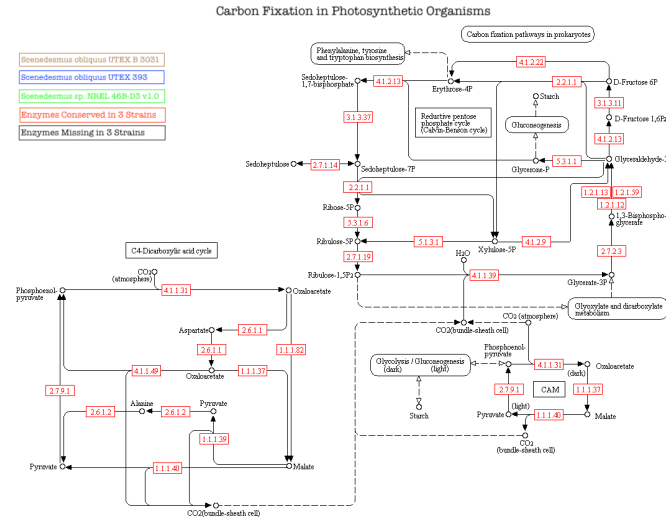
Completed a literature review to identify carbon metabolism genes/proteins and known growth response(s) in wastewater sources.  
**\*Published as a minireview in *Algal Research* -**  
<https://doi.org/10.1016/j.algal.2020.101911>.



Construct Custom C Metabolism Model



Curate Metabolic Maps with Phenotyping Data



\*Initial characterization of C metabolism in *Scenedesmus obliquus* 152Z submitted for publication.

# Progress

*Key Milestone:* Custom, *in silico* model constructed for organic and inorganic carbon uptake, regulation, and metabolism conserved in all sequenced *Scenedesmus* strains with zero gaps in proposed pathways.

## Task 1: Annotation Curation & Metabolic Pathway Construction

Literature Search

What is known about C metabolism in *Scenedesmus*?

Construct Custom C Metabolism Model

Curate Metabolic Maps with Phenotyping Data

### Manual Curation / “Gap” filling of pathways

**Reductive carboxylate cycle (CO<sub>2</sub> fixation):** 1.2.7.1; 1.3.99.1; 1.4.1.1; 2.7.9.2; 6.2.1.5  
**Photosynthesis:** 1.10.9.1 (mislabelled in the table as 1.10.99.1).  
**Carbon Fixation in Photosynthetic Organisms:** 1.2.1.59; 2.7.1.14; 2.7.9.1; 4.1.2.22; 4.1.2.9  
**Amino sugar and nucleotide sugar metabolism:** 2.7.1.6; 2.7.7.22; 2.7.7.23; 3.2.1.55; 3.5.1.33 (3.5.1.108 same description); 5.1.3.18.  
**Ascorbate and aldarate metabolism:** 1.8.5.1.  
**utanoate metabolism:** 1.2.7.1; 1.3.99.1; 3.1.1.75; 4.1.1.5; 5.1.2.3; 6.2.1.16.  
**C5-Branched dibasic acid metabolism:** 4.1.1.5; 4.1.3.25; 6.2.1.5.  
**Citrate cycle (TCA cycle):** 4.1.1.32; 6.2.1.5.  
**Fructose and mannose metabolism:** 2.7.1.105; 2.7.1.28; 3.1.3.46 (only found in UTEX B 3031 and NREL 46B-D3); 3.2.1.80.  
**Galactose metabolism:** 2.7.1.6.  
**Glycolysis / Gluconeogenesis:** 1.2.1.59; 1.2.7.1; 4.1.1.32; 5.4.2.1.  
**Glyoxylate and dicarboxylate metabolism:** 1.1.1.26; 1.1.1.29; 1.1.1.60 (); 1.1.1.79 (1.1.1.81); 1.12.7.2; 1.5.1.5; 3.1.3.18; 3.5.1.31 (3.5.1.88 same description); 3.5.4.9  
**Pentose and glucuronate interconversions:** 2.7.1.16 (2.7.1.47 same description); 2.7.1.45.  
**nositol phosphate metabolism:** 1.2.1.27; 2.7.1.134; 2.7.1.140 (2.7.1.151); 2.7.1.158 ; 3.1.3.26; 3.1.3.57; 3.1.3.62 (3.1.3.8 same description); 3.1.3.66; 3.1.3.67.  
**Pentose phosphate pathway:** 1.1.1.44; 5.4.2.7.  
**Propanoate metabolism:** 2.3.1.8.  
**Starch and sucrose metabolism:** 2.4.1.12; 2.4.1.15; 3.1.3.24; 3.2.1.3.



# Progress

*Key Milestone:* Studies completed to screen for improved growth in mixotrophic and heterotrophic growth on C2-C6 monomers and HMW polymers over autotrophic conditions.

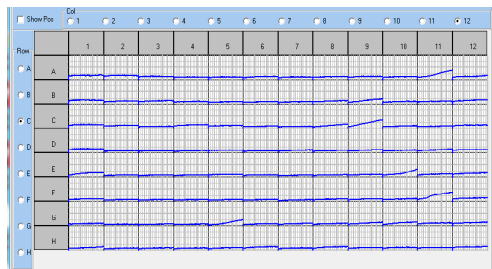
## Task 2: Phenotyping and Validation Studies



Carbon Utilization Studies  
C2-C6, Polymers



### HTP Screening



Measure systems level response +/- C source, High and low CO<sub>2</sub>

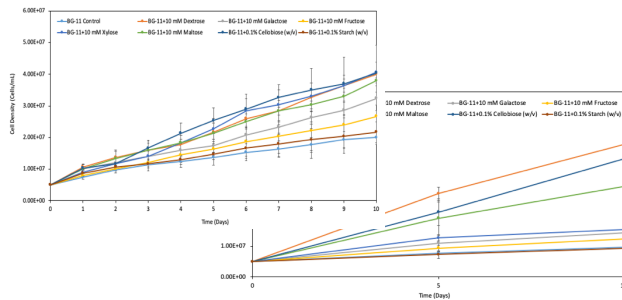


Targeted Metabolic Analysis  
<sup>13</sup>C-CO<sub>2</sub> or Bicarbonate



Curate Metabolic Maps with Phenotyping Data

### + Validation Studies



BG 11 Medium (+Substrate)	Light	Dark
10 mM Xylose		†
10 mM Ribose		
10 mM Sorbitol		
10 mM Fructose	†	
10 mM Galactose		†
10 mM D Glucose		‡
10 mM Maltose		
0.1% Cellobiose (w/v)		
0.1% α Cellulose (w/v)	†	†
0.1% Starch (w/v)		
0.1% Pectin (w/v)		
0.01% Guar (w/v)		
0.1% N acetyl D galactosamine (w/v)		†
0.1% Chitin (w/v)	†	
0.1% Chitosan (w/v)		

**Go/No Go:** 30% improved growth in mixotrophic or heterotrophic growth over autotrophic conditions



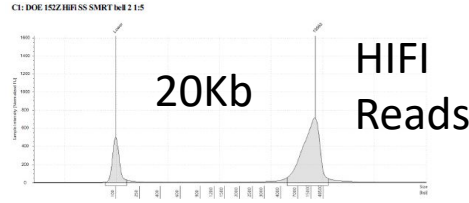
# Progress



## Task 3: Advancement of Genome Assembly Methods



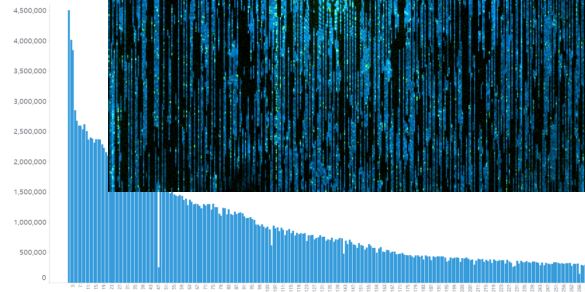
Extraction/Purification of  
HMW DNA



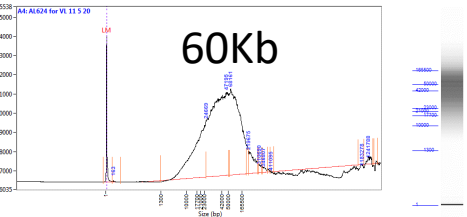
Construct Scaffolding  
Libraries



Bionano



Scaffold *S. obliquus* 152Z  
with each technology



HiC

Publish Results – Complete  
Genome + Curated  
Annotation



\*Published meta-analysis of algal genome quality and diversity

<https://doi.org/10.1016/j.algal.2020.101968>

# Progress

Key Milestone: 100% complete assembly of an algal genome

## Task 3: Advancement of Genome Assembly Methods



Extraction/Purification of  
HMW DNA



Construct Scaffolding  
Libraries



*Scaffold S. obliquus 152Z*  
with each technology



Publish Results – Complete  
Genome + Curated  
Annotation



Statistics	Improved Pacbio Assembly	Bionano Assembly Scaffolded	Hybrid Scaffolded + Unscaffolded
Count	427	86	326
Min length (Mbp)	0.003	0.179	0.003
Median length (Mbp)	0.138	1.659	0.029
Mean length (Mbp)	0.465	2.203	0.62
N50 length (Mbp)	1.297	3.292	3.04
Max length (Mbp)	4.75	6.204	6.204
Total length (Mbp)	198.517	189.434	202.206

	Draft assembly (Starkenburger et. al., 2017)	+ New HiFi Reads hifiasm	HiFi Hifiasm+ purge_dups
Scaffolds	2705	81	<b>42</b>
Contigs	2705	81	<b>42</b>
Gaps	N/A	62	<b>23</b>
Assembly Size	207,967,116 bp	111,225,458 bp	107,176,229 bp
Scaffold N50	155,544 bp	4,973,244 bp	<b>5,167,611 bp</b>
Largest scaffold	2,334,183 bp	<b>7,844,655 bp</b>	<b>7,844,655 bp</b>
BUSCO Complete %	97.3%	<b>97.4%</b>	<b>97.4%</b>
Duplicated BUSCO %	N/A	3.5%	2.0%



## Summary

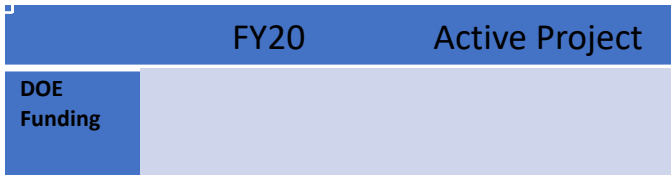
- Curated carbon metabolism pathways for a leading production strain
- Validated mixotrophic & heterotrophic carbon utilization in *Scenedesmus* 152Z
- Setting new gold standard in algal genomics
  - Published meta-analysis on the state of the art of algal genomics
  - Evaluated new assembly techniques to construct (nearly!) 100% complete genomes
- Ported Genomes from Greenhouse to JGI-Phycocosm to create a unified resource for all stakeholders
- Forms the foundation for genetic manipulation and breeding of production strains to improve biomass yields/traits



# Quad Chart Overview

## Timeline

- Project start: October 2018
- Project end date: September 2021



## Project Partners\*

- New Mexico Consortium (25%)

## Barriers addressed

- Lack of Biomass Genetics and Physiological Knowledge of Production Strains (AftC)
- Working towards MYPP goals to improve productivity in mass cultivation

## Project Goal

Use genomics to accurately characterize the metabolic potential of leading production strains to enable performance improvements

## End of Project Milestones

- Curated/validated metabolic models of inorganic and organic carbon uptake
- 100% Complete Assembly of a Diploid Algal Genome

## Funding Mechanism

Annual Operating Plan Expenditures



**ADDITIONAL SLIDES**

# Publications & Presentations

## PUBLICATIONS (6)

Joseph Msanne, Juergen Polle, S.R. Starkenburg. An assessment of heterotrophy and mixotrophy in *Scenedesmus* and its utilization in wastewater treatment. *Algal Research*, Volume 48, 2020, ISSN 2211-9264, <https://doi.org/10.1016/j.algal.2020.101911>.

Erik Hanschen, B.T. Hovde, S.R. Starkenburg. An evaluation of methodology to determine algal genome completeness. Volume 51, October 2020. *Algal Research*. <https://doi.org/10.1016/j.algal.2020.102019>.

Jenna Schambach, Anna Finck, Peter Kitin, Christopher Hunt, Erik Hanschen, Brian Vogler, Shawn R. Starkenburg, Amanda Barry. Growth, total lipid, and omega-3 fatty acid production by *Nannochloropsis* spp. cultivated with raw plant substrate. 2020. *Algal Research*. <https://doi.org/10.1016/j.algal.2020.102041>

Erik Hanschen and S.R. Starkenburg. The state of algal genome quality and diversity. *Algal Research*, Volume 50, ISSN 2211-9264, <https://doi.org/10.1016/j.algal.2020.101968>.

J. A. Ohan, B. T. Hovde, X. L. Zhang, K. Davenport, O. Chertkov, C. Han, S. Twary, **S.R. Starkenburg**. Nuclear Genome Assembly of the Microalga *Nannochloropsis salina* CCMP1776 .

Christina Steadman, Blake Hovde, Hajnalka Daligault, Xiang Zhang, Yuliya Kunde, Babetta Marrone, Scott Twary, **S.R. Starkenburg**. 2019. High-quality draft genome of the green algae *Tetraselmis striata* (Chlorophyta) generated from PacBio sequencing. *Microbiology Resource Announcements* Oct 2019, 8 (43) e00780-19; DOI: 10.1128/MRA.00780-19.

## PRESENTATIONS

- 2020, Understanding the Rules of Life: Complexity in Algal Systems Workshop (Virtual). Title: The state and quality of algal genomics. Cosponsored by NSF and DOE-EERE.
- 2020. Understanding the Rules of Life: Complexity in Algal Systems Workshop (Virtual). Title: An Evaluation of Methodology to Determine Genome Completeness.
- 2020. Sequencing, Finishing and Analysis in the Future. Genomics Conference. Title: The state and quality of algal genomics.
- 2020. ABO Summit (Virtual). Title: Analysis of Algal Genomics Demonstrates Declining Quality and Gaps in Species Distribution
- 2019, Department of Energy Peer Review, "Algal Translational Genomics", Denver, CO



# Responses to Previous Reviewers' Comments

- This AOP was a new start in FY19. No significant recommendations or suggestions for re-direction resulted from the 2019 Peer Review.



<b>ALL FY19-FY2021 ATG Milestones</b>	<b>End Date</b>
Comprehensive literature search to identify validated carbon metabolism genes/proteins and known growth response(s) in wastewater sources.	12/31/2018
Gene Inventory for all Sequenced Scenedesmus Genomes and Public Transcriptome Datasets Collected from NCBI	3/31/2019
DNA Extracted From Scenedesmus obliquus 152Z - 50 ug of High Molecular Weight (50 kbp average length) to support construction of 3 scaffolding libraries.	6/30/2019
Custom, in silico model constructed for organic and inorganic carbon uptake, regulation, and metabolism conserved in all sequenced Scenedesmus strains with zero gaps in proposed pathways.	9/30/2019
New publicly available genomes on NCBI deposited into Greenhouse (stretch)	3/31/2019
Replicated studies completed to screen for 30% improved growth in mixotrophic and heterotrophic growth on C2-C6 monomers over autotrophic conditions.	12/31/2019
New publicly available genomes on NCBI deposited into Phycosm	3/31/2020
Replicated (n=3) experiments conducted to assess for growth on high molecular weight carbon molecules (e.g. lignocellulosics, pectin, starch).	6/30/2020
Comparison of Contiguity (Telomere ends, N50, Gap Count, SNP analysis) from HiC, Bionano, and 10X Scaffolding Technologies.	9/30/2020
Transcriptome differential expression analysis comparing growth under high vs. low CO <sub>2</sub> levels in triplicate.	12/31/2020
Genome Announcement published reporting 100% complete assembly of an algal genome	3/31/2021
<sup>13</sup> C labeling of C4 metabolites and biomass production rates (AFDW over 2 week time course in triplicate) are higher than N. salina under C limiting conditions	6/30/2021
Curated/validated metabolic models of inorganic and organic carbon uptake displayed on the Greenhouse website.	9/30/2021