GREAT LAKES BIOENERGY RESEARCH CENTER

Managing Omics Data at GLBRC

Yury V Bukhman

Bioinfo-Core Workshop, ISMB 2016

U.S. DEPARTMENT OF ENERGY

DOE Bioenergy Research Centers

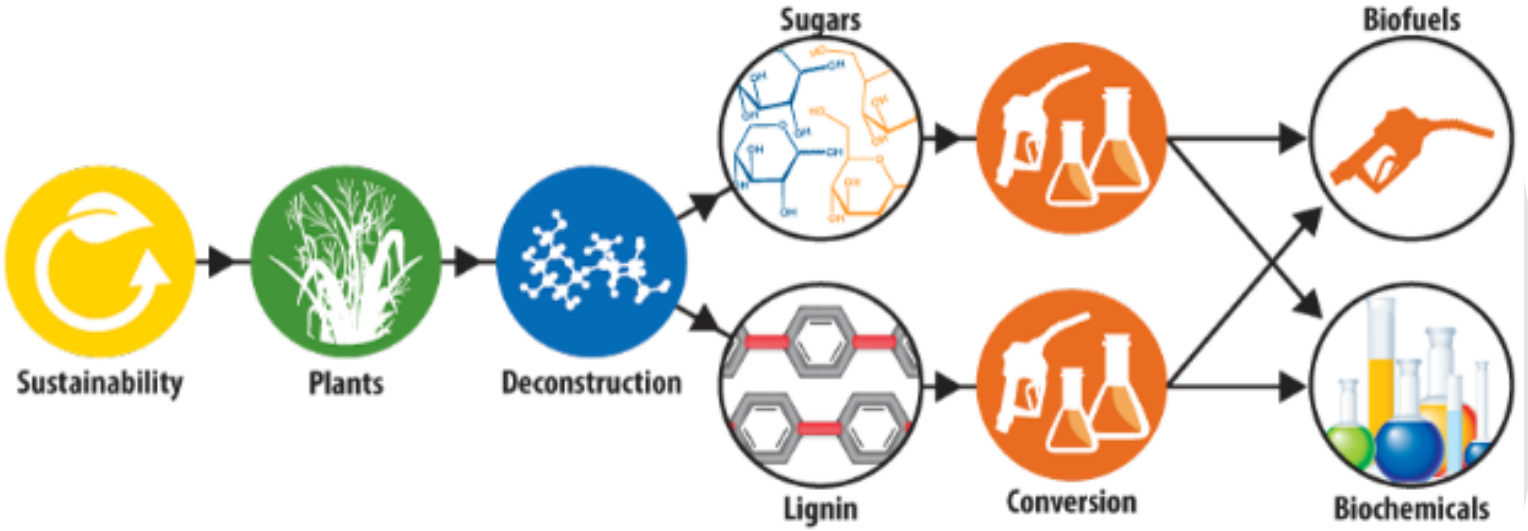
GLBRC = Great Lakes Bioenergy Research Center

- The mission of the Great Lakes Bioenergy Research Center is grand, but simply stated: to perform the basic research that generates technology to convert cellulosic biomass to ethanol and other advanced biofuels.
- DOE-funded, university-based center
- UW – Madison, Michigan State and other partners

GLBRC Structure

- Multiple projects done by academic labs led by professors
- Enabling technologies
  - Core facilities
  - Informatics and Information Technology (IIT)
    - IT
    - LIMS
    - Computational biology
      - Yury the scientist
    - MSU informatics group

GLBRC Data Management Challenges

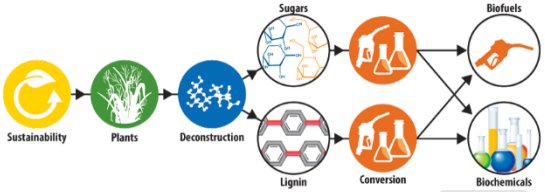


- A wide variety of research projects → many different data types, including omics
- University labs have a high degree of independence, different needs, different practices
- Geographic separation

GLBRC Data Management Solutions

- Each project has a data management plan, reviewed by IIT
  - Several projects have their own databases
- Centralized informatics resources
  - Shared file servers
  - SharePoint
  - LIMS
  - GLOW metadata database
  - GxSeq: Genome Browser, Expression Viewer etc.

Omics Data Flow



Lab samples and materials: field samples, biomass, hydrolysates…

GLBRC LIMS

Sequencing Facilities

bigdata file server

GLOW metadata database

Bioinformatics workflows

GxSeq Genome Suite

GLBRC LIMS

STARLIMS | About | Add/Remove Content | Logout

Dashboard | GLBRC Sample Viewer

Folders List

Search | Export To Excel | Submitter's Lab: GLBRC Product Chain Management

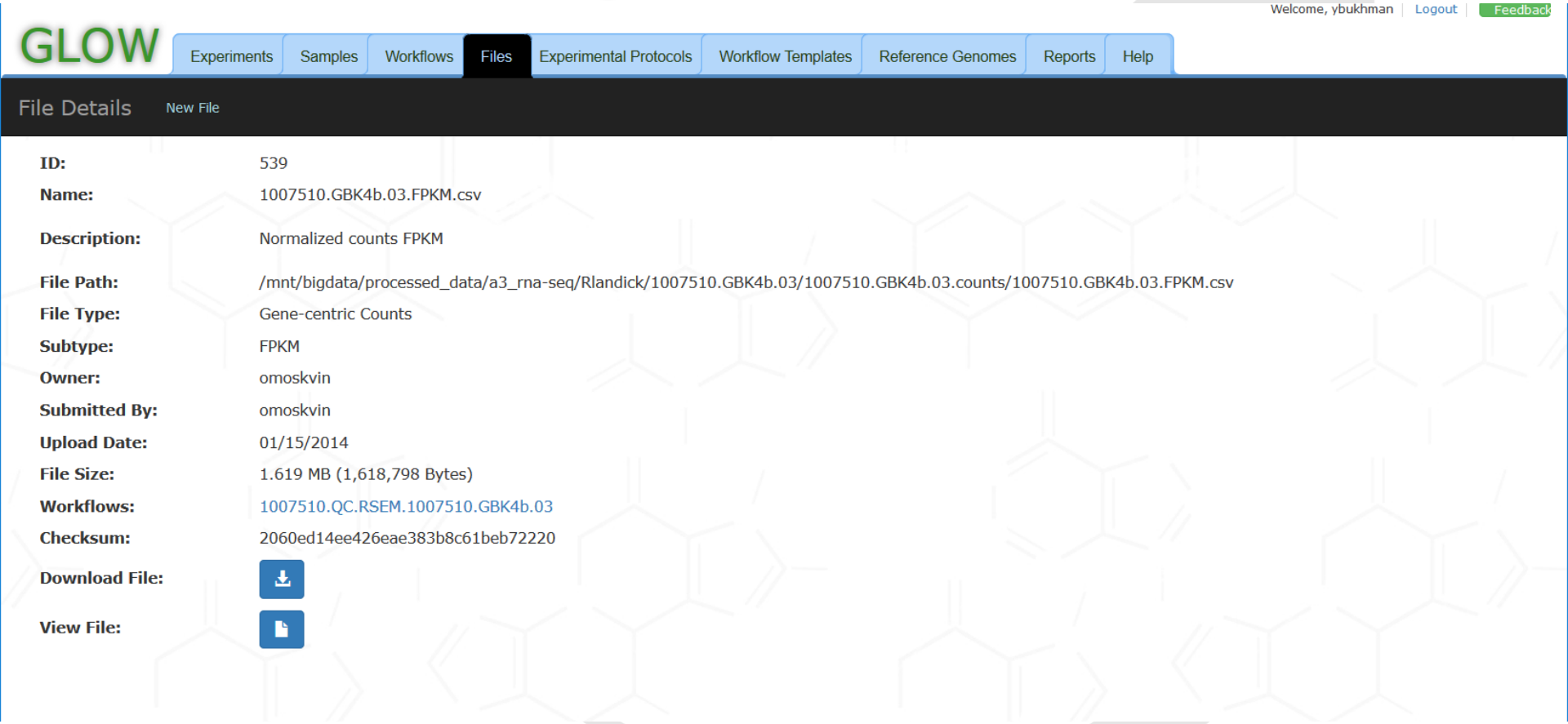
Approval S	Approved	Approved	Status	Folder #	Exp #	Service	Test Plan	Submitted By	Project #	Service Lab	Notes
Releas...	Auto	06/11...	Done	UFS140054	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/11...	Done	UFS140054	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/11...	Done	UFS140053	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/11...	Done	UFS140053	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/11...	Done	UFS140052	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/11...	Done	UFS140052	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/02...	Done	UFS140051	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/02...	Done	UFS140051	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	
Releas...	Auto	06/02...	Done	UFS140050	0	Cell Wall Analyses	Cell Wall Compositio...	RRANATUN...	5.5.2	Cell Wall ...	

Test Results | Attachments

Sample Info	Test	Analyte	Rep	Result	Final	Units	Status	Test Stat	Analyzed Da	Analyzed by	HIGH A	HIGH E	LC	LC
UFS140054001 - FSUWM000000...	Digestibility...	Glucose	1	12.957	12.957	%Yiel...	Done	Done	06/02/2...	CWLIMS				
UFS140054001 - FSUWM000000...	Digestibility...	Glucose	2	9.655	9.655	%Yiel...	Done	Done	06/02/2...	CWLIMS				

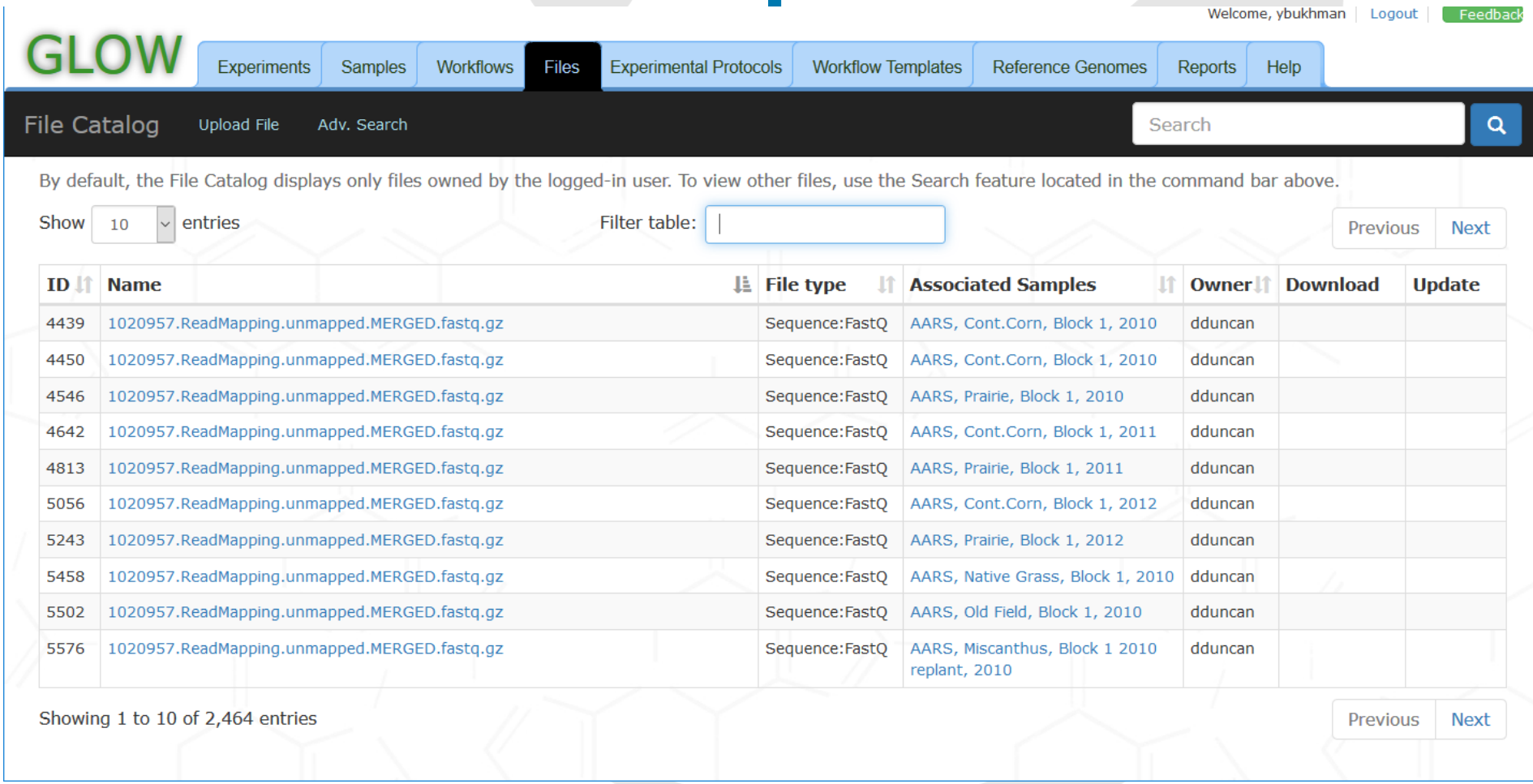
Ying Gao, Quansheng Yang, Feiran Yu, LiRong Tao, et al.

GLOW: the File Metadata Database



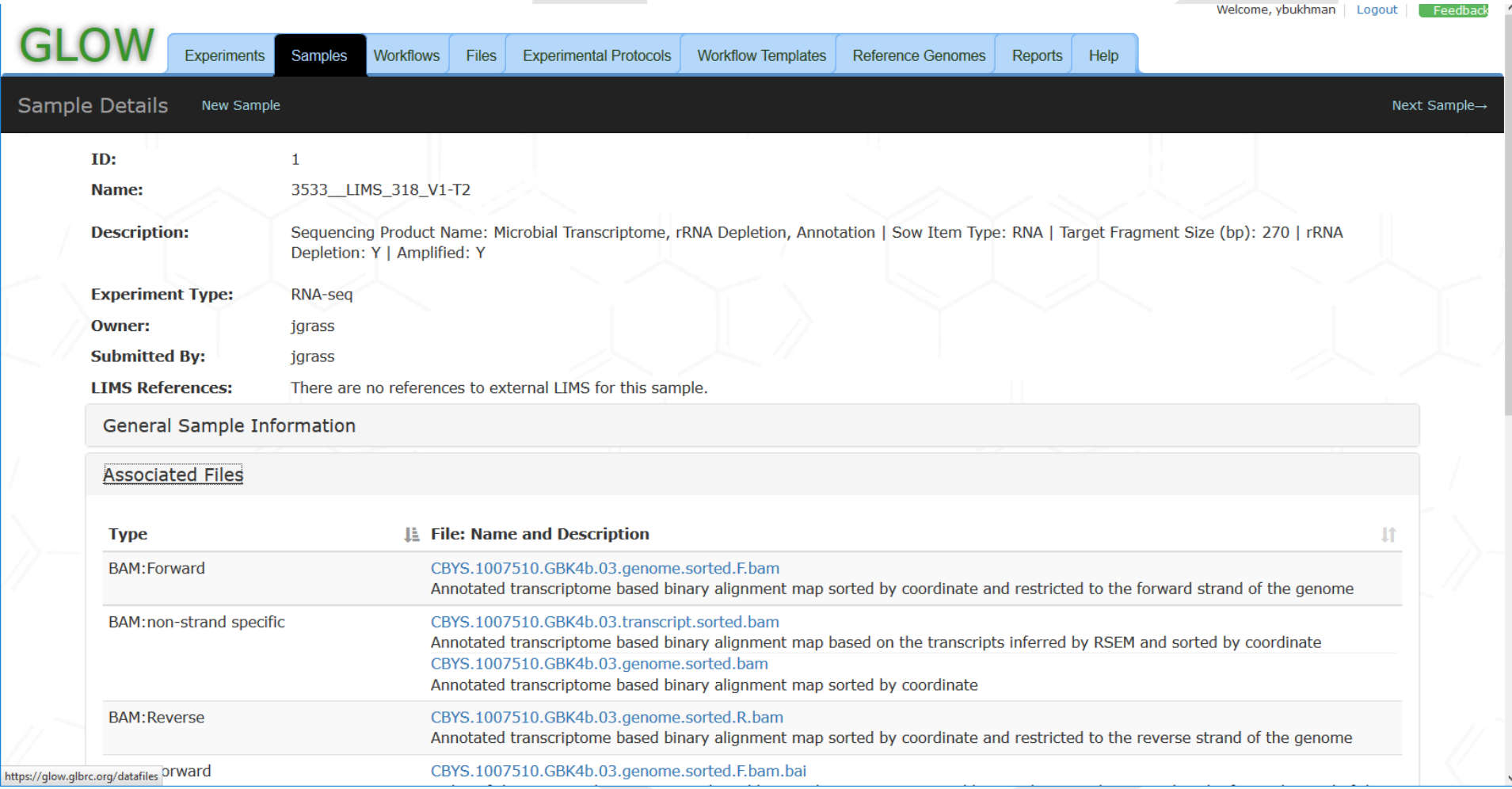
Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

Files are Associated With Samples



Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

Sample and Its Files



Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

Samples are Grouped Into Experiment Sets

GLOW

Experiments | Samples | Workflows | Files | Experimental Protocols | Workflow Templates | Reference Genomes | Reports | Help

Experiment Set Details Add To... ←Previous Experiment Set

ID: 1

Name: Escherichia coli K-12 MG1655 3.1.4 Multiomics #1

Description: Project 1007510 uploaded from JGI metadata

Owner: jgrass

Submitted By: jgrass

Samples in this Experiment Set:

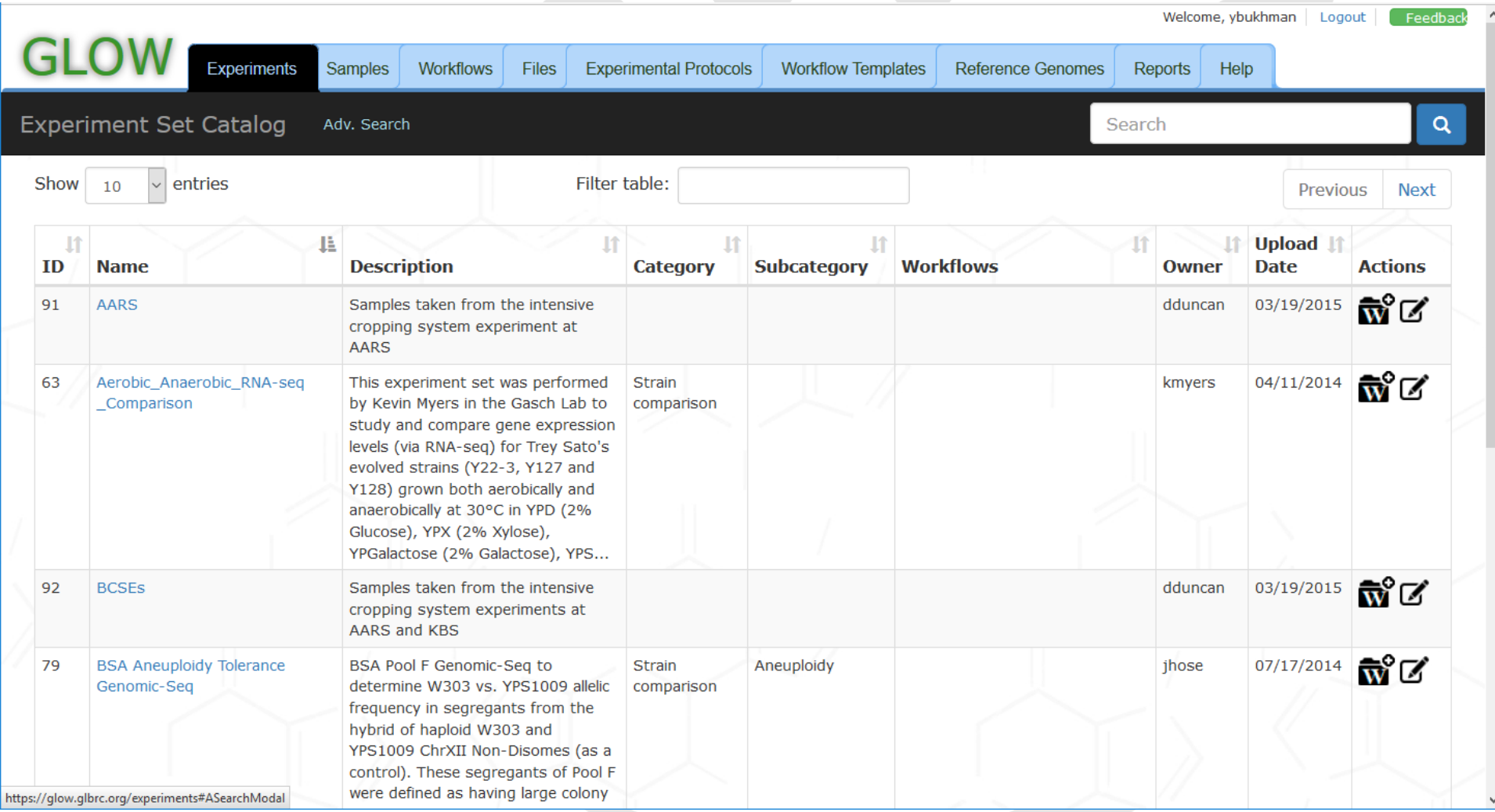
Show 10 entries Filter table: Previous Next

ID	Name	Description	Experiment Sets	Experiment Type	Category	Subcategory	Workflows	File Types	Owner	Upload Date	Action
1	3533__LIMS_318_V1-T2	Sequencing Product Name: Mi...	Escherichia coli K-12 MG1655 3.1.4 Multiomics #1;	RNA-seq			1007510.QC.RSEM.1007510.GBK4b.03;	BAM:Forward BAM:Reverse BAM:non-strand specific BAMI:Forward BAMI:Reverse BAMI:non-strand specific Sample-centric Cou...	jgrass	01/14/2014	Add to set • Create workflow •
	3534__LIMS_318_V1-T3	Sequencing	Escherichia coli	RNA-seq			1007510.QC.RSEM.1007510.G	BAM:Forward	jgrass	01/14/2014	Add to set

https://glow.glbrc.org/datafiles

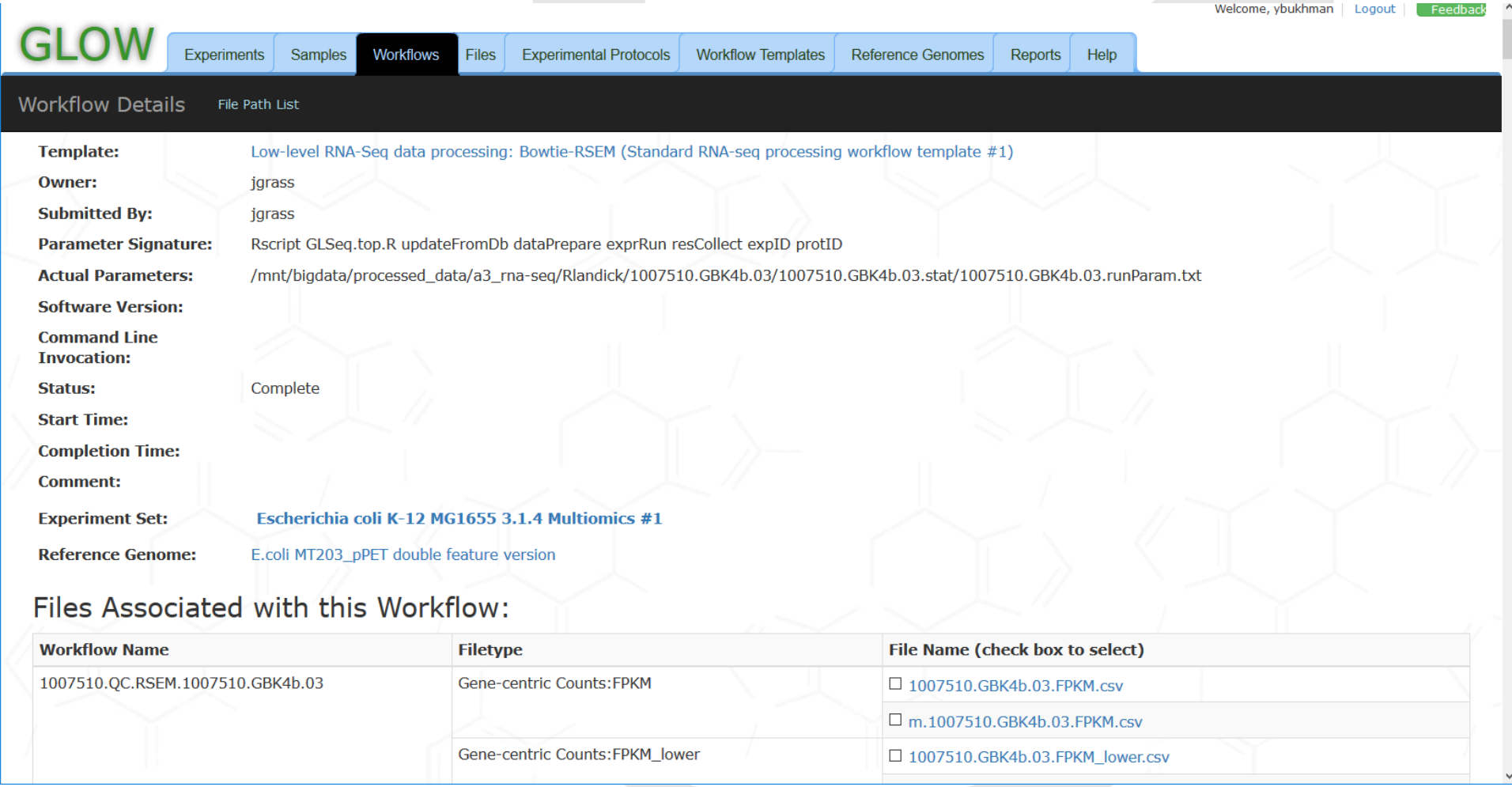
Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

GLOW Experiments



Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

Workflows are Run on Experiment Sets, Consume and Produce Files



Darin Kalisak, Nathan DiPiazza, Adam Halstead, et al.

GxSeq Gene Expression Viewer

GLBRC GxSeq

ybukhman | My Account | Log Out

Help | Tools | Samples | Features | Sequence

Expression Viewer - Matrix for: Brachypodium distachyon (Bd21)

Update Selection | Download

Filter Favorites: None

Search definition and locus: Go

(Advanced Options)

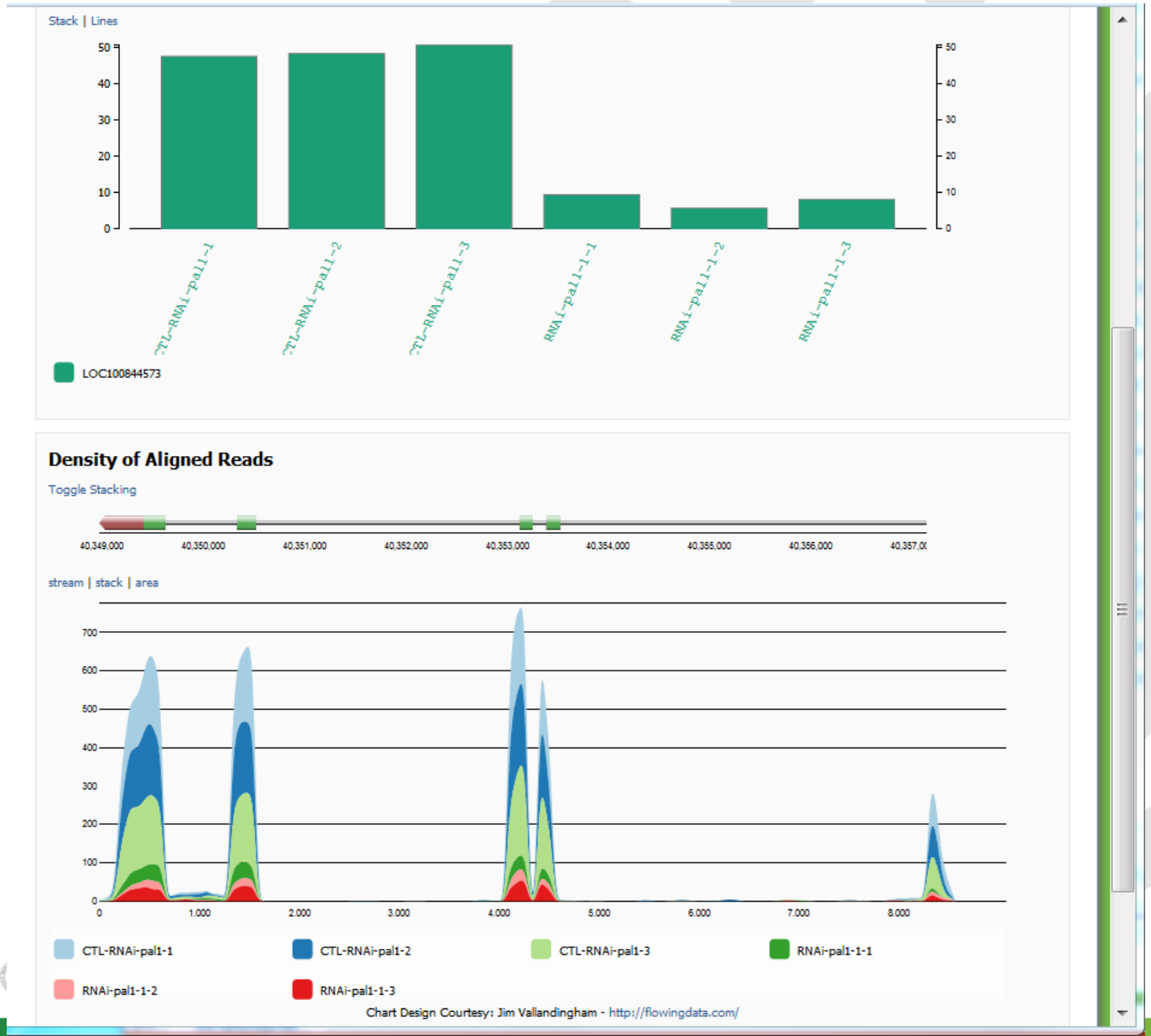
← Previous 1 2 3 4 5 6 7 8 9 ... 521 522 Next →

26,091 Matching Results

	Bd_192	tair10	Sbicolor 1.4	CTL-RNAi-pal1-1	CTL-RNAi-pal1-2	RNAi-pal1-1-1	RNAi-pal1-1-2	Sum	Options
| GeneID:100839285 | |				9435.72	6826.8	6553.75	11019.86	33836.13	Details | Graph | Browser
onein-like protein 2C-like	Bradi2g38120.1		Sb05g022840	6039.65	3105.31	2871.86	9070.81	21087.63	Details | Graph | Browser
se small chainPW9,	Bradi3g26391.1	AT1G67090.1	Sb05g003480	4264.48	4556.47	4398.18	3662.85	16881.98	Details | Graph | Browser
se small chainPWS4.3,	Bradi1g65564.1	AT1G67090.1	Sb02g001725	4172.7	4684.72	3873.82	3445.35	16176.59	Details | Graph | Browser
se small chainPWS4.3,	Bradi4g08500.1	AT1G67090.1	Sb02g042765	4193.09	4534.38	3965.85	3044.95	15738.27	Details | Graph | Browser
s; PF03953.7.fs;	Bradi1g10150.1	AT4G14960.2	Sb01g009560	2935.65	3541.07	3257.52	3087.65	12821.89	Details | Graph | Browser
osylmethionine synthase	Bradi2g12160.1	AT4G01850.2	Sb03g013170	2243.15	2260.26	2906.21	2533.62	9943.24	Details | Graph | Browser
842305 |	Bradi1g12787.1	AT4G39260.1	Sb01g012300	2431.77	2192.41	1965.51	2842.94	9432.63	Details | Graph | Browser
; PF00891.8.ls;	Bradi3g16530.1	AT5G54160.1	Sb07g003860	2129.64	1782.27	1959.39	3405.29	9276.59	Details | Graph | Browser
ethyltransferase-like |	Bradi1g13290.1	AT5G17920.2	Sb01g012960	1835.59	2214.82	2885.17	2284.07	9219.65	Details | Graph | Browser
	Bradi4g19457.1	AT4G13940.1	Sb05g014470	1818.18	1888.53	2270.99	2540.1	8517.8	Details | Graph | Browser
02496.6.ls | |	Bradi4g24650.1		Sb08g004190	2084.03	1813.89	1706.95	2717.46	8322.33	Details | Graph | Browser
439.8.ls | |	Bradi2g06220.1	AT3G15353.1		1911.09	2016.92	2088.74	1668.07	7684.82	Details | Graph | Browser
polyubiquitin-like	Bradi3g04737.1	AT5G03240.3	Sb04g004270	2065.79	1757.84	1875.13	1923.75	7622.51	Details | Graph | Browser
ethyltransferase-like |	Bradi4g01200.2	AT3G03780.3	Sb08g022210	1502.49	1755.13	2052.45	1845.14	7155.21	Details | Graph | Browser
LOC100846426 |	Bradi3g01470.1	AT4G35090.1	Sb04g001130	1434.33	1369.15	1790.25	2377.0	6970.73	Details | Graph | Browser
rm 1: chlorophyll a-b	Bradi2g16290.1	AT2G34420.1	Sb03g027030	2201.31	1962.69	1445.91	1066.33	6676.24	Details | Graph | Browser

Nick Thrower, Matt Larson, Curt Wilkerson

GxSeq Gene View



Nick Thrower,
Matt Larson,
Curt Wilkerson

Open Source

- GxSeq: https://github.com/Michigan-State-University/gxseq (version 1)
- GLOW: coming soon

Acknowledgements

David Benton, retiring IIT Director

Nihar Sheth, new IIT Director

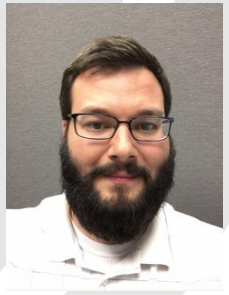
WEI IT: Dirk Norman
Kevin Leigeb, Branden Timm, …

LIMS: Ying Gao
Quansheng Yang, Feiran Yu, LiRong Tao,…

QA/QC: Enhai Xie

GLOW: Darin Kalisak,
Spencer Thiel, Nathan DiPiazza, Adam Halstead, …

Curt Wilkerson, MSU Informatics Group Lead



GxSeq: Nick Thrower,
Matt Larson

THANK YOU

Growing Core Data Management Systems

- Thoughtful design ➔ ability to extend systems and create customized solutions
  - Needs database and software design expertise
  - Needs scientific domain expertise
- Sound software development practices: documentation, version control, testing etc. ➔ maintainable and extensible systems, robust to staff turnover
- Agile: release early, release often ➔ get user input as frequently as possible
- Do well by early adopters ➔ gradually extend reach