

The Greenhouse: A Comprehensive Knowledge Base of Algal Feedstocks

**U.S. Department of Energy (DOE)
Bioenergy Technologies Office (BETO)
2017 Project Peer Review**

Advanced Algal Systems

Shawn Starckenburg

March 7, 2017



Operated by Los Alamos National Security, LLC for the U.S. Department of Energy's NNSA

LA-UR-17-21228

Overview- History and Motivation

- **Born out of sequencing and functional genomics studies in NAABB (8 genomes, 150+ transcriptomes) and desire to genetically modify algae to improve productivity**

Limitations/Challenges in Algal Biology and Genomics

- **Limited metabolic models**
 - Few Genomes, poor annotations
 - older genomes are highly fragmented, alleles not assembled
 - Impairs accurate GMO construction and rational target selection
- **Many (Most?) 'axenic' cultures are not clonal**
 - See both distinct strains & distinct genera of algae, and lots of bacteria
- **No APPS or workflows for algae**
 - heavily biased towards bacteria and euk model systems
 - Lack of tools to handle of multi-omics experimentation, time course analysis
- **Comparative analysis difficult b/c genomes are scattered**

Goals and Outcomes

- **Create a multifunctional web portal to organize and integrate metadata with -omics profiles to accelerate strain improvement**
- **OUTCOMES**
 - Enables rational genetic engineering for productivity improvements
 - Tools for accurate species tracking in open production systems and maintenance cultures
 - High molecular weight DNA extraction methods = HQ genome assembly from long read sequencing technologies
 - Centralized Data Storage, Dissemination, Standardization
 - Improved Baseline Annotation and Physiological Potential of New Production Strains

Algae Characterization & Improvement “Org Chart”

Characterization of Productivity and Robustness

DISCOVER



Genome Sequencing, Functional ‘Omics, Metabolic Mapping

Greenhouse: Comprehensive Knowledge Base of Algal Feedstocks



Genetic Blueprint of Microalgae Carbon Productivity



Strain Improvement: Genetic Modification & Non-GM Strategies

Algae Biotechnology Partnership

Functional Characterization of Cellular Metabolism

Multi-scale Characterization of Improved Algae Strains

Breeding Algae for Long Term Stability & Enhanced Biofuel Prod'n



Strategy

Use a multi-lab consortium approach to establish a state-of-the-art platform for the deep *characterization of new strains under outdoor-relevant conditions*

Leverage the expertise at LANL and LBNL/JGI to deliver the *data and tools* required to understand algae Metabolism, Production Ecology and develop GM tools

Use a range of expertise & approaches across Labs to deliver *improved algae strains, and a suite of tools* that are effective in our top strains of interest and are also broadly applicable to new strains of interest (from DISCOVER or external stakeholders)

GOAL: *Deliver deeply characterized and improved strains, with accompanying data and tools, to stakeholders including industry, academics, and other BETO projects (e.g. BioFoundry)*

Overview

Timeline

- Project start: October 2015
- Project end : September 2018
- Percent complete: 50%

Budget

	Total Costs FY 12 – FY 15	FY 16 Costs	FY 17 Costs	Total Planned Funding (FY 18-Project End Date)
DOE Funded	\$450K	\$350K	\$350K	\$1100K
Project Cost Share	\$0	\$0	\$0	\$0

Barriers

- Lack of Biomass Genetics and Physiological Knowledge of Production Strains (AftC) and Production Ecosystems
- Supports unit-scale algae yield of 5,000 gallons biofuel/acre/year and \$3/GGE

Partners/ Collaborations

- Michael Guarnieri, NREL (0%)
- John Magnuson, PNNL (0%)
- Juergen Polle, CUNY (0%)
- Sapphire Energy (0%)

TECHNICAL APPROACH

INPUT DATA

Genomes

Functional Omics

16s/18s Metagenomes

PROCESSING

Formatting/
Standardization

ANALYSIS
&
CONTROLS

INTEGRATION

WORKFLOW DEVELOPMENT

STORAGE & LIMS

MANAGEMENT APPROACH

Support Staff:
0.5 FTE Postdoc
0.5 FTE Tech,
Weekly Meetings

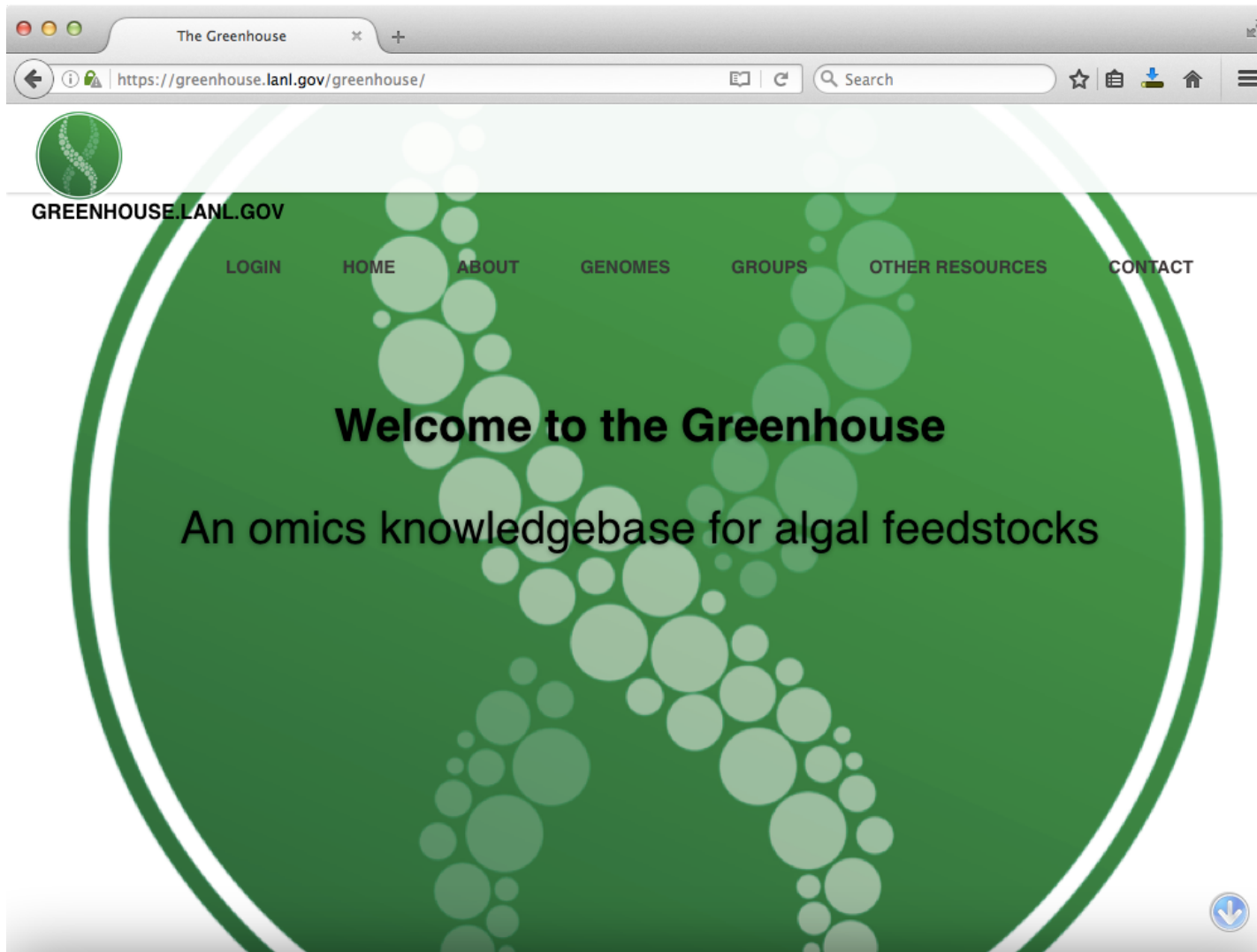
Hosted By
LANL 'Green'
Network

SHARING

PATHWAY PREDICTIONS,
ANNOTATIONS, CULTIVATION PROFILES
FORMATTED DATA

GREENHOUSE

Accomplishments –Web Interface (FY16)

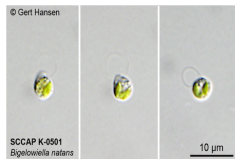
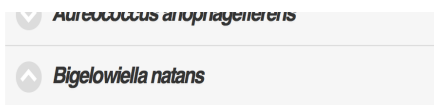




HOSTING 38 ALGAL GENOMES (FY16-17)

- Largest eukaryotic algal genome collection available online

- Species information
- Genome stats
- Downloads
- other meta-data based on availability



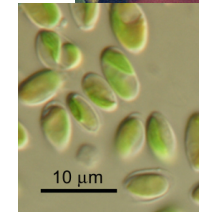
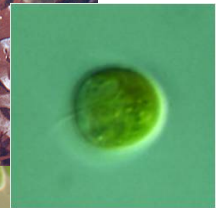
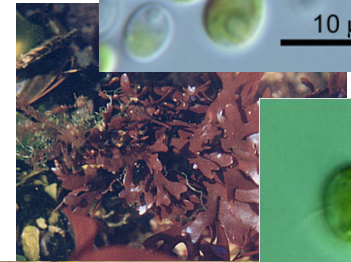
Bigelowiella natans is an amoeboflagellate algae that obtained its chloroplast by engulfing a photosynthetic eukaryote and retaining its chloroplast, a process called secondary endosymbiosis. The host cell also contains the nucleus, called nucleomorph, and the cytoplasm of the engulfed alga, though in reduced form. It belongs to family Chlorarachniophyceae. Members of this family are mixotrophic marine flagellates and amoeboflagellates with green plastids surrounded by four membranes.

Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Curtis BA, et al. Nature 2012 Dec 6. [\[LINK\]](#)

The complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Rogers MB, et al. Mol Biol Evol 2007 Jan. [\[LINK\]](#)

Complete nucleotide sequence of the chlorarachniophyte nucleomorph: nature's smallest nucleus. Gilson PR, et al. Proc Natl Acad Sci U S A 2006 Jun 20. [\[LINK\]](#)

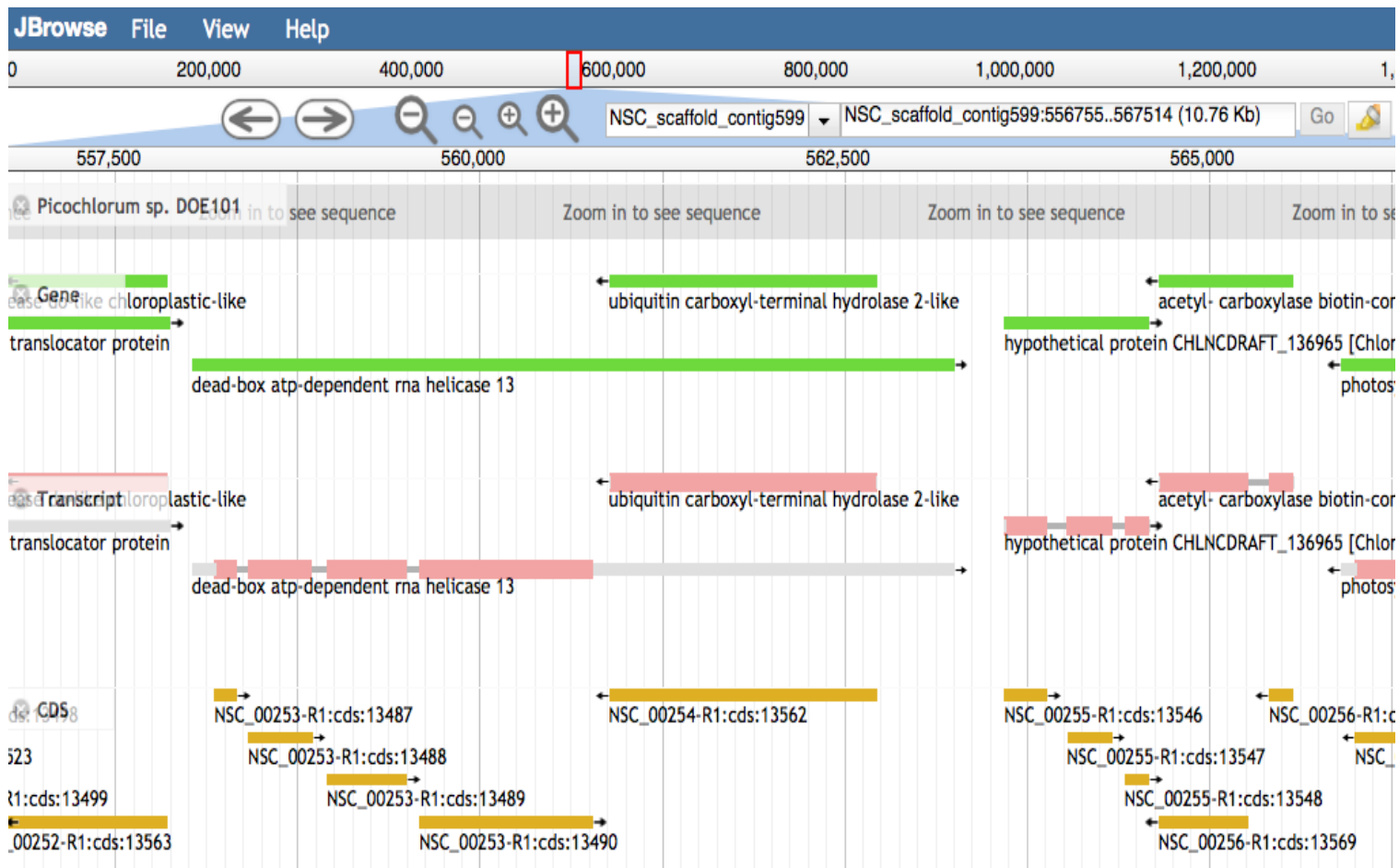
Image source <http://www.ncbi.nlm.nih.gov/genome/?term=Bigelowiella%20natans>



We are always looking to add more genomes!

Deployed a Genome Centric Browser/Viewer

- browsing and annotation visualization via Jbrowse



Enabled BLAST Search Functionality (FY16)

- BLAST and Genome Browsers for all algal species in the database

BLAST

Choose program to use and database to search:

Program Database

Enter sequence below in FASTA format

Or load it from disk No file selected.

Set subsequence: From To

The query sequence is filtered for low complexity regions by default.

Filter Low complexity Mask for lookup table only

Expect Matrix Perform ungapped alignment

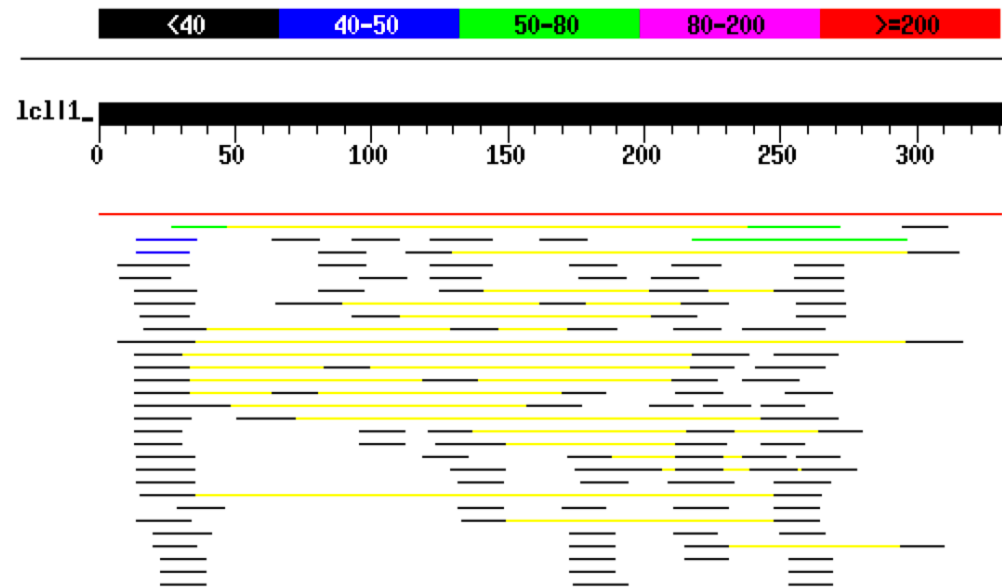
Query Genetic Codes (blastx only)

Database Genetic Codes (tblast[nx] only)

Database Genetic Codes (tblast[nx] only)

Frame shift penalty for blastx

Other advanced options:



Sequences producing significant alignments:	Score (bits)	E Value
C_sorokiniana-UTEX_1230_Contig1054	658	0.0
C_sorokiniana-UTEX_1230_Contig846	54	1e-05
Chlorella_sp._1412_CSJ.fasta.screen.Contig3336	54	1e-05
C_tobin-CCMP291_JWZX01001142.1_Chrysochromulina_sp._CCMP291_CAF....	40	0.16
C_reinhardtii-CC-503_chromosome_3	40	0.16
S_obliquus-DOE0152Z_unitig_945	38	0.61
S_obliquus-DOE0152Z_unitig_2278	38	0.61
S_obliquus-DOE0152Z_unitig_3212	38	0.61
S_obliquus-DOE0152Z_unitig_3300	38	0.61

ANNUAL MILESTONE: USER MANAGEMENT SYSTEM (FY16)

- allows groups of users to access private data sets:

User Management System

- Update Account
- Logout
- Admin Tools
- Manage Users
- Manage Groups
- Manage Projects

Login

Required to access restricted data

Email *

Password *

No account yet? [Register](#)

Private User Groups:

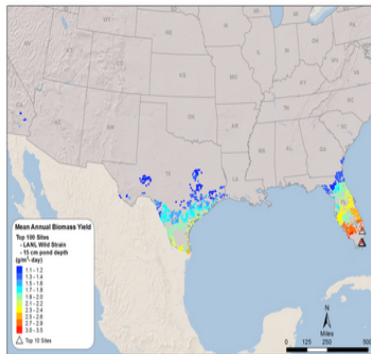
NREL
CUNY
LANL
PNNL
PACE
DISCOVER
SAPPHIRE

Stretch Goal: Create Ability to Host additional Metadata

published or unpublished species specific production data

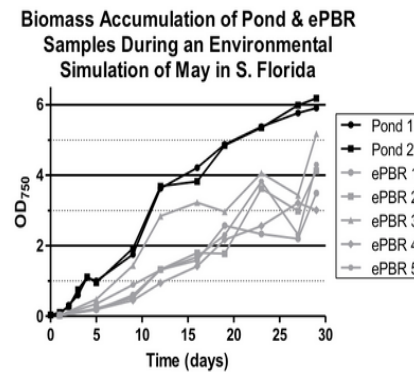
Metadata

Areal Productivity

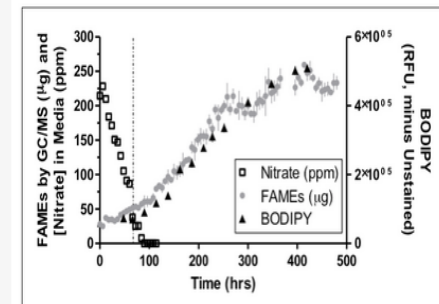


Predicted annual areal productivities for *Picochlorum* sp, using the Strain Specific Biomass Assessment Tool (Wigmosta and Huesemann, PNNL)

Biomass Accumulation



Lipid Accumulation Profile



Lipid accumulation for *Picochlorum* sp, upon nitrogen depletion (as monitored by nitrate concentration in the media) – (T. Dale and S. Twary, LANL)

Data courtesy of M. Huesemann, T. Dale

Additional Metadata

PLASMID MAPS

ePBR SIMULATIONS

MICROBIAL PROFILES

...

Home	Organisms	Tools	Browser	Omics Pathway Viewer	Blast	QIIME	Contact Us	Resources	Lo
------	-----------	-------	---------	----------------------	-------	-------	------------	-----------	----

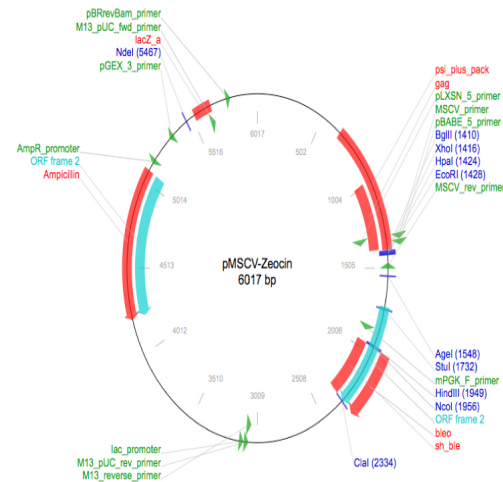
Tool(s)

Example Group 1

Example user group: (description)

Example user group files:

Mammalian Zeocin cassette vector: Mammalian expression vector: Encodes a Zeocin resistance cassette. More information at: [AddGene](#)



a. Download

Download	Link
Zeocin vector - Plasmid Map	[JPEG]

Genome Projects delivered in FY16

	NNA	4C12	14F2
Quality	Draft	Draft	Draft
Size (Mbp)	14.4	40.7	69.5
Contigs*	31	278	424
N50	1.1 Mb	617 kb	781 kb
Max	1.5 Mb	2.7 Mb	3.3 Mb

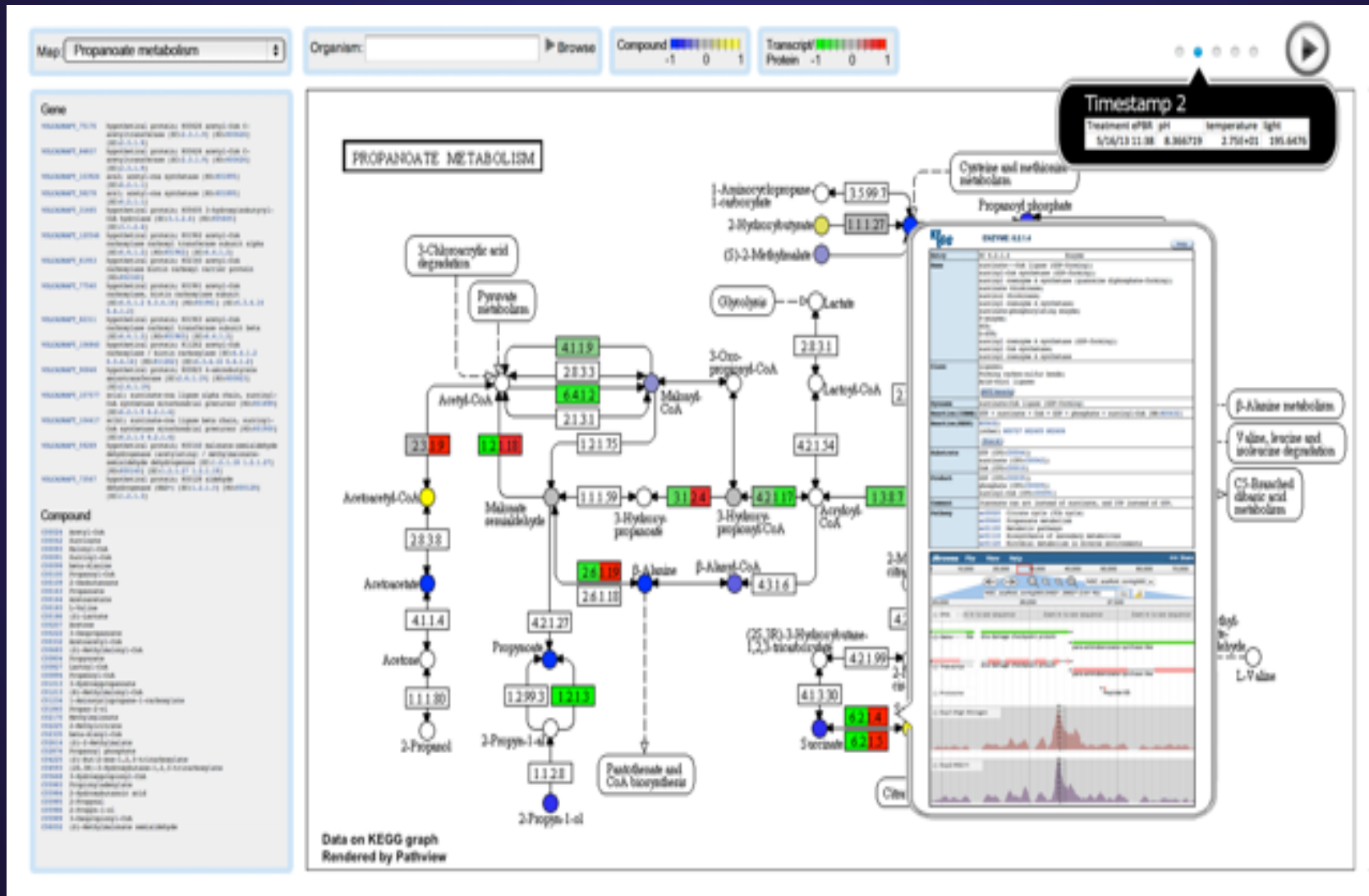
* >5kbp cutoff

Joint milestone with M. Guarnieri, NREL (Algal Biotechnology Partnership)

Chlorella sorokiniana genome improvements (FY15-16)

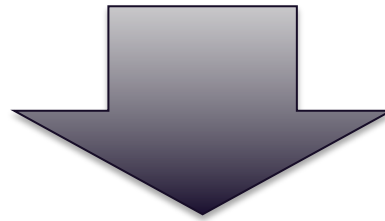
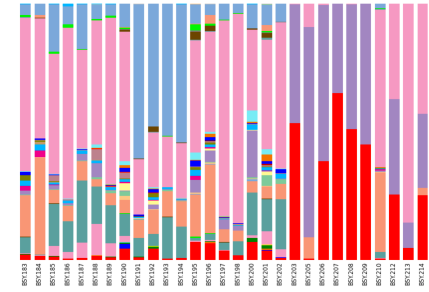
Strain	1230	1230 V2	1228	1412	1412 V2
Quality	Draft	Improved Draft	Improved Draft	HQ Draft	Improved Draft
Platforms	Illumina	Illumina + Pacbio	Opgen + Pacbio	Illumina	PacBio + Illumina
Size	56.2 Mb	59.7 Mb	61Mb	59.3	57.8
Scaffolds/ Chromosomes	N.D.	N.D.	13/12	N.D.	N.D
Contigs	10042	22	64	5949	65
N50	10.8 kb	3818 kb	2395 kb	19.5 kb	2025 kb
Max	75 kb	5.1 Mb	4.56 Mb	122 kb	5.4 Mb

Future Work: Time Course Pathway Viewer (Q3 FY17)



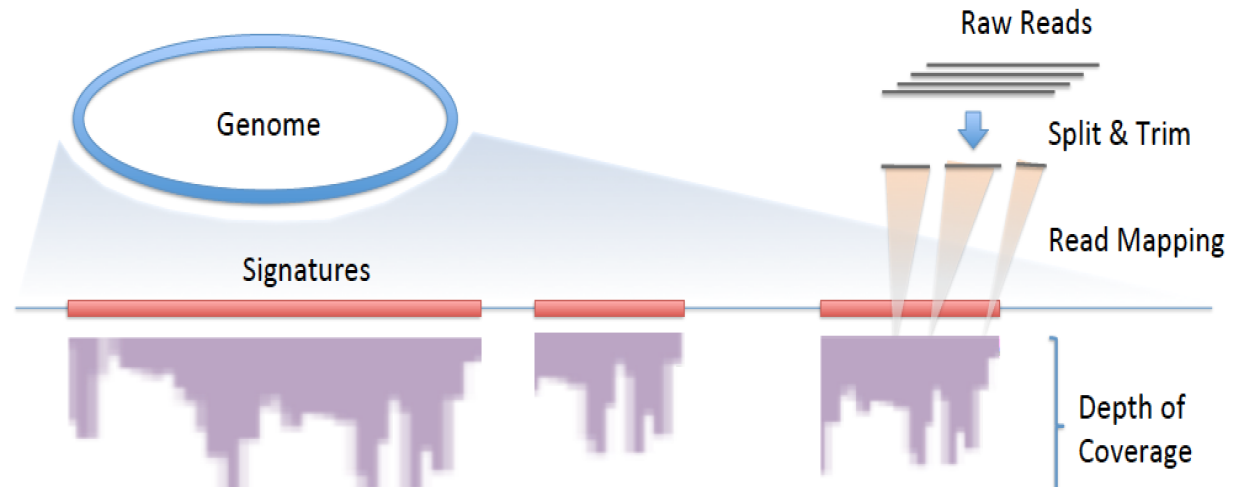
Future Work: Metagenomics Cultivation Diagnostics Tool Deployment (Q4 FY17)

- Currently utilize QIIME for 16S/18S culture diagnostics



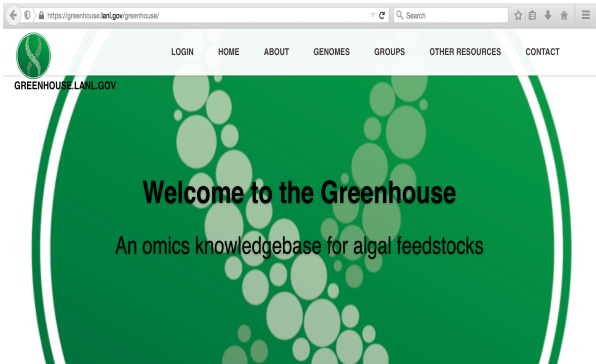
NEW: GOTTCHA

Enables Species & Strain Level Identification from all genomic signatures

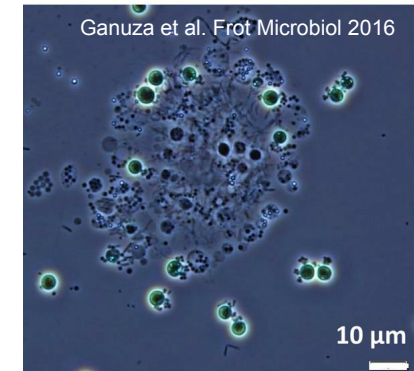


Paul Li and Patrick Chain

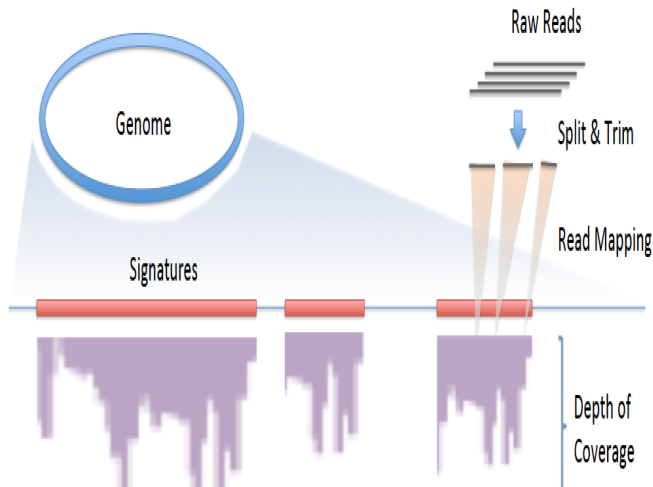
FY 18 ACTIVITY



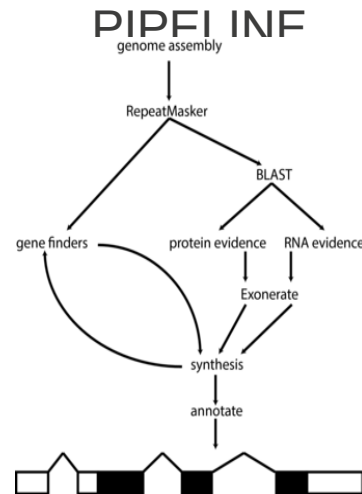
PREDATOR & PHYCOSPHERE GENOMES



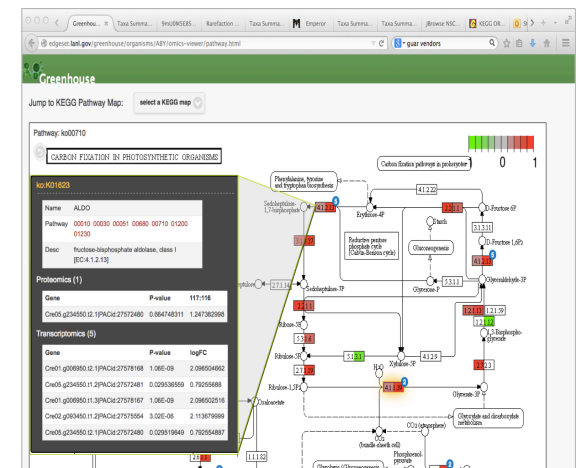
CULTIVATION DIAGNOSTICS



V2.0 ALGAL GENOME ANNOTATION



TIMECOURSE OMICS PATHWAY VIEWER





TOOLS AND KNOWLEDGE FOR INDUSTRY, ACADEMIA, GOV'T

- Building Fundamental Knowledge of Algal Biology to enable genetic manipulation of production strains to improve biomass yields
- accurate species and contaminant tracking in production systems and polycultures to enable the development of low-cost, scalable cultivation systems
- Combines -omics information with metadata from multiple resources to centralize and standardize data storage, dissemination to accurately benchmark strain improvement

Acknowledgements



Collaborators:

- Michael Guarnieri, NREL
- John Magnuson, PNNL
- Juergen Polle, CUNY
- Richard Sayre, LANL
- Taraka Dale, LANL
- A. Corcoran, Sapphire

Greenhouse Crew:

- Blake Hovde
- Patrick Chain
- Chienchi Lo
- Paul Li

Sequencing Technologies Team:

- Yuliya Kunde Hajni Daligault
- Cheryl Gleasner Karen Davenport
- Kim McMurry

BLANK

- REQUIRED ADDITIONAL SLIDES FOLLOW

Milestones Progress Summary

MEASURE	DESCRIPTION	DATE	STATUS
ML.1 QPM	Sequencing and assembly of two novel halo-tolerant algae strains.	31-Dec-15	COMPLETE
ML.2 QPM	Complete annotation for <i>S. obliquus</i> 152Z and disseminate/share via the Greenhouse website	31-Mar-16	COMPLETE
ML.3 QPM	Design of the User Management System	30-Jun-16	COMPLETE
ML.4 Annual	Deploy User Management System on Greenhouse	30-Sep-16	COMPLETE
ML.5 QPM	Manuscript Submitted for Publication Comparing the Genomes of 3 Chlorella strains	31-Dec-16	IN PROGRESS
ML.6 QPM	Database containing Unique Genome Regions for all Publicly Available Algae and Algal Predators/Prey	31-Mar-17	COMPLETE
ML.7 QPM	Software that enables integrated analysis and co-display of metabolomics, transcriptomics, and proteomics data	30-Jun-17	IN PROGRESS
ML.8 Annual	Diagnostics pipeline to monitor algae and predator/prey populations in production systems from shotgun metagenomic datasets	30-Sep-17	IN PROGRESS
ML.9 QPM	User Initiated and Web-based Genome Annotation Service	31-Dec-17	
ML.10 QPM	Sequence and Assemble 3 Predator/Pathogen Genomes from Partnering Companies/Labs	31-Mar-18	
ML.11 QPM	Link Biocyc Metabolic Models and Annotation Curation Tool to Greenhouse	30-Jun-18	
GN1	Functional User Management System	30-Sep-16	PASS

Publications, Patents, Presentations, Awards, and Commercialization

Publications:

B.T. Hovde, Y. Kunde, K. Davenport, J. Msanne, and S.R. Starkenburg. A Comparative Genomic Analysis of Three Biofuel-Relevant Strains of *Chlorella sorokiniana* Reveals High in Species Genetic Diversity. *Algal Research*. In prep.

K.R. Fixen*, **S.R. Starkenburg***, **B.T. Hovde**, S.L. Johnson, C.R. Deodato, C. S. Harwood, and R.A. Cattolico. 2016. Genome sequences of 8 bacterial species found in co-culture with the haptophyte *Chrysochromulina tobin*. *Genome Announcements*. 2016 Nov 3;4(6). pii: e01162-16. doi: 10.1128/genome A.01162-16. PubMed PMID: 27811091;

Presentations:

'Greenhouse: an omics resource for algal feedstocks'. Algal Biomass, Biofuels and Bioproducts. San Diego, CA, June 27th, 2016

'Greenhouse: an omics resource for algal feedstocks'. Algal Biomass Summit, Glendale AZ, Oct 25th, 2016

"Algal Omics Databases and Resources", Sharpening Our Tools: Algal Biology Toolbox Workshop, DOE-Bioenergy Technology Office, May 2016, San Diego, CA

"Advances in Algal Genomics", Lecturer, Spring ATP3 Workshop; Principles and Processes: Algae Culture Maintenance, Production and Downstream Processing, April 2016, Los Alamos, NM

Responses to Previous Reviewers' Comments

- **This AOP was a new start in FY16 and was not apart of the 2015 Peer Review**
- **The 2016 GO/NO GO was completed successfully and on time**