KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

The DOE Systems Biology Knowledgebase

KBase Progress and Goals

Adam Arkin
Physical Biosciences Division
Lawrence Berkeley National Laboratory

BERKELEY LAB
Argonne NATIONAL LABORATORY
OAK RIDGE National Laboratory
BROOKHAVEN NATIONAL LABORATORY
CSH Cold Spring Harbor Laboratory

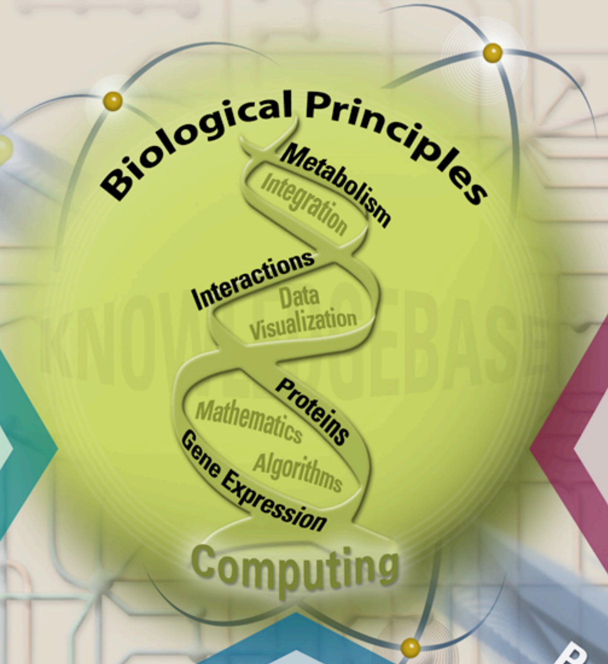
DOE Office of Science
Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY

DOE Systems Biology Knowledgebase

Components Enabling Its Development

Plants and Microbes for Energy and Environment



2010 Knowledgebase R&D Project

Year-long effort funded by American Recovery and Reinvestment ACT (ARRA).

Results from this project, which was completed in September 2010, are underpinning Knowledgebase development:

- **DOE Systems Biology Knowledgebase Implementation Plan**
- **ARRA pilot projects**

User Community Data and Resources

The Knowledgebase will leverage and establish critical partnerships with other synergistic efforts:

- **DOE Joint Genome Institute**
- **National Center for Biotechnology Information**
- **iPlant Collaborative**
- **DOE Office of Advanced Scientific Computing Research**
- **Others**

University-Led Projects to Develop Computational Biology and Bioinformatic Methods Enabling the Knowledgebase

In 2010, the DOE Office of Biological and Environmental Research awarded funding to 11 projects for basic research that will support Knowledgebase development in four areas:

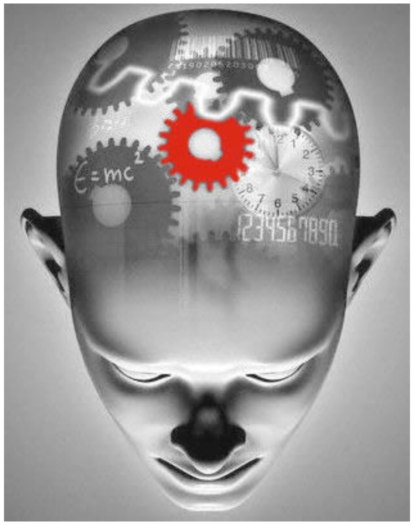
- **Omic data integration**
- **Integrated pathway reconstruction**
- **Genomic annotation**
- **Whole cellular simulations**

Predictive Understanding

Complex Missions, Sophisticated Mission Analysis

- **DOE mission**: *predict, control and design* the biological components of energetic processes and environmental balance.

- Complex missions with rapidly expanding, intricately related diverse data types require ways to augment scientists' ability to:
  - Filter information
  - Focus attention
  - Ask the right questions
  - Leverage other minds
  - Make the impossible possible. ☺



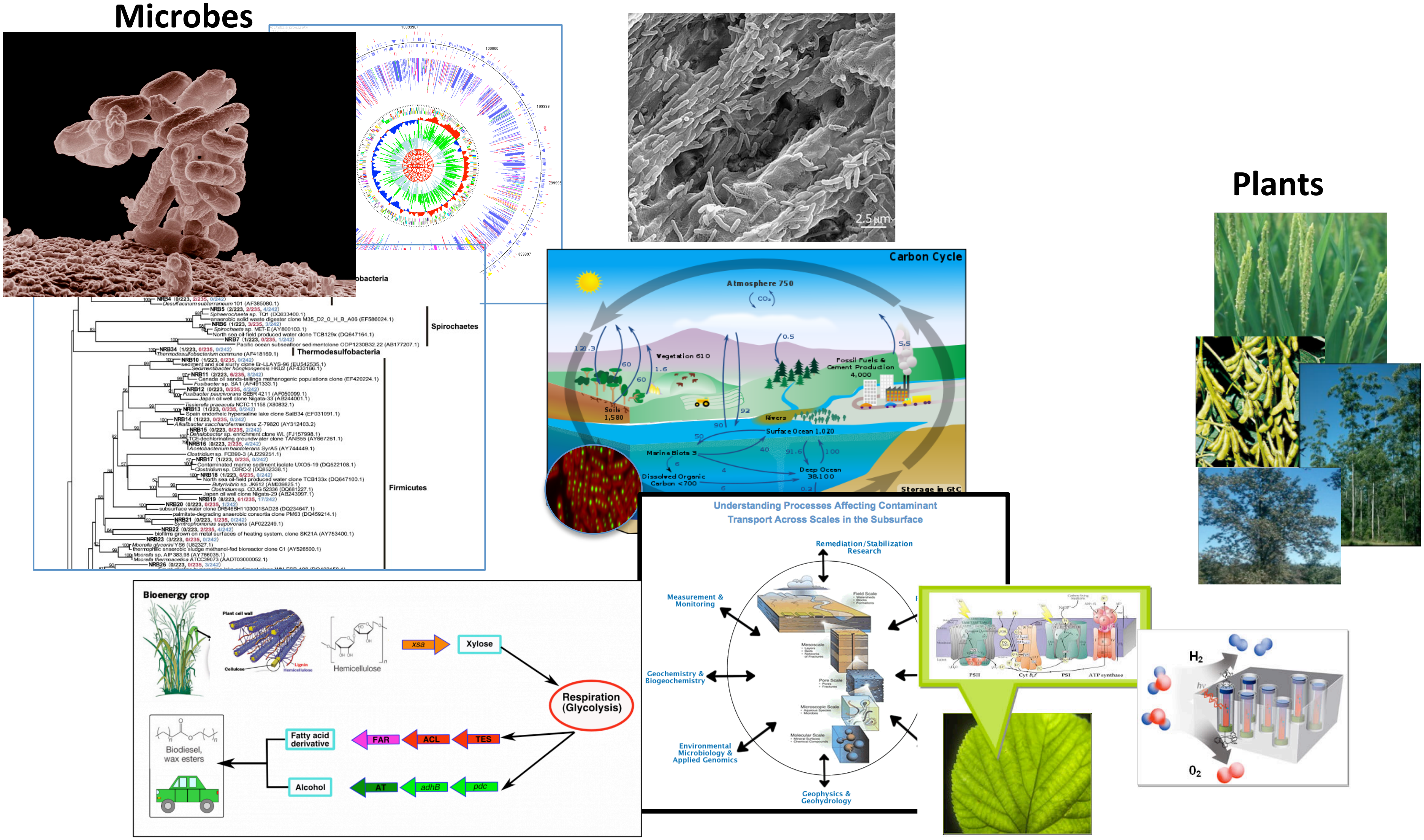
KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Enable DOE Mission Science

Microbes

Communities

Plants



DOE Office of Science
Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY

Before I tell you how…

DID WE MEET SPECIFICATIONS?

Launch and Build

Distributed Development

- Four major National Labs and more than 10 collaborating universities/research centers
- 50 people with diverse cultures and expertise and with different institutional and group alliances
- Enthusiasm to create a system to revolutionize understanding and application of biological systems.
- Deliver version 1 in 18 months and prove it can support DOE scientists!

Building an airplane in the sky

Community Specifications for KBase: 2011

- A Federated infrastructure and 10GBit/s transfer capabilities.
- 2 petabytes data storage and 2000 cores for data processing including interface and between high performance computing and cloud computational resources
- Development of core Knowledgebase integrated data and workflow analysis management tools including Application Programming Interfaces, semantic and user Interfaces.

i. For microbial systems, from 100-1000 microbes: Reconstruct and Predict Metabolic and Gene Expression Regulatory Networks to Manipulate Microbial Function
ii. For plant systems, for 10 key plants related to DOE missions: Integrate Phenotypic and Experimental Data and Metadata to Predict Biomass Properties from Genotype and Assemble Regulatory Data to Enable Analysis, Cross- Comparisons, and Modeling
iii. For Microbial Communities: Model Metabolic Processes within Microbial Communities and Mine Metagenomic Data to Identify Unknown Genes

KBase 2013: 18 Months Deliverables Exceed Spec.

- A Federated infrastructure and 10GBit/s transfer capabilities.
  - Built for high speed data transfer over ESNET using 100 GBit/s rates.
- 2 petabytes data storage and 2000 cores for data processing including interface and between high performance computing and cloud computational resources
  - KBase Magellan has 12,000 cores for data processing via both Open Stack Cloud interfaces and Cluster Services
  - KBase has >3 Petabytes of storage capacity
- Development of core Knowledgebase integrated data and workflow analysis management tools including Application Programming Interfaces, semantic and user Interfaces.
  - Integrated KBase API specified and operational. *Used by third parties to integrate function and build apps.*
  - Integrated data model aware of 925 data types encompassing sequence reads, contigs, genomes, genome features, transcription data, fitness data and more.
  - 40 Interface description documents leading to 821 functions that can be compiled into use for PERL, Python, Java, and R.
  - Prototype Search, Workflow and Novel Narrative/Notebook interfaces for navigating, analyzing and building knowledge in KBase

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

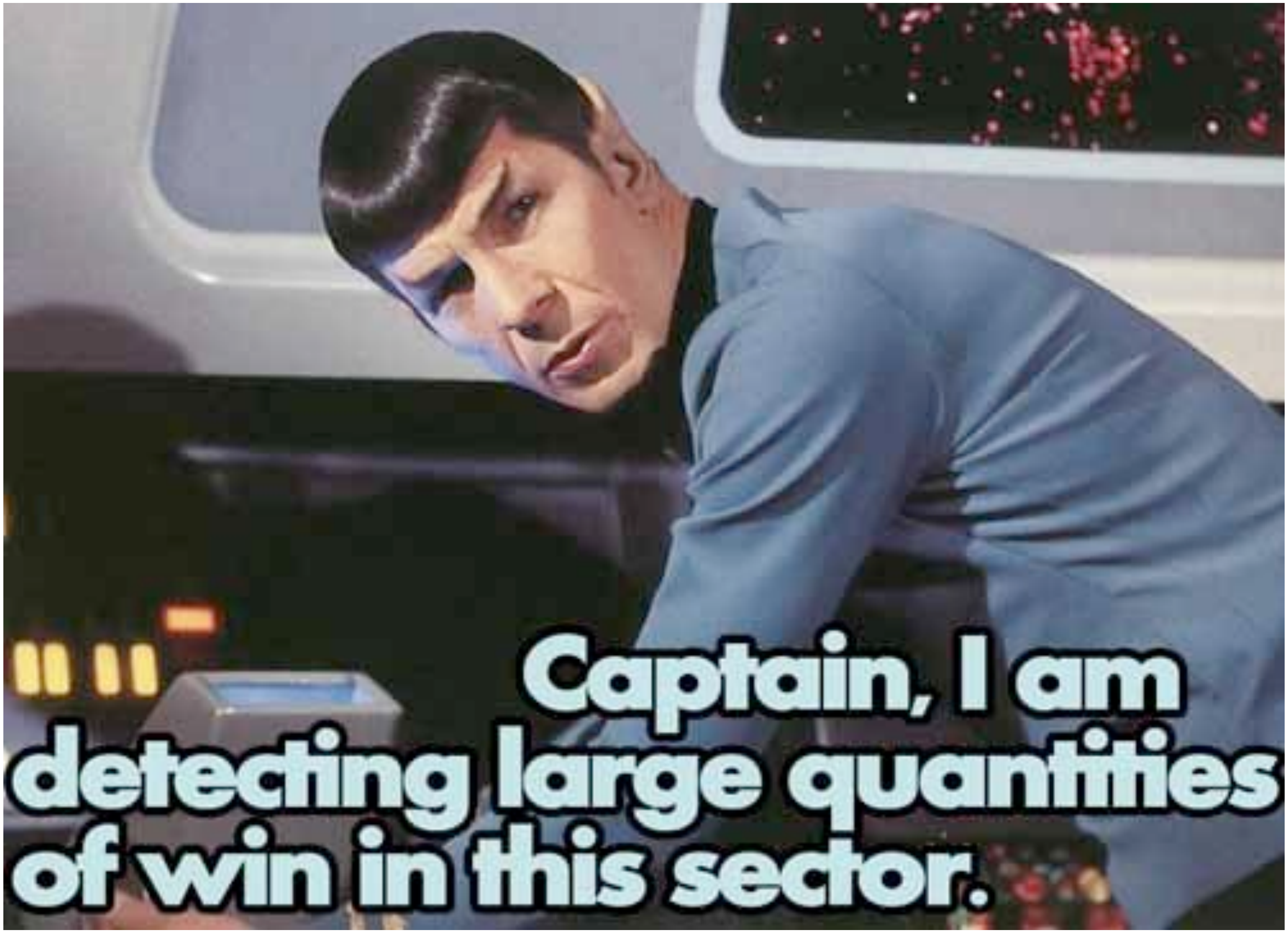
KBase 2013: 18 Months Deliverables Exceed Spec.

i. For microbial systems, from 100-1000 microbes: Reconstruct and Predict Metabolic and Gene Expression Regulatory Networks to Manipulate Microbial Function
    i. Metabolic and regulatory reconstructions for 5534 prokaryotic and 161 archaeal genomes
    ii. 7830 genome annotations, 23,058,670 features predicted
    iii. 12,620 regulons with 266,345 protein families inferred
    iv. 4985 metabolic models including a total of 16,196 compounds and 13,428 reactions
    v. 6202 growth curves, 1,947,690 strain fitness measurements; 3227 gene expression data sets
    vi. Services for assembly, annotation, phylogenomics, regulatory and metabolic networks inference, FBA and PROM modeling of metabolism, reconciliation and improvement of models against data

ii. For plant systems, for 10 key plants related to DOE missions: Integrate Phenotypic and Experimental Data and Metadata to Predict Biomass Properties from Genotype and Assemble Regulatory Data to Enable Analysis, Cross-Comparisons, and Modeling
    i. Over 175 eukaryotic genomes including many variants of Poplar, Arabidopsis, Sorghum, Chlamydomonas, Brachypodium, Miscanthus, Switchgrass as well as many other algae and fungi.
    ii. Phenotypes for genome variants of plants and services for calling the genetic variation among individuals.
    iii. Services for variation calling, mapping genotype-to-phenotype via GWAS style analysis and tools for candidate gene filtering, trait modeling, and pathway enrichment,
    iv. 731 gene expression experiments in Arabidopsis and Poplar; Plant co-expression network analysis for all.
    v. Initial plant metabolic modeling

iii. For Microbial Communities: Model Metabolic Processes within Microbial Communities and Mine Metagenomic Data to Identify Unknown Genes
    i. Access to 11,000 metagenomes(>21 TBp)
    ii. Integrated KBase access to QIIME functionality
    iii. New tools for metagenome sequence quality assessment and experimental design
    iv. Services for taxonomic and gene identification, abundance, and a host of other functions

U.S. DEPARTMENT OF
ENERGY


Captain, I am detecting large quantities of win in this sector.

The community document is not a specification, though…

WHY DO WE NEED KBASE?

Our goals drive our specification

- We believe it is necessary to solve the grand challenges in biological control of environmental and energy process.
- We believe it is necessary to deal with deep issues in scientific communication and reproducibility
- We need to lower the barriers to computationally efficient use of advanced algorithms and data from diverse producers.
- We believe we need to do this together-- leveraging many minds.

The number of bioinformatics resources is huge

- ☐ An increasing number of data warehouses since biology is becoming a big data discipline (e.g., NCBI, Ensembl, etc.)
- ☐ Specialized applications and databases for relatively generic analyses (e.g., MG-RAST and MicrobesOnline)
- ☐ Evolving libraries of sophisticated computational biology algorithms for use in programming environments (e.g., Bioconductor)
- ☐ Workflow tools that allow the chaining of these algorithms together by non-programmers (e.g., Galaxy and Taverna)
- ☐ Workflow sharing tools to allow people to use each other’s work products
- ☐ Open-access publication of journal articles with increasing use of semantic tagging
- ☐ Scientific social networks (e.g., ResearchGate, Epernicus, etc.)

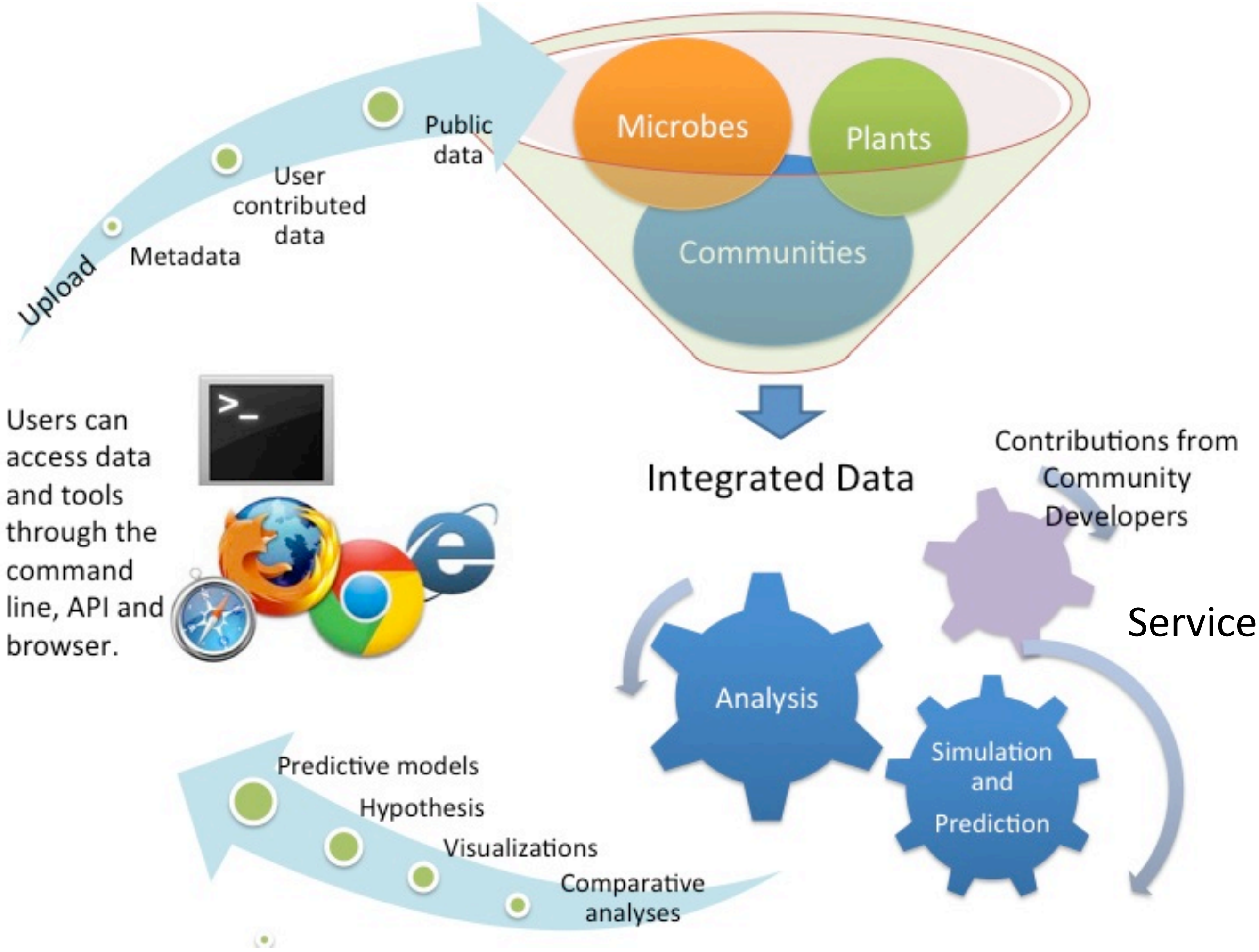
So what is missing?l

KBase Concepts

- KBase drives data through models to predictions and experimental design.
- KBase accelerates reproducible, reusable, and transparent science.
- KBase deeply enables scientists to work together to approach complex biological problems.
- KBase gives credit where it is due and privacy where it is needed.
- KBase is an open software and data environment to which others can contribute and with which others can build.

Empowering Users

User Access

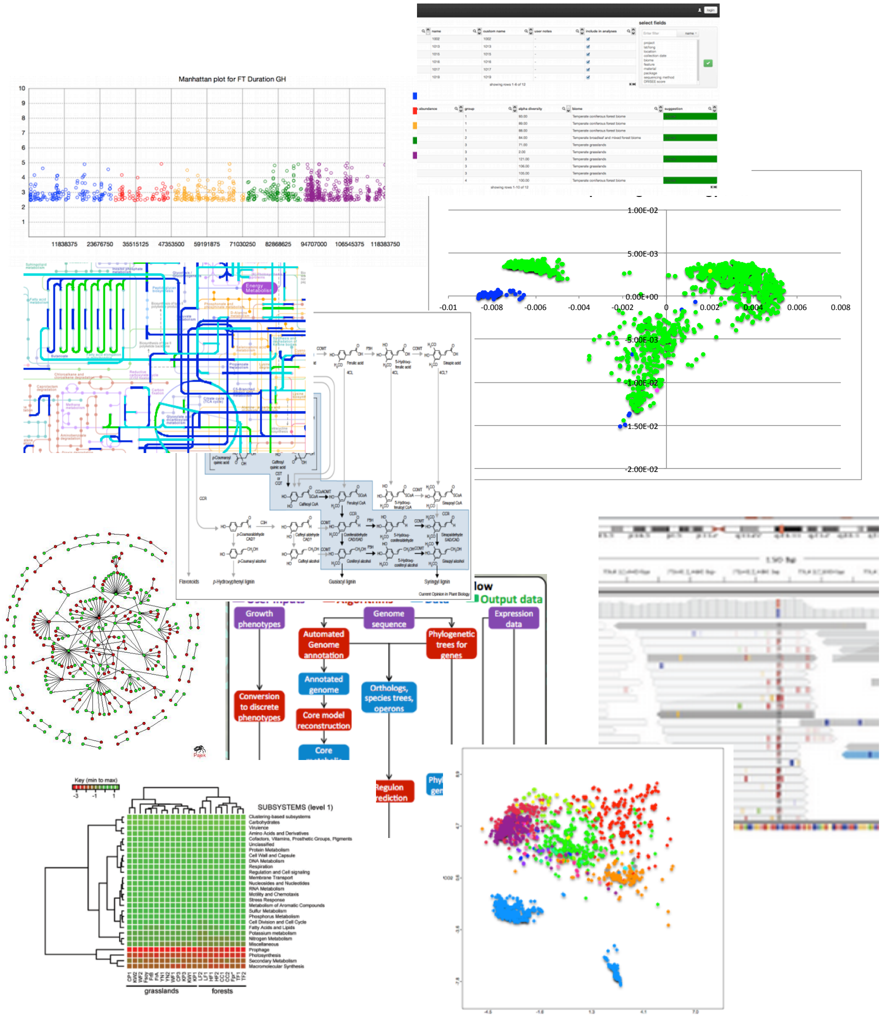


Users can access data and tools through the command line, API and browser.

Services and Tools

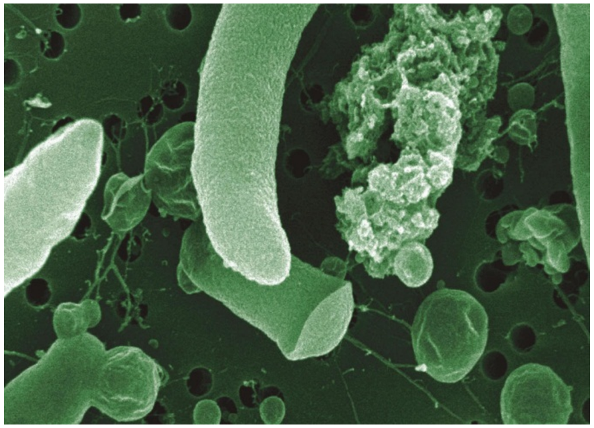
What can we do that was hard before?

- Transparently access multiple heterogeneous datasets and bioinformatics tools.
- Efficiently annotate new microbial genomes and infer metabolic and regulatory networks.
- Transform network inferences into metabolic models and map missing reactions to genes using novel data reconciliation tools.
- Design effective sequencing strategies for complex multi-sample metagenomic projects
- Test microbial ecological hypotheses through taxonomic and functional analysis of quality-assessed metagenomic data
- Predict plant gene function and molecular phenotype via navigation and analysis of tissue-specific co-expression networks.
- Discover genetic variations within plant populations and map these to complex organismal traits.
- Share data, analysis tools, workflows and scientific conclusions with the community.

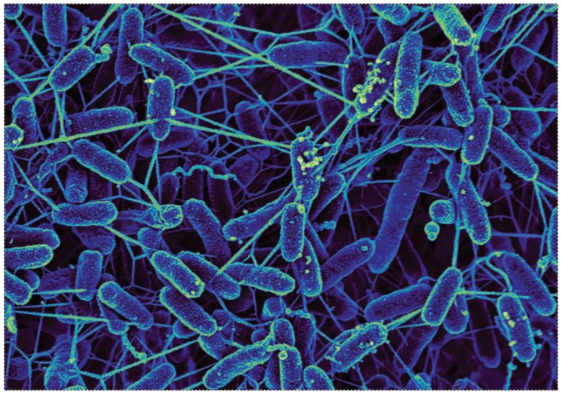


Microbes

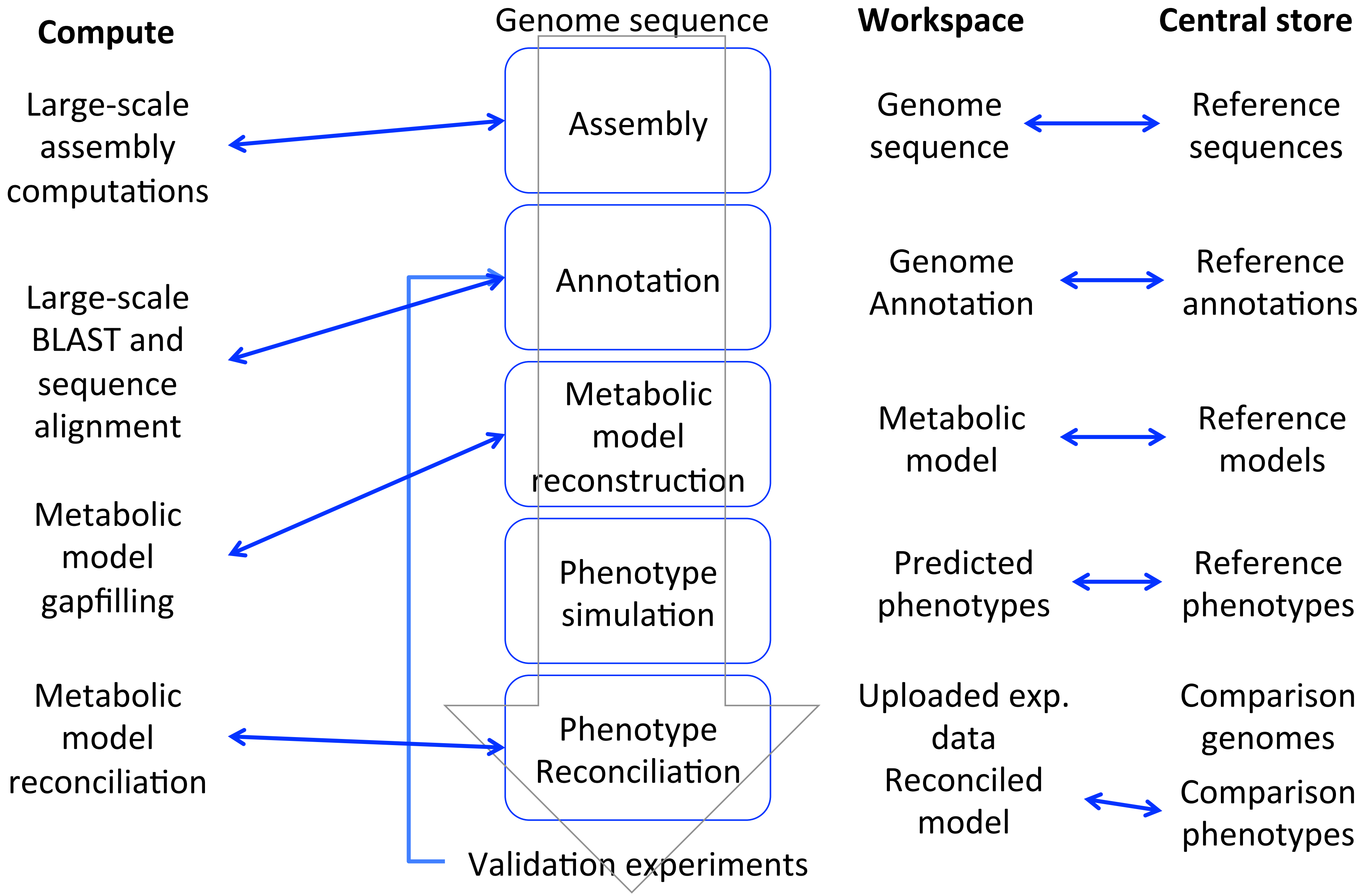
Our overall goals are to:



- Reconstruct and predict metabolic and gene expression regulatory networks to manipulate microbial function
- Vastly increase the capability of the scientific community to communicate and utilize their existing data
- Enable the planning of effective experiments and maximize our understanding of microbial system function



Rapidly building understanding of microbe behavior



Compute

Large-scale assembly computations

Large-scale BLAST and sequence alignment

Metabolic model gapfilling

Metabolic model reconciliation

Genome sequence

Assembly

Annotation

Metabolic model reconstruction

Phenotype simulation

Phenotype Reconciliation

Validation experiments

Workspace

Genome sequence

Genome Annotation

Metabolic model

Predicted phenotypes

Uploaded exp. data

Reconciled model

Central store

Reference sequences

Reference annotations

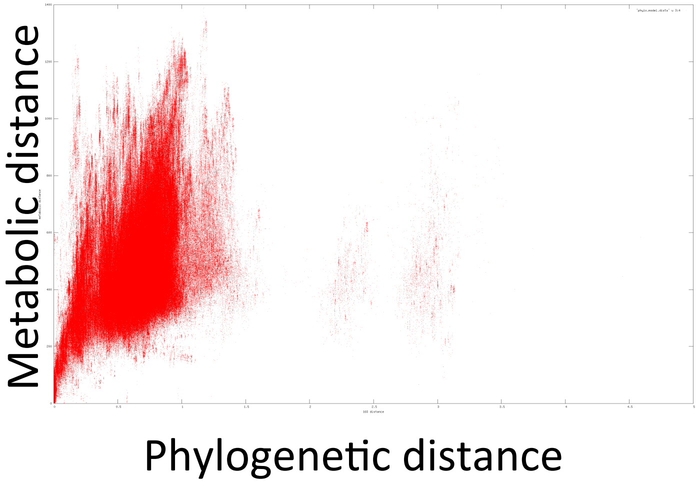
Reference models

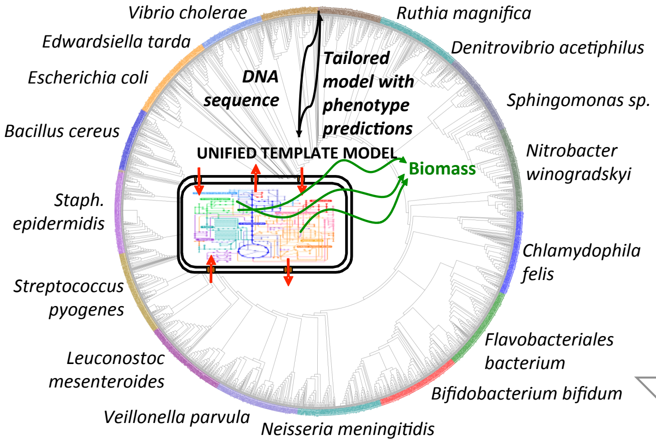
Reference phenotypes

Comparison genomes

Comparison phenotypes

Biological trends across the tree of life





Assessing “completeness” of our understanding

Genome sequence

- Large-scale query for genomes
- Extract annotations/ models
- Construction application of templates
- Statistical analysis of data

Workspace		**Central store**
Genome sequence	↔	Reference sequences
Genome Annotation	↔	Reference annotations
Metabolic model	↔	Reference models
Predicted phenotypes	↔	Reference phenotypes
Uploaded exp. data		Comparison genomes
Reconciled model	↔	Comparison phenotypes

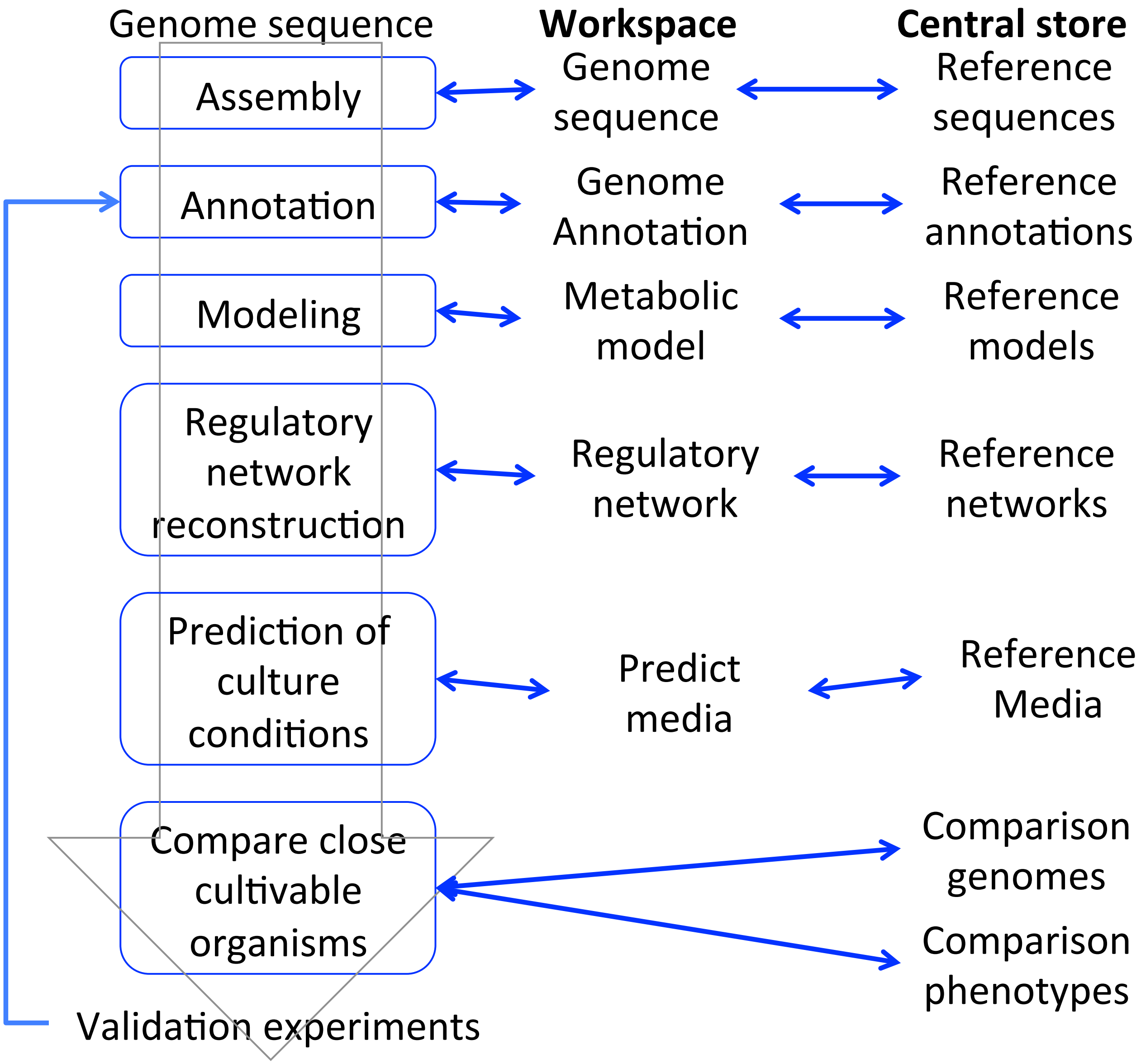
Driving discovery in experimental biology

Predict nutrients required for biomass biosynthesis

Predict potential nutrients that may inhibit growth

Compare models and predicted growth behavior of nearby cultivable microbes with genome of interest

Explore microbe-microbe codependency or microbe-host dependency hypotheses



Genome sequence

Workspace

Central store

Genome sequence	Workspace	Central store
Assembly	Genome sequence	Reference sequences
Annotation	Genome Annotation	Reference annotations
Modeling	Metabolic model	Reference models
Regulatory network reconstruction	Regulatory network	Reference networks
Prediction of culture conditions	Predict media	Reference Media
Compare close cultivable organisms		Comparison genomes Comparison phenotypes

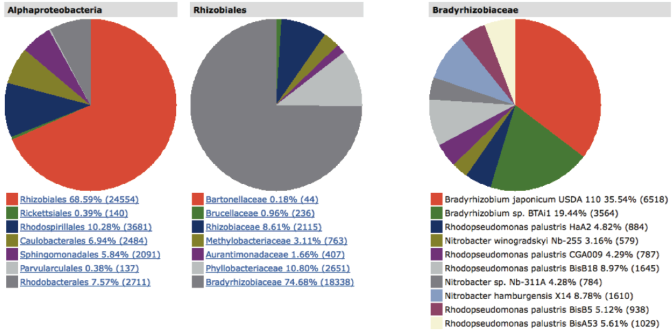
Validation experiments

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

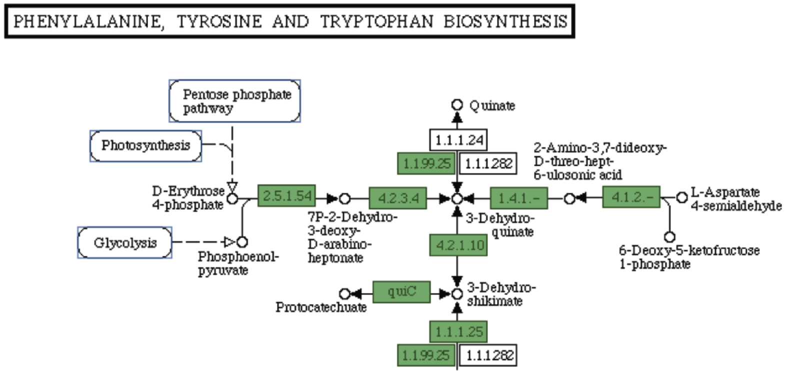
KBase Communities Supports Massive Metagenomics

- Automated processing and analysis of metagenomic data (16s/18s, shotgun metagenome, meta transcriptome)
- Incorporation of MG-RAST and QIIME functionality
- Novel sequence QC pipelines (DRISEE)
- Evidence-based design of metagenomic experiments

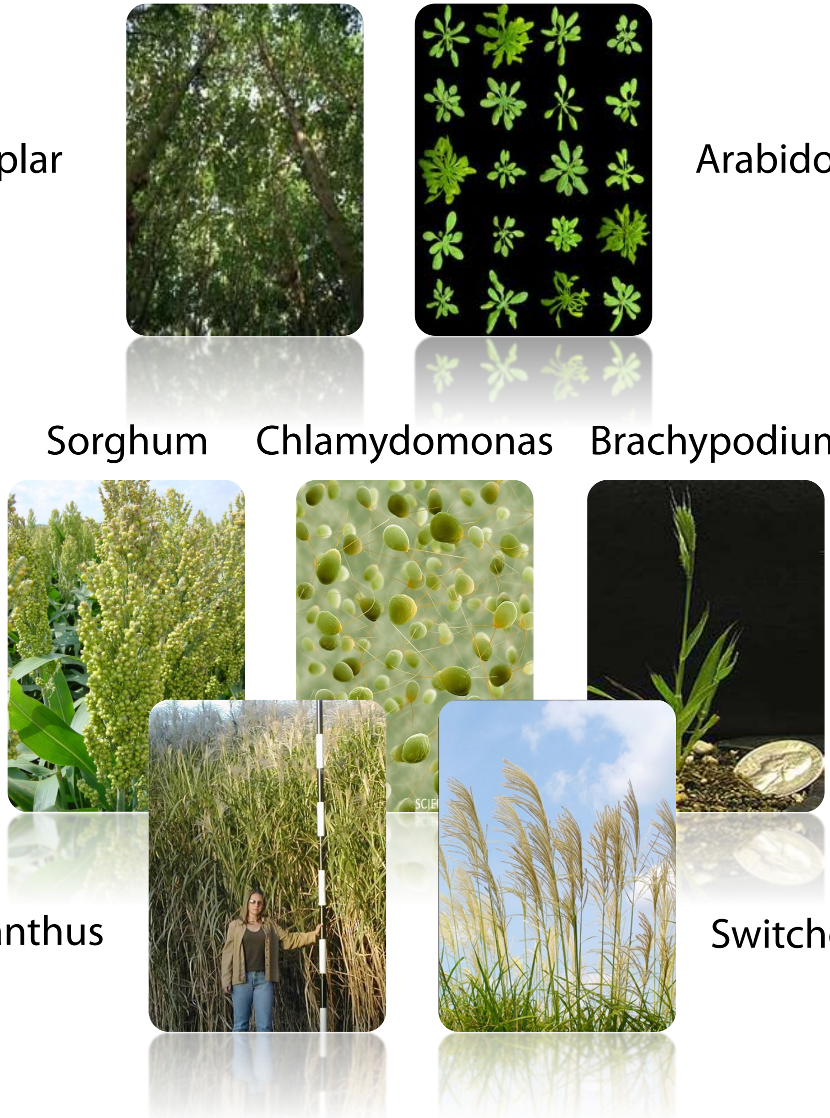
Who are they?



What are they doing?

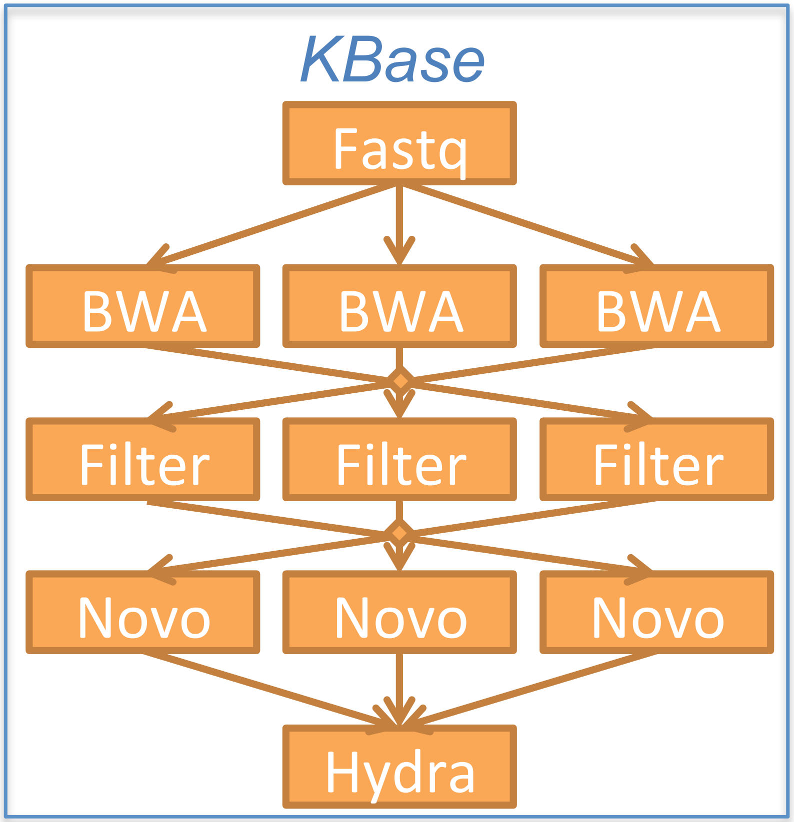


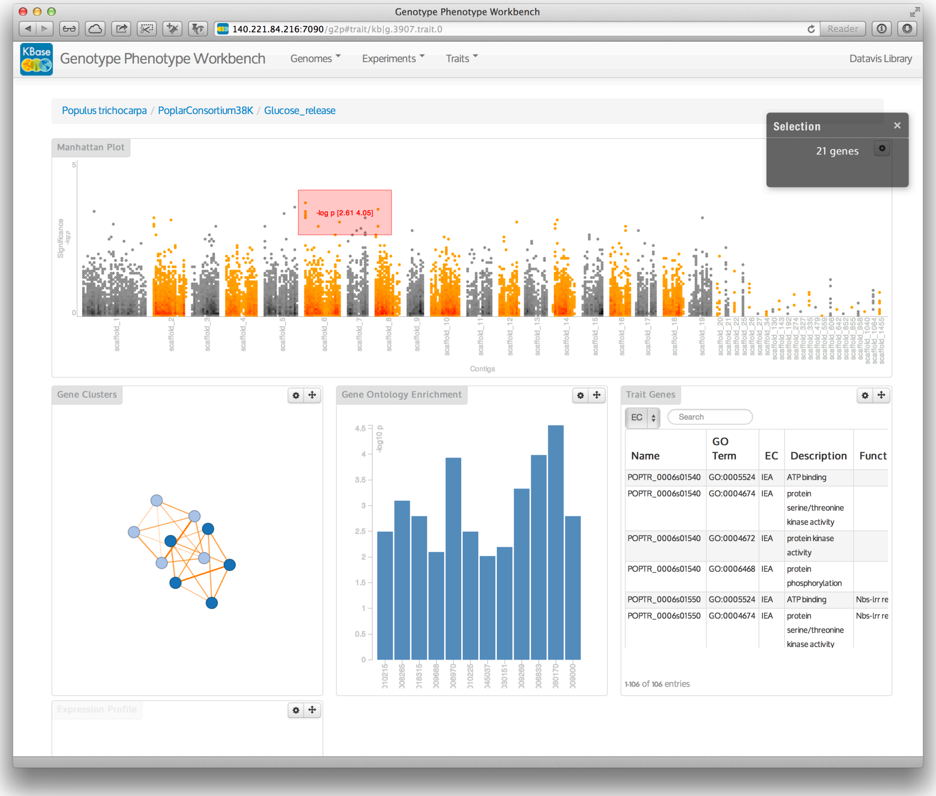
KBase-Plants: Targeted genomes



KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

From reads through variants to phenotypes





Name	GO Term	EC	Description	Funct
POPTR_0006s01540	GO:0005524	IEA	ATP binding	
POPTR_0006s01540	GO:0004674	IEA	protein serine/threonine kinase activity	
POPTR_0006s01540	GO:0004672	IEA	protein kinase activity	
POPTR_0006s01540	GO:0006468	IEA	protein phosphorylation	
POPTR_0006s01550	GO:0005524	IEA	ATP binding	Nbs-lrr re
POPTR_0006s01550	GO:0004674	IEA	protein serine/threonine kinase activity	Nbs-lrr re

1-106 of 106 entries

Maize Test Run

Align and call SNPs from 35M 80bp (14Gbp) reads with maize genome (zmb73v2)

Identified 372k high confidence SNPs

	Serial	Multitcore	KBase Cloud
Config	1 core (1 node)	44 core (1 node)	118 cores (15 nodes)
Bowtie2	45 h*	1h 10m	23 m
Sort	2 hr	2 hr	N/A
Samtools	2 hr	2 hr	12 m
End-to-End Speedup	50h* 1x	5h 10m 9.6x	35 m 86x

*estimated time

Maize Population Analysis

Align and call SNPs from 131 maize samples
1TB fastq / 408Gbp input data

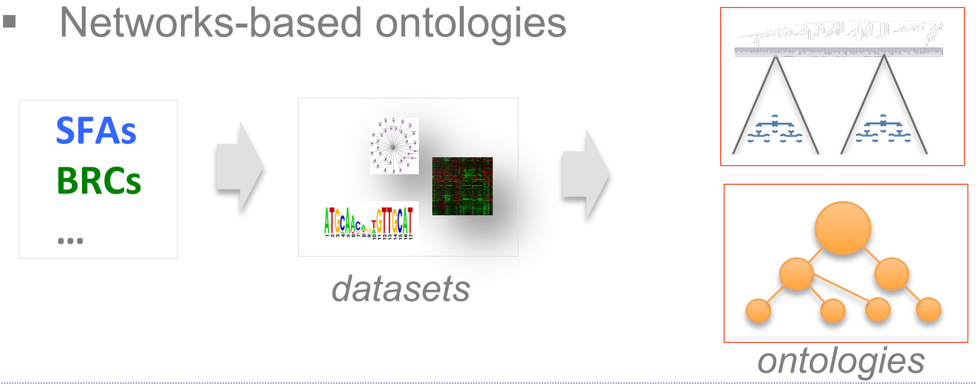
	Serial	KBase cloud (small)	KBase Cloud (large)
Config	1 core (1 node)	210 cores (15 nodes)	854 cores (61 nodes)
Bowtie2	1311 hr*	19.5 hr	5 hr
Sort	58 hr*	N/A	N/A
Samtools	58 hr*	3.5 hr	1.5 hr
End-to-End Speedup	1427 hr* 1x	23 hr 62x	6.5 hr 219x

*estimated time

Networks: Cross-cut example

I. Networks-based knowledge discovery

- Networks-based ontologies



SFAs
BRCs
…

datasets

ontologies

- Gene-ontologies matcher

Algorithms and UI tools to search and analyze best matched network components for a user specified gene set

II. Facilitate building predictive models

- Networks-based reliable orthologs

Plant genomes
co-expression based functional orthologs (Mutwil et al. Plant Cell 2011)

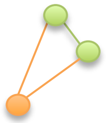
Microbial genomes
co-localization based reliable orthologs

III. Infrastructure for scientific social networks

- Networks of scientific communities and projects

Support narrative interface. Social networks connecting users and joint projects

- Networks of algorithms similarity



Allows KBase to diversify its menu of algorithms and avoid algorithms producing nearly identical results

- Networks quality control

Assign quality measures to KBase networks to allow healthy competition between algorithms and teams

Datasets and Networks API

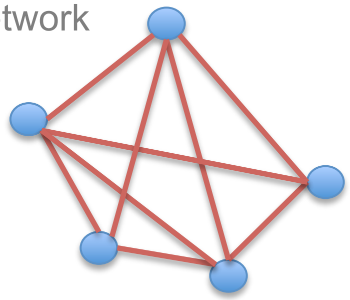
I. Datasets: collected, processed, classified

KBase domain	Dataset sources	Datasets
Microbes	7	5099
Plants	6	46
Communities	1	1

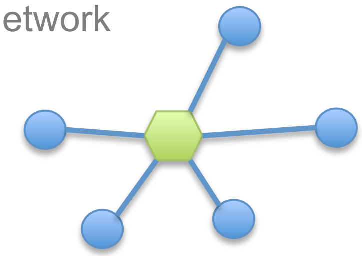
Network types	
REGULATORY_NETWORK	CO_FITNESS
PROT_PROT_INTERACTION	CO_EXPRESSION
METABOLIC_SUBSYSTEM	FUNCTIONAL_ABUNDANCE
FUNCTIONAL_ASSOCIATION	PHYLOTYPE_ABUNDANCE

II. Networks API: provide heterogeneous networks in unified format

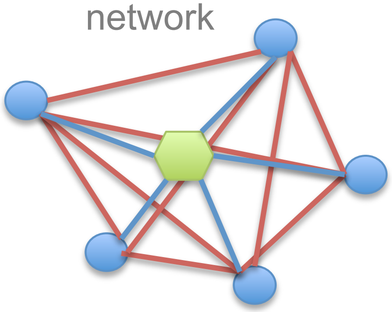
GENE-GENE network



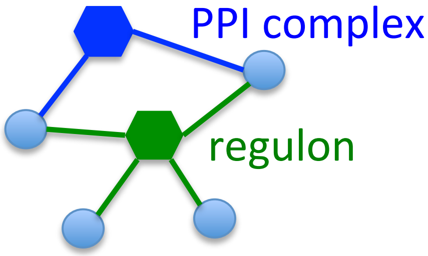
GENE-CLUSTER network



MIXED network



INTEGRATED network



Networks build methods

```
Network buildFirstNeighborNetwork(
        list<string> datasetIds,
        list<string> entityIds,
        list<EdgeType> edgeTypes)
```

```
Network buildInternalNetwork(
        list<string> datasetIds,
        list<string> geneIds,
        list<EdgeType> edgeTypes)
```

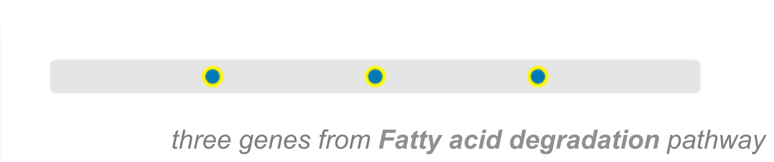
Datasets management methods

```
list<Dataset>  allDatasets()
list<DatasetSource> allDatasetSources()
list<NetworkType>  allNetworkTypes()
…
```

Networks Workbench

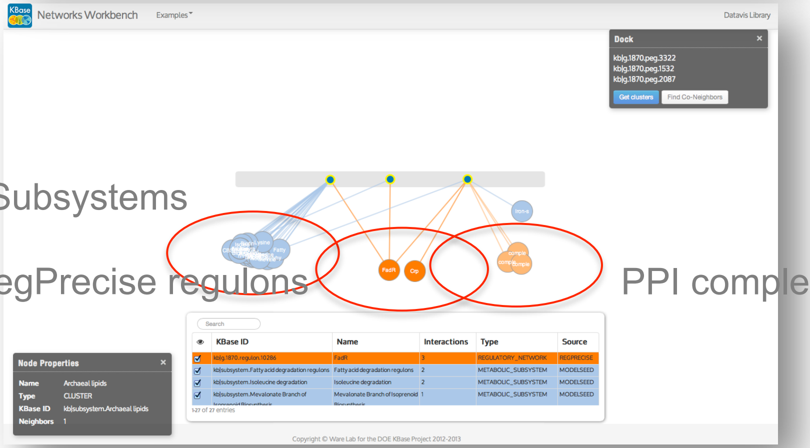
Iterative exploration and building network components

powered by the developed ***Networks API***



three genes from ***Fatty acid degradation*** *pathway*

Gene clusters associated with genes from dock panel



Dock

kb|g.1870.peg.3322
kb|g.1870.peg.1532
kb|g.1870.peg.2087

SEED Subsystems

RegPrecise regulons

PPI complexes

	KBase ID	Name	Interactions	Type	Source
☑	kb|g.1870.regulon.10286	FadR	3	REGULATORY_NETWORK	REGPRECISE
☑	kb|subsystem.Fatty acid degradation regulons	Fatty acid degradation regulons	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Isoleucine degradation	Isoleucine degradation	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Mevalonate Branch of Isoprenoid Biosynthesis	Mevalonate Branch of Isoprenoid Biosynthesis	1	METABOLIC_SUBSYSTEM	MODELSEED

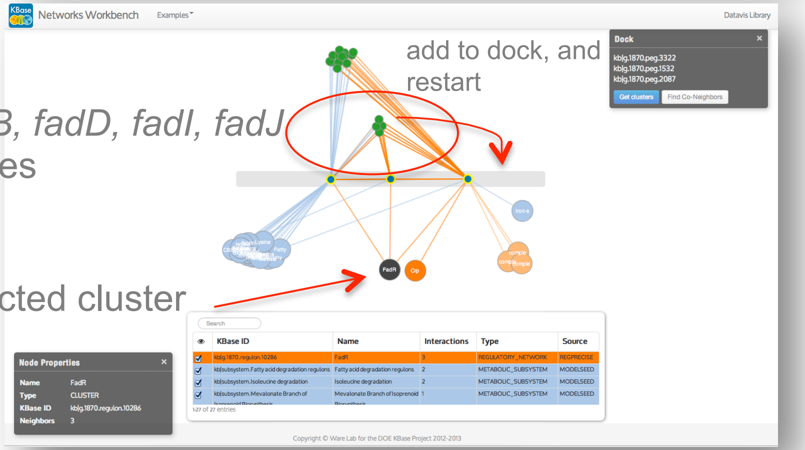
1-27 of 27 entries

Node Properties

Name Archaeal lipids
Type CLUSTER
KBase ID kb|subsystem.Archaeal lipids
Neighbors 1

Copyright © Ware Lab for the DOE KBase Project 2012-2013

Putative members of the same network component (docked genes)



add to dock, and restart

fadB, fadD, fadI, fadJ genes

selected cluster

Dock

kb|g.1870.peg.3322
kb|g.1870.peg.1532
kb|g.1870.peg.2087

	KBase ID	Name	Interactions	Type	Source
☑	kb|g.1870.regulon.10286	FadR	3	REGULATORY_NETWORK	REGPRECISE
☑	kb|subsystem.Fatty acid degradation regulons	Fatty acid degradation regulons	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Isoleucine degradation	Isoleucine degradation	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Mevalonate Branch of Isoprenoid Biosynthesis	Mevalonate Branch of Isoprenoid Biosynthesis	1	METABOLIC_SUBSYSTEM	MODELSEED

1-27 of 27 entries

Node Properties

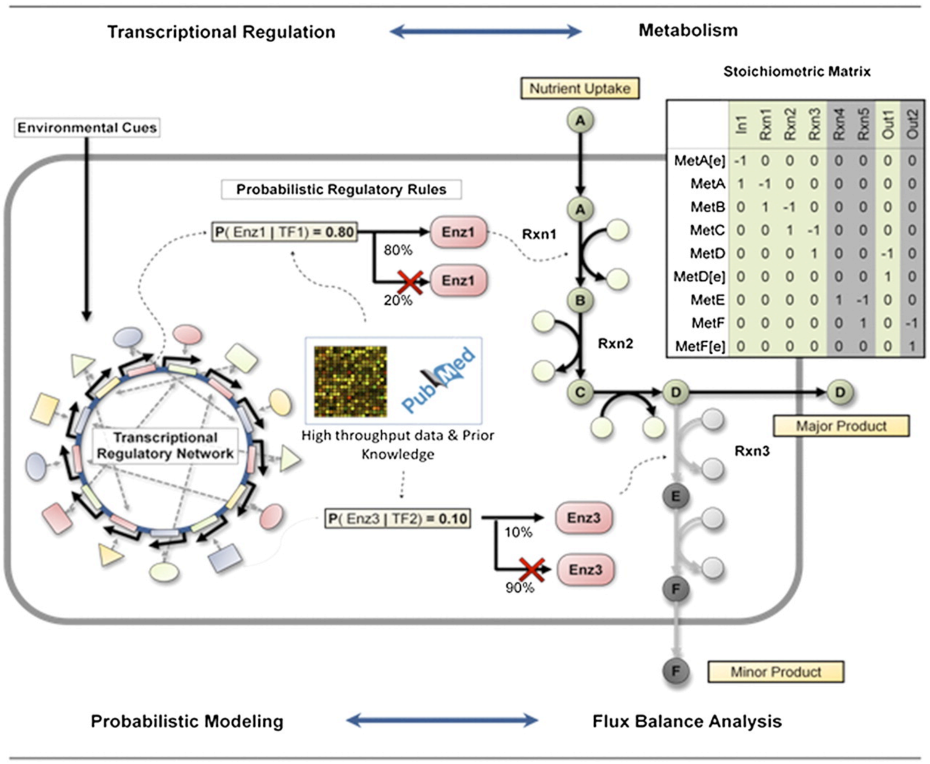
Name FadR
Type CLUSTER
KBase ID kb|g.1870.regulon.10286
Neighbors 3

Copyright © Ware Lab for the DOE KBase Project 2012-2013

Working with collaborators early

ENSURING IMPACT

Using KBase framework to enable PROM

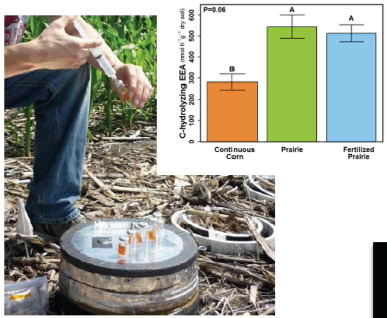


Stoichiometric Matrix

	In1	Rxn1	Rxn2	Rxn3	Rxn4	Rxn5	Out1	Out2
MetA[e]	-1	0	0	0	0	0	0	0
MetA	1	-1	0	0	0	0	0	0
MetB	0	1	-1	0	0	0	0	0
MetC	0	0	1	-1	0	0	0	0
MetD	0	0	0	1	0	0	-1	0
MetD[e]	0	0	0	0	0	0	1	0
MetE	0	0	0	0	1	-1	0	0
MetF	0	0	0	0	0	1	0	-1
MetF[e]	0	0	0	0	0	0	0	1

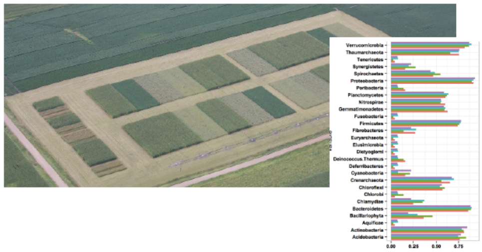
- Nathan Price's probabilistic regulation of metabolism (PROM) integrates metabolic models and functional data to make better predictions of growth given genetic or transcription variation.
- Difficult to access and for people to use.
- Previously applied to just two organisms.
- Now a KBase service that can be applied to any genome for which there is expression or variation data.
- Testing on *Shewanella oneidensis* MR-1 with transcription, TF knockout and growth/fitness data available in KBase.

KBase Microbial Communities Analysis Tools to Discover Niche Selection in Crop Soil Aggregates

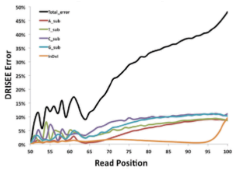


Biochemical characterization of soil aggregates in various crop treatments

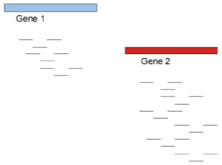
Deep sequencing of soil aggregates in various crop treatments



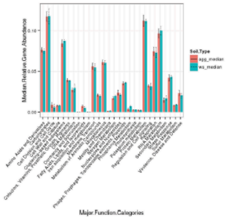
Quality Evaluation
Metadata Integration



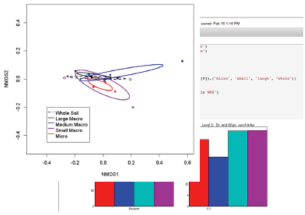
Assembly of reads



Predict coding regions
Annotation against
known databases



Integrative analysis comparing metadata, replicates, treatments, and aggregates (MatR, KBase Notebook)



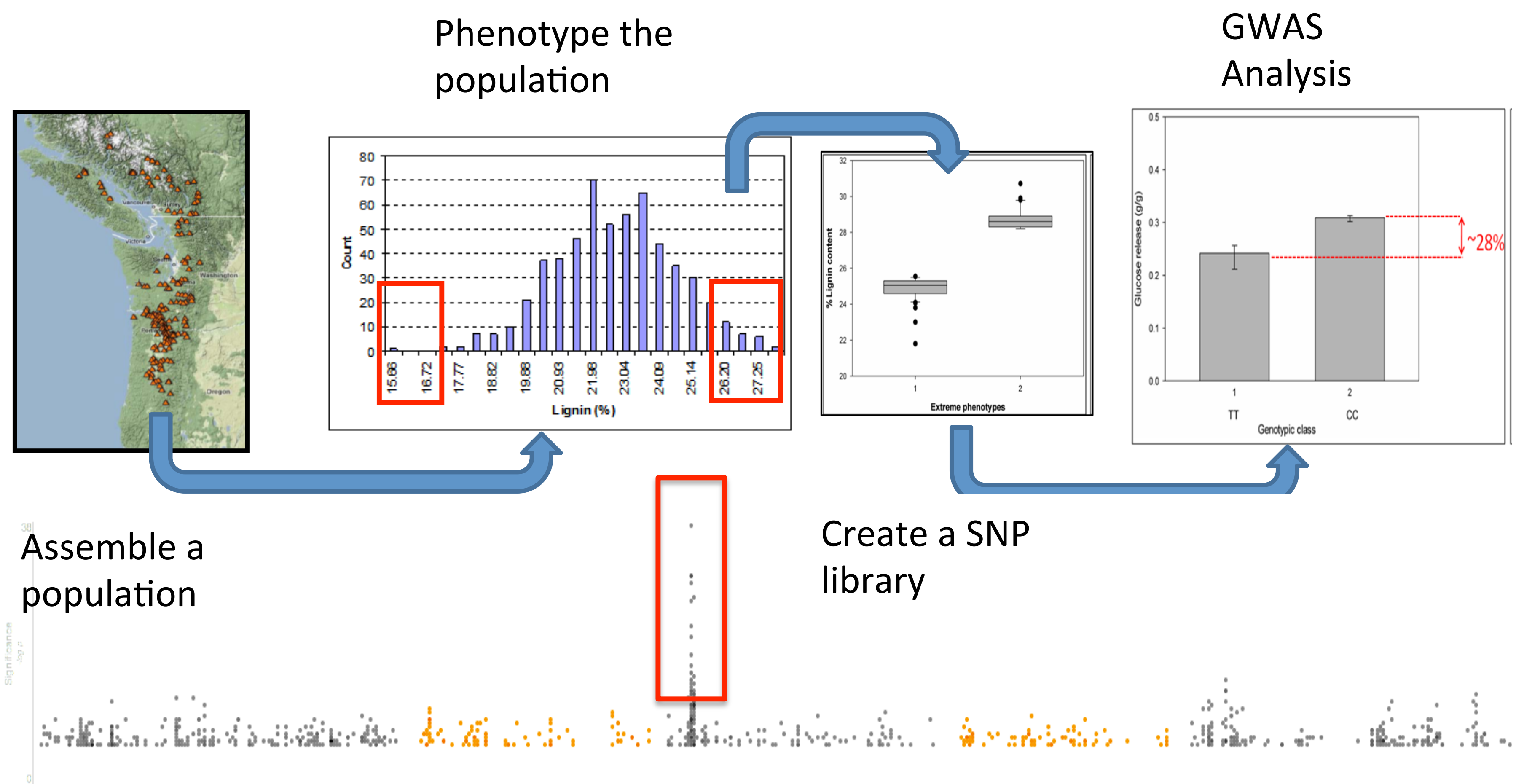
Kirsten Hofmockel, Iowa State, used KBase services to analyze and compare metagenomes from different sized soil aggregates across different crop treatments.

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Work with Tuskan (ORNL) & DiFazio (WVU) Identifies important SNPs

Variation in Lignin Composition and Content

Extreme variation contained in native populations of *Populus* detected in common garden experiments linked to genes using Association Genetics


Assemble a population

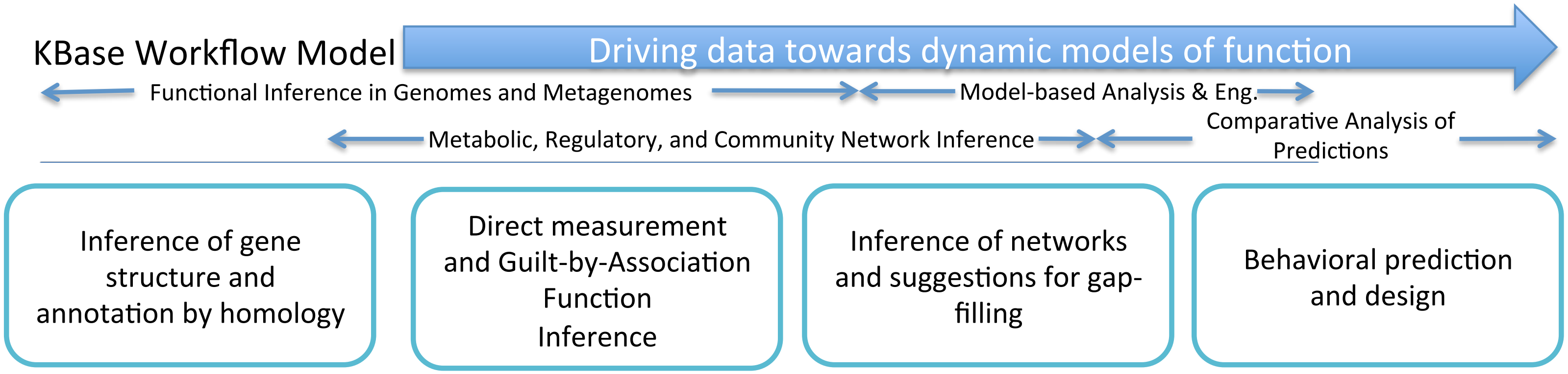
Phenotype the population

Create a SNP library

GWAS Analysis

HOW ARE WE BUILDING KBASE?

Define over research "paths"



KBase Workflow Model

Driving data towards dynamic models of function

Functional Inference in Genomes and Metagenomes

Model-based Analysis & Eng.

Metabolic, Regulatory, and Community Network Inference

Comparative Analysis of Predictions

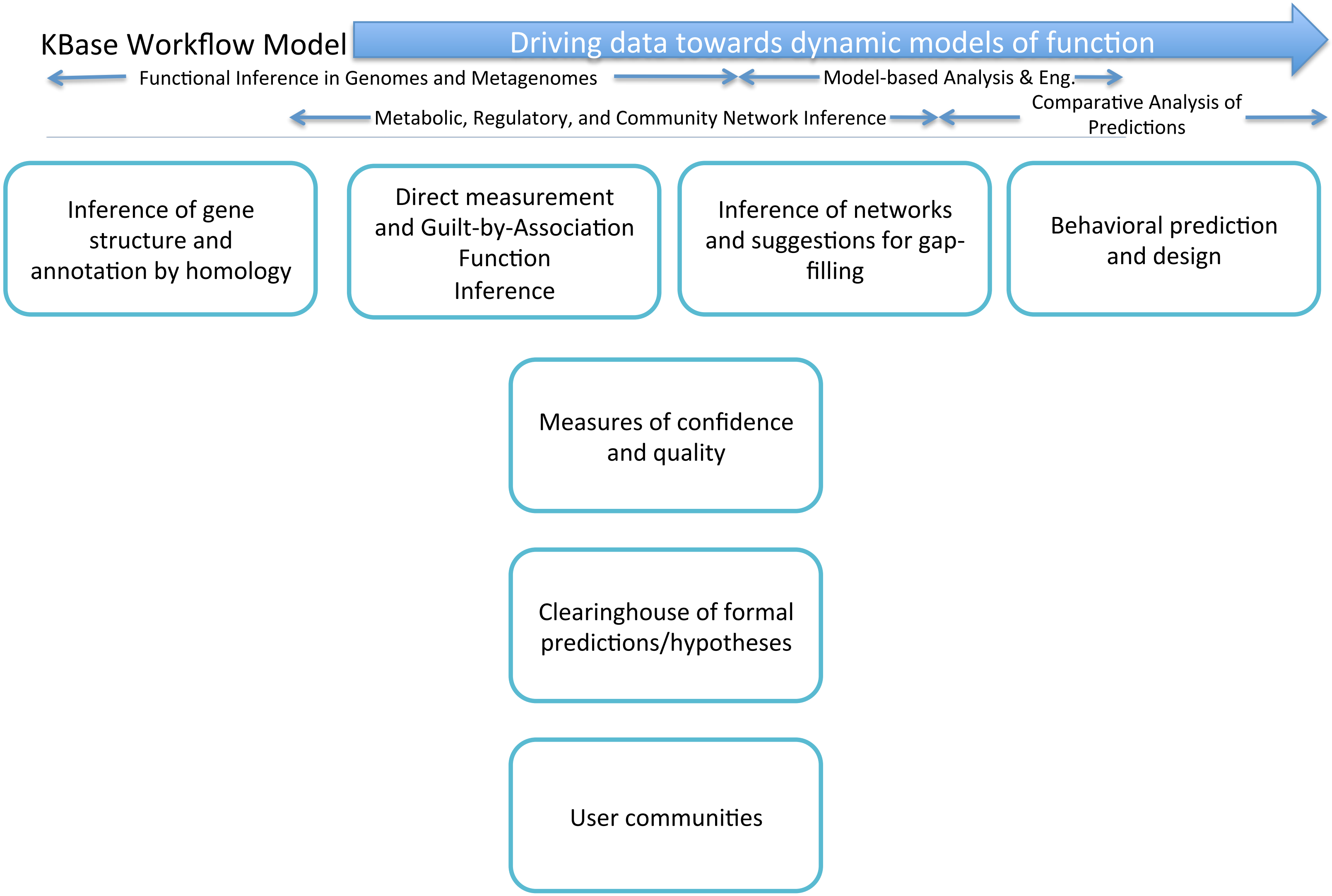
Inference of gene structure and annotation by homology

Direct measurement and Guilt-by-Association Function Inference

Inference of networks and suggestions for gap-filling

Behavioral prediction and design

Link design to KBase Concepts



KBase Workflow Model

Driving data towards dynamic models of function

← Functional Inference in Genomes and Metagenomes → ← Model-based Analysis & Eng. →

← Metabolic, Regulatory, and Community Network Inference → ← Comparative Analysis of Predictions →

Inference of gene structure and annotation by homology

Direct measurement and Guilt-by-Association Function Inference

Inference of networks and suggestions for gap-filling

Behavioral prediction and design

Measures of confidence and quality

Clearinghouse of formal predictions/hypotheses

User communities

Inference of gene structure and annotation by homology

Direct measurement and Guilt-by-Association Function Inference

Inference of networks and suggestions for gap-filling

Measures of Confidence and Quality

Clearinghouse of Predictions/Hypotheses

User Communities

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Three “Products” highlighted on Home Page

The Department of Energy Systems Biology Knowledgebase (KBase) is an emerging software and data environment designed to enable researchers to collaboratively generate, test and share new hypotheses about gene and protein functions, perform large-scale analyses on a scalable computing infrastructure, and model interactions in microbes, plants, and their communities. KBase provides an open, extensible framework for secure sharing of data, tools, and scientific conclusions in predictive and systems biology.

Try KBase Now
Use a web-based command-line interface—no installation necessary

Download the Tools
Install and run KBase command-line tools on your computer

Visit KBase Labs
Sneak a peek at KBase applications in development

KBase includes

- 5695 prokaryotic genomes
- 175 eukaryotic genomes
- 4985 models
- 12 services

What can KBase do?

- Efficiently annotate new microbial genomes and infer metabolic and regulatory networks.
- Transform network inferences into metabolic models and map missing reactions to genes using novel data reconciliation tools.
- Test microbial ecological hypotheses through taxonomic and functional analysis of quality-assessed metagenomic data
- Discover genetic variations within plant populations and map these to complex organismal traits.

Search the database:

Latest News

KBase Developer's Minor Release
Posted by nlharris Jan 18, 2013

KBase at International Plant and Animal Genome XXI
Posted by salazar Jan 09, 2013

KBase Team at Argonne for November Build
Posted by salazar Nov 30, 2012

View news

Upcoming Events

2013-02-18
BERAC Presentations

2013-02-22
Microbes Webinar

2013-02-24
DOE/NIFA Plant Feedstocks Genomics for Bioenergy

2013-02-24
Genomic Science Contractors-Grantees Meeting

View calendar

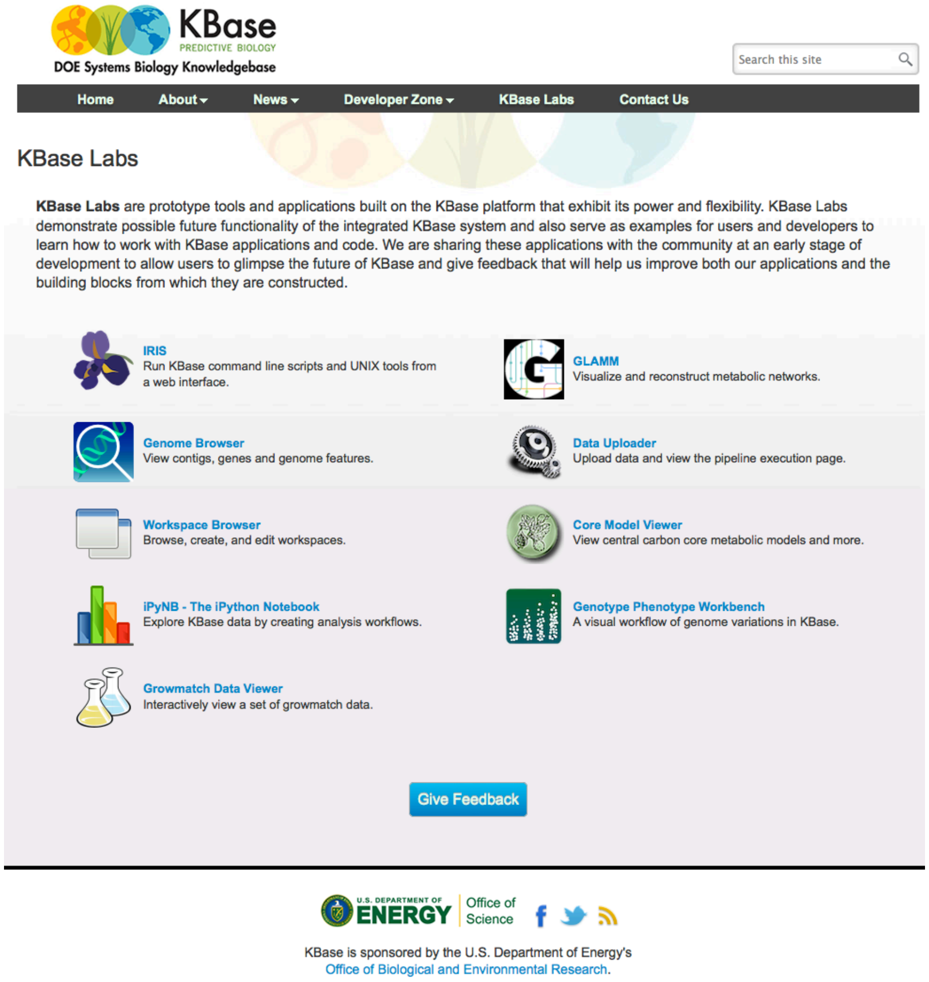
Sign up for a KBase account

KBase is sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research

Acknowledgements
Privacy and Security

KBase Labs

- Labs are prototypes and applications that demonstrate KBase present and possible future function.
- Early stage functionality both provides utility and a way for the community to provide feedback.



KBase Labs

KBase Labs are prototype tools and applications built on the KBase platform that exhibit its power and flexibility. KBase Labs demonstrate possible future functionality of the integrated KBase system and also serve as examples for users and developers to learn how to work with KBase applications and code. We are sharing these applications with the community at an early stage of development to allow users to glimpse the future of KBase and give feedback that will help us improve both our applications and the building blocks from which they are constructed.

IRIS
Run KBase command line scripts and UNIX tools from a web interface.

GLAMM
Visualize and reconstruct metabolic networks.

Genome Browser
View contigs, genes and genome features.

Data Uploader
Upload data and view the pipeline execution page.

Workspace Browser
Browse, create, and edit workspaces.

Core Model Viewer
View central carbon core metabolic models and more.

iPyNB - The iPython Notebook
Explore KBase data by creating analysis workflows.

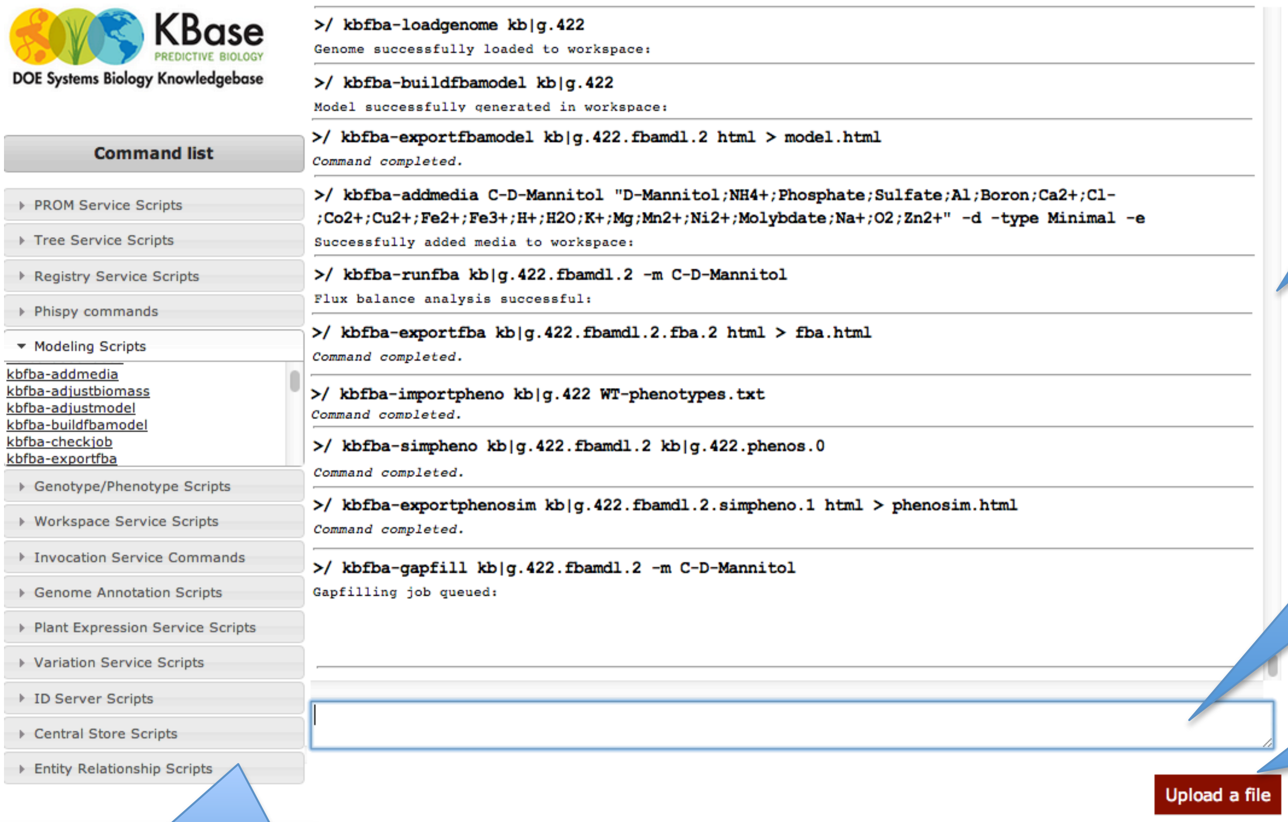
Genotype Phenotype Workbench
A visual workflow of genome variations in KBase.

Growmatch Data Viewer
Interactively view a set of growmatch data.

KBase is sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research.

IRIS: Command-Line On the Web

Metabolic workflow in IRIS



```
>/ kbfba-loadgenome kb|g.422
Genome successfully loaded to workspace:
>/ kbfba-buildfbamodel kb|g.422
Model successfully generated in workspace:
>/ kbfba-exportfbamodel kb|g.422.fbamdl.2 html > model.html
Command completed.
>/ kbfba-addmedia C-D-Mannitol "D-Mannitol;NH4+;Phosphate;Sulfate;Al;Boron;Ca2+;Cl-
;Co2+;Cu2+;Fe2+;Fe3+;H+;H2O;K+;Mg;Mn2+;Ni2+;Molybdate;Na+;O2;Zn2+" -d -type Minimal -e
Successfully added media to workspace:
>/ kbfba-runfba kb|g.422.fbamdl.2 -m C-D-Mannitol
Flux balance analysis successful:
>/ kbfba-exportfba kb|g.422.fbamdl.2.fba.2 html > fba.html
Command completed.
>/ kbfba-importpheno kb|g.422 WT-phenotypes.txt
Command completed.
>/ kbfba-simpheno kb|g.422.fbamdl.2 kb|g.422.phenos.0
Command completed.
>/ kbfba-exportphenosim kb|g.422.fbamdl.2.simpheno.1 html > phenosim.html
Command completed.
>/ kbfba-gapfill kb|g.422.fbamdl.2 -m C-D-Mannitol
Gapfilling job queued:
```

Command and result history

User enters commands

User data upload

Directory of services

IRIS is a web-based terminal to give you access to the full KBase command line without the need to install anything!

Command Line

- Command-line environment for programmers
- Available via downloadable Macintosh DMG or Ubuntu (linux) image
- Installs clean environment for KBase application development

Populus reads to SNPs in Five Easy Steps

1. Identify reference genome

```
$ all_entities_Genome -f scientific_name | grep -i ’Populus’
```

2. Upload Reads to KBase cloud

```
$ jk_fs_put_pe populus.1.fq.gz populus.2.fq.gz populus
```

3. Align Reads with Bowtie2

```
$ jk_compute_bowtie -in=populus.pe -org=populus -out=populus_align
```

4. Call SNPs with SAMTools

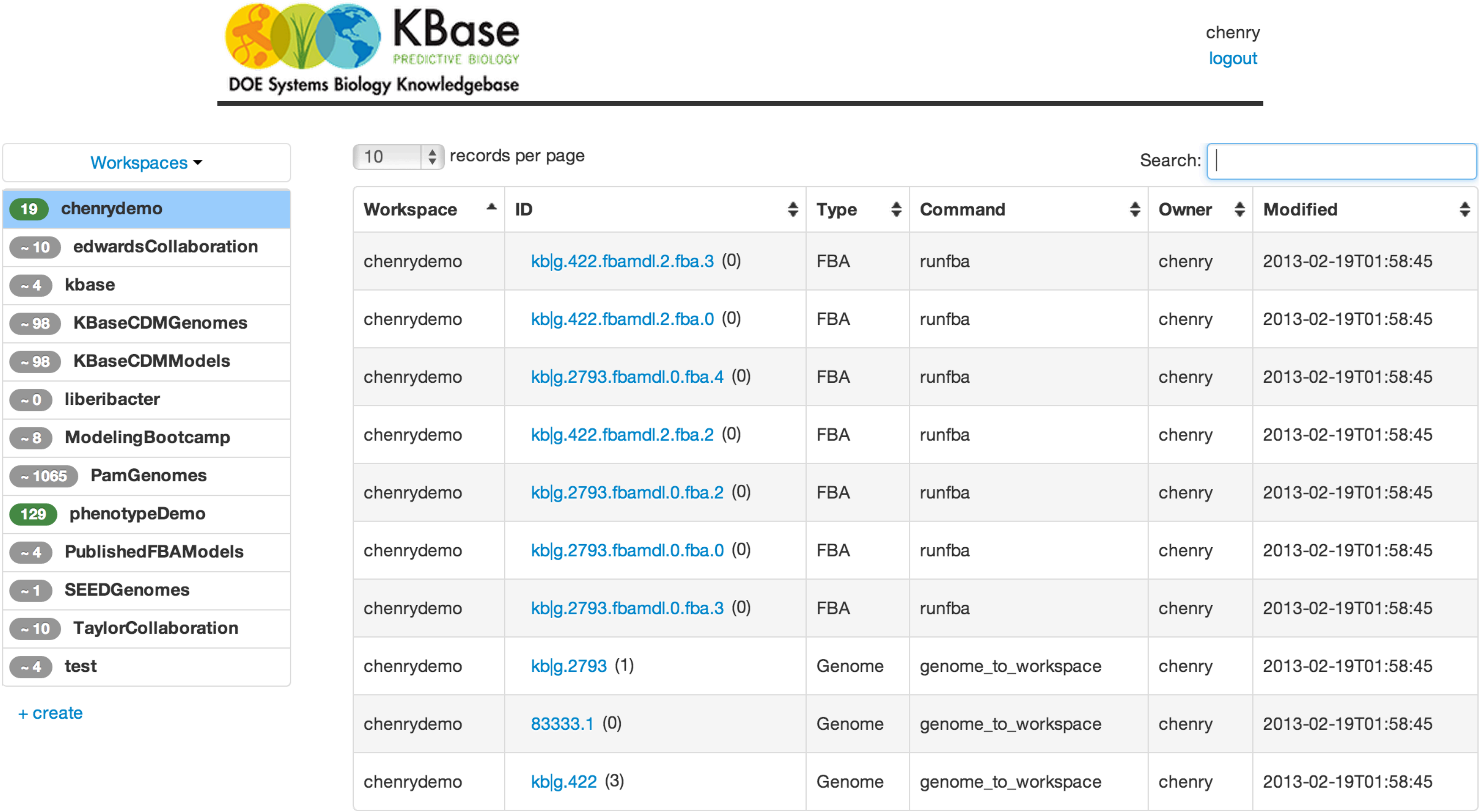
```
$ jk_compute_samtools_snp -in=populus_align -org=populus -out=populus_snps
```

5. Merge and Download VCF files

```
$ jk_compute_vcf_merge -in=populus_snps --alignments=populus_align -out=populus.vcf
$ jk_fs_get populus.vcf
```

Prototype collaborative workspaces

View Objects Produced by Analysis in an Online Collaborative Workspace



KBase PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

chenry
logout

Workspaces ▾

19	chenrydemo
~10	edwardsCollaboration
~4	kbase
~98	KBaseCDMGenomes
~98	KBaseCDMModels
~0	liberibacter
~8	ModelingBootcamp
~1065	PamGenomes
129	phenotypeDemo
~4	PublishedFBAModels
~1	SEEDGenomes
~10	TaylorCollaboration
~4	test

+ create

10 records per page

Search:

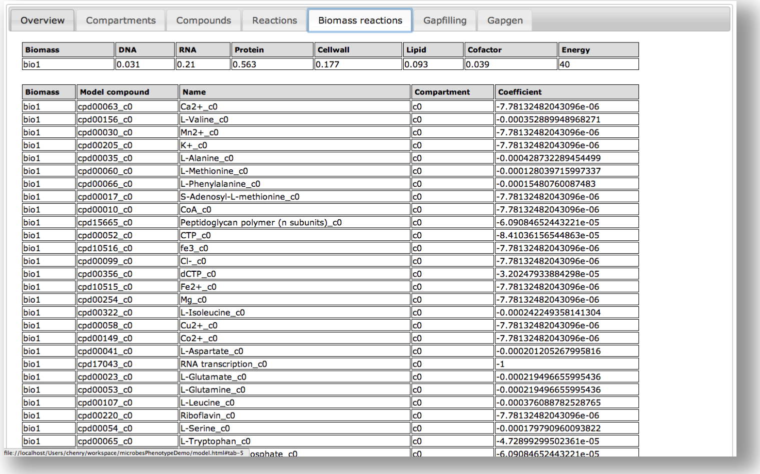
Workspace	ID	Type	Command	Owner	Modified
chenrydemo	kb|g.422.fbamdl.2.fba.3 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.422.fbamdl.2.fba.0 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.2793.fbamdl.0.fba.4 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.422.fbamdl.2.fba.2 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.2793.fbamdl.0.fba.2 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.2793.fbamdl.0.fba.0 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.2793.fbamdl.0.fba.3 (0)	FBA	runfba	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.2793 (1)	Genome	genome_to_workspace	chenry	2013-02-19T01:58:45
chenrydemo	83333.1 (0)	Genome	genome_to_workspace	chenry	2013-02-19T01:58:45
chenrydemo	kb|g.422 (3)	Genome	genome_to_workspace	chenry	2013-02-19T01:58:45

- Object model supports provenance, meaning all previous versions of the objects can be retrieved
- Workspace browser provides direct links to view any object contained in the workspace

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Visualizing Metabolic Models and FBA in KBase

All views may be used for all public and private models and FBA solutions

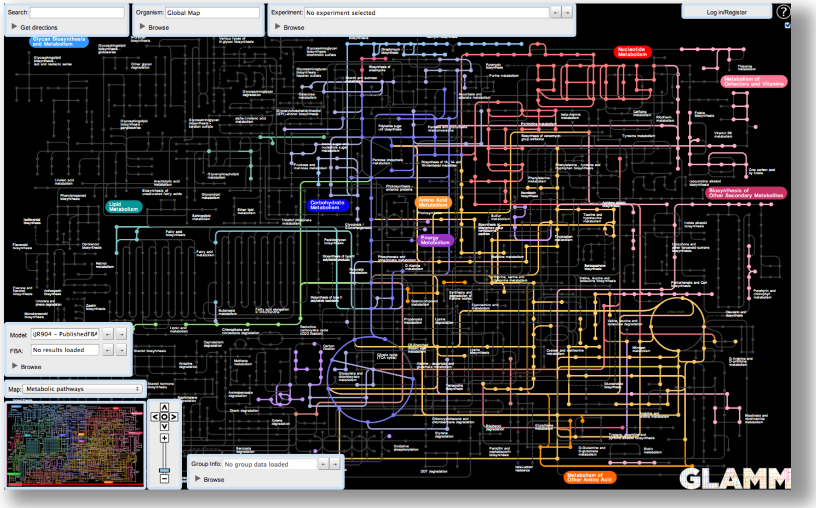


Biomass	DNA	RNA	Protein	Cellwall	Lipid	Cofactor	Energy
bio1	0.031	0.21	0.563	0.177	0.093	0.039	40

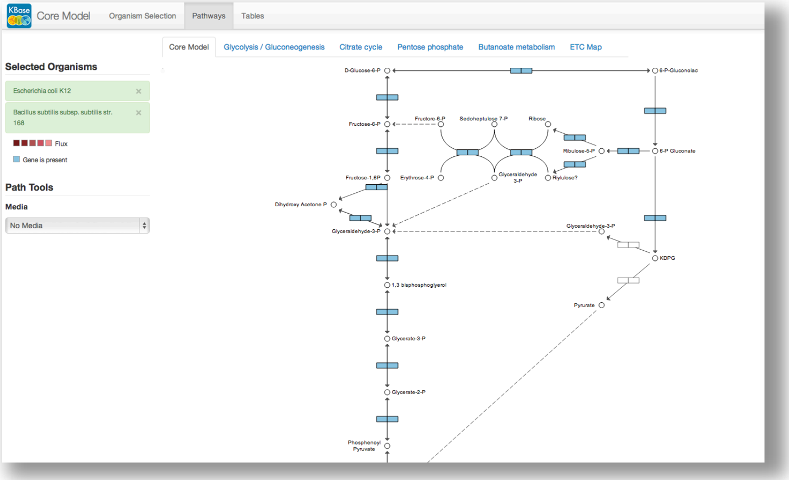
Biomass	Model compound	Name	Compartment	Coefficient
bio1	cpd00063_c0	Ca2+_c0	c0	-7.78132482043096e-06
bio1	cpd00156_c0	L-Valine_c0	c0	-0.000352889948968271
bio1	cpd00030_c0	Mn2+_c0	c0	-7.78132482043096e-06
bio1	cpd00205_c0	K+_c0	c0	-7.78132482043096e-06
bio1	cpd00035_c0	L-Alanine_c0	c0	-0.000428732289454499
bio1	cpd00060_c0	L-Methionine_c0	c0	-0.000128039715997337
bio1	cpd00066_c0	L-Phenylalanine_c0	c0	-0.00015480760087483
bio1	cpd00017_c0	S-Adenosyl-L-methionine_c0	c0	-7.78132482043096e-06
bio1	cpd00010_c0	CoA_c0	c0	-7.78132482043096e-06
bio1	cpd15665_c0	Peptidoglycan polymer (n subunits)_c0	c0	-6.09084652443221e-05
bio1	cpd00052_c0	CTP_c0	c0	-8.41036156544863e-05
bio1	cpd10516_c0	fe3_c0	c0	-7.78132482043096e-06
bio1	cpd00099_c0	Cl-_c0	c0	-7.78132482043096e-06
bio1	cpd00356_c0	dCTP_c0	c0	-3.20247933884298e-05
bio1	cpd10515_c0	Fe2+_c0	c0	-7.78132482043096e-06
bio1	cpd00254_c0	Mg_c0	c0	-7.78132482043096e-06
bio1	cpd00322_c0	L-Isoleucine_c0	c0	-0.000242249358141304
bio1	cpd00058_c0	Cu2+_c0	c0	-7.78132482043096e-06
bio1	cpd00149_c0	Co2+_c0	c0	-7.78132482043096e-06
bio1	cpd00041_c0	L-Aspartate_c0	c0	-0.000201205267995816
bio1	cpd17043_c0	RNA transcription_c0	c0	-1
bio1	cpd00023_c0	L-Glutamate_c0	c0	-0.000219496655995436
bio1	cpd00053_c0	L-Glutamine_c0	c0	-0.000219496655995436
bio1	cpd00107_c0	L-Leucine_c0	c0	-0.000376088782528765
bio1	cpd00220_c0	Riboflavin_c0	c0	-7.78132482043096e-06
bio1	cpd00054_c0	L-Serine_c0	c0	-0.000179790960093822
bio1	cpd00065_c0	L-Tryptophan_c0	c0	-4.72899299502361e-05
		osphate_c0	c0	-6.09084652443221e-05

file://localhost/Users/chenry/workspace/microbesPhenotypeDemo/model.html#tab-5

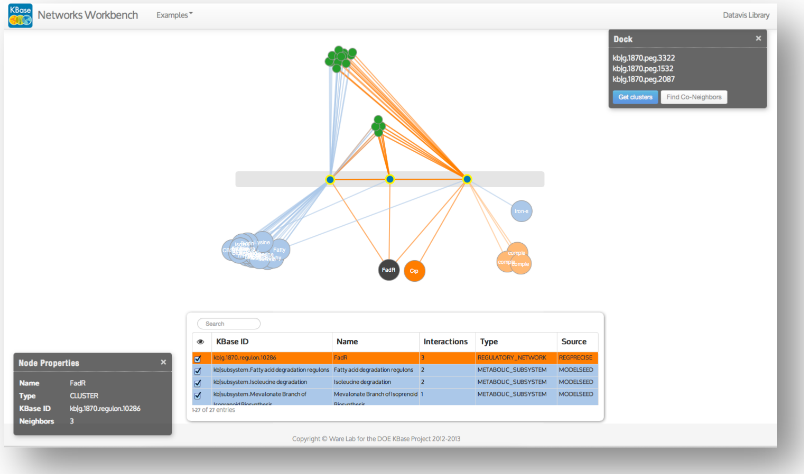
Simple Tabular Views



Interactive metabolic maps and models



Comparative Model Viewer



Node Properties

Name	FadR
Type	CLUSTER
KBase ID	kb|g.1870.regulon.10286
Neighbors	3

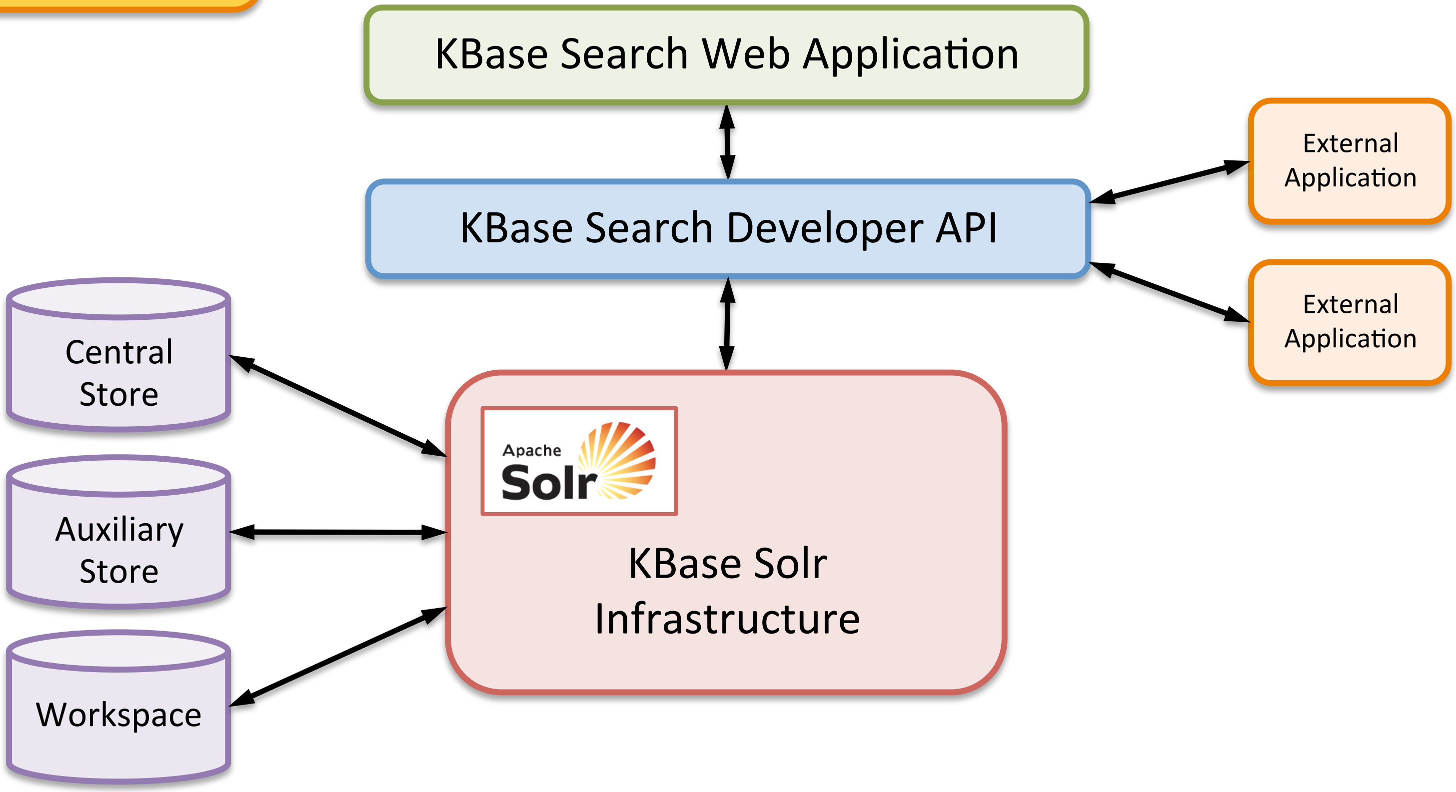
	KBase ID	Name	Interactions	Type	Source
☑	kb|g.1870.regulon.10286	FadR	3	REGULATORY_NETWORK	REGPRECISE
☑	kb|subsystem.Fatty acid degradation regulons	Fatty acid degradation regulons	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Isoleucine degradation	Isoleucine degradation	2	METABOLIC_SUBSYSTEM	MODELSEED
☑	kb|subsystem.Mevalonate Branch of Isoprenoid Biosynthesis	Mevalonate Branch of Isoprenoid Biosynthesis	1	METABOLIC_SUBSYSTEM	MODELSEED

1-27 of 27 entries

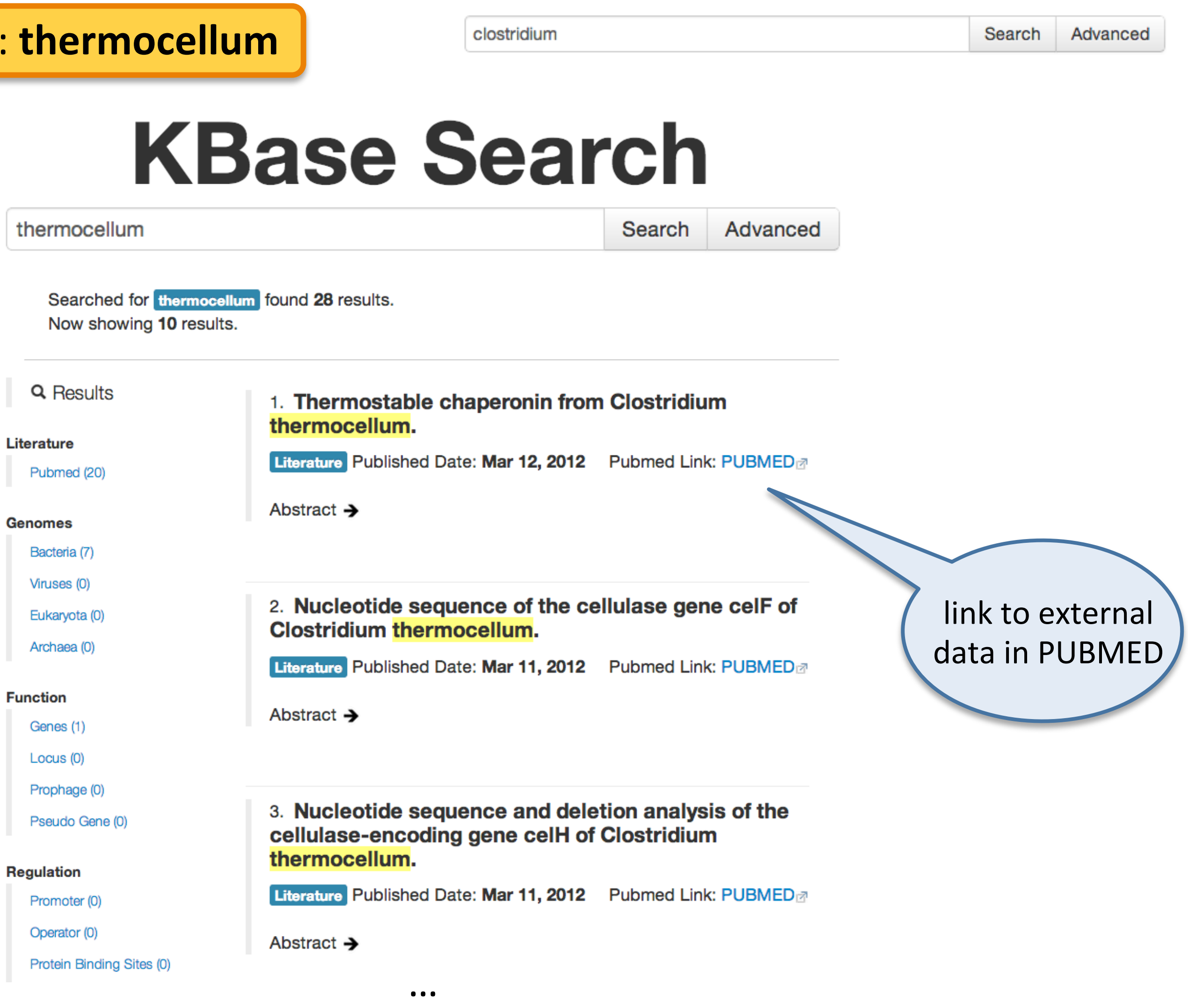
Network Analysis Workbench

Search Prototype

Search Architecture



Search with keyword: **thermocellum**



KBase Search

Searched for thermocellum found **28** results.
Now showing **10** results.

1. **Thermostable chaperonin from Clostridium thermocellum.**
   Literature Published Date: **Mar 12, 2012** Pubmed Link: PUBMED
   Abstract →

2. **Nucleotide sequence of the cellulase gene celF of Clostridium thermocellum.**
   Literature Published Date: **Mar 11, 2012** Pubmed Link: PUBMED
   Abstract →

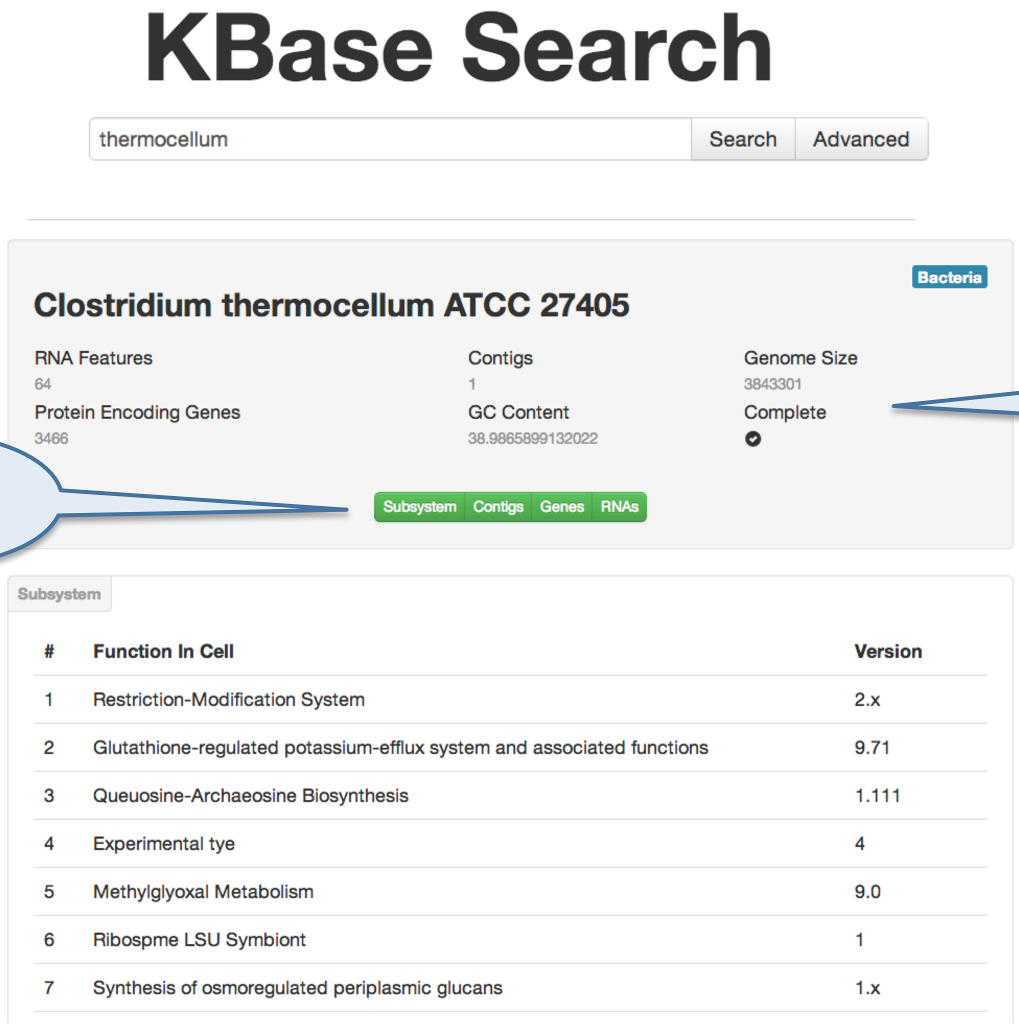
3. **Nucleotide sequence and deletion analysis of the cellulase-encoding gene celH of Clostridium thermocellum.**
   Literature Published Date: **Mar 11, 2012** Pubmed Link: PUBMED
   Abstract →

…

results organized using central data model

link to external data in PUBMED

Genome page



KBase Search

thermocellum | Search | Advanced

Bacteria

Clostridium thermocellum ATCC 27405

RNA Features
64
Protein Encoding Genes
3466

Contigs
1
GC Content
38.9865899132022

Genome Size
3843301
Complete
✓

Subsystem | Contigs | Genes | RNAs

Subsystem

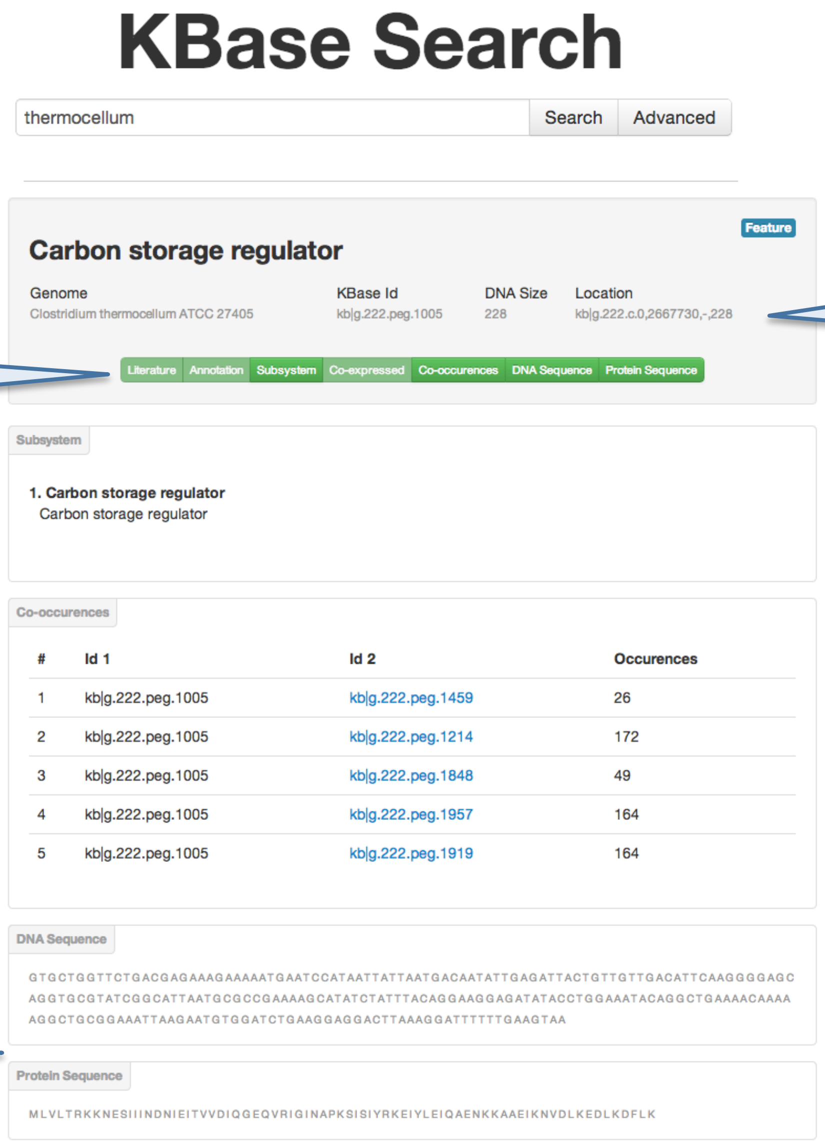
#	Function In Cell	Version
1	Restriction-Modification System	2.x
2	Glutathione-regulated potassium-efflux system and associated functions	9.71
3	Queuosine-Archaeosine Biosynthesis	1.111
4	Experimental tye	4
5	Methylglyoxal Metabolism	9.0
6	Ribospme LSU Symbiont	1
7	Synthesis of osmoregulated periplasmic glucans	1.x

genome details

links to features

…

Gene page



KBase Search

Carbon storage regulator

Genome	KBase Id	DNA Size	Location
Clostridium thermocellum ATCC 27405	kb|g.222.peg.1005	228	kb|g.222.c.0,2667730,-,228

Subsystem

1. **Carbon storage regulator**
   Carbon storage regulator

Co-occurences

#	Id 1	Id 2	Occurences
1	kb|g.222.peg.1005	kb|g.222.peg.1459	26
2	kb|g.222.peg.1005	kb|g.222.peg.1214	172
3	kb|g.222.peg.1005	kb|g.222.peg.1848	49
4	kb|g.222.peg.1005	kb|g.222.peg.1957	164
5	kb|g.222.peg.1005	kb|g.222.peg.1919	164

DNA Sequence

```
GTGCTGGTTCTGACGAGAAAGAAAAATGAATCCATAATTATTAATGACAATATTGAGATTACTGTTGTTGACATTCAAGGGGAGC
AGGTGCGTATCGGCATTAATGCGCCGAAAAGCATATCTATTTACAGGAAGGAGATATACCTGGAAATACAGGCTGAAAACAAAA
AGGCTGCGGAAATTAAGAATGTGGATCTGAAGGAGGACTTAAAGGATTTTTTGAAGTAA
```

Protein Sequence

```
MLVLTRKKNESIIINDNIEITVVDIQGEQVRIGINAPKSISIYRKEIYLEIQAENKKAAEIKNVDLKEDLKDFLK
```

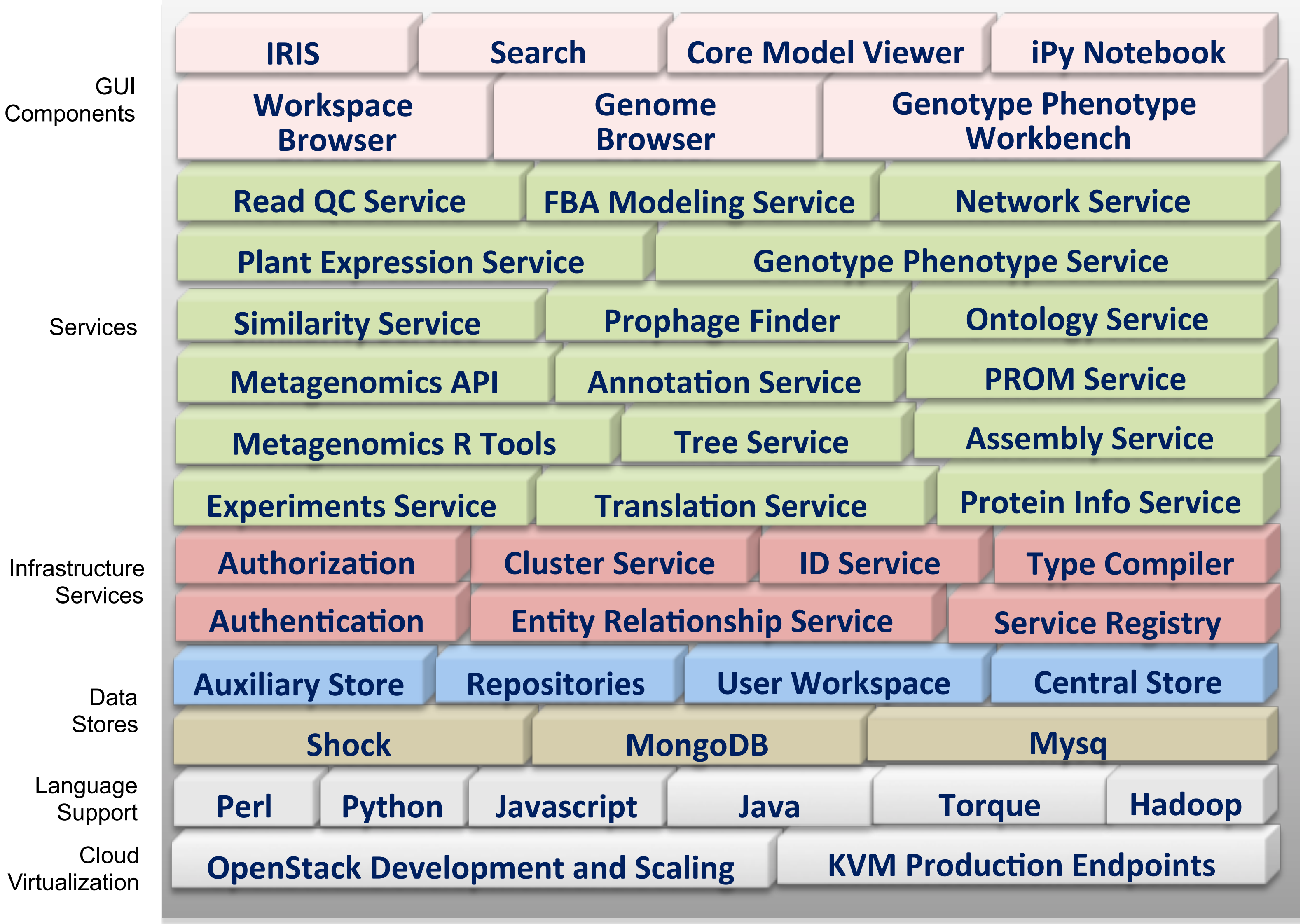
links to
related content

gene
details

sequences

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Initial Platform Modules



GUI Components
IRIS | Search | Core Model Viewer | iPy Notebook
Workspace Browser | Genome Browser | Genotype Phenotype Workbench

Services
Read QC Service | FBA Modeling Service | Network Service
Plant Expression Service | Genotype Phenotype Service
Similarity Service | Prophage Finder | Ontology Service
Metagenomics API | Annotation Service | PROM Service
Metagenomics R Tools | Tree Service | Assembly Service
Experiments Service | Translation Service | Protein Info Service

Infrastructure Services
Authorization | Cluster Service | ID Service | Type Compiler
Authentication | Entity Relationship Service | Service Registry

Data Stores
Auxiliary Store | Repositories | User Workspace | Central Store
Shock | MongoDB | Mysq

Language Support
Perl | Python | Javascript | Java | Torque | Hadoop

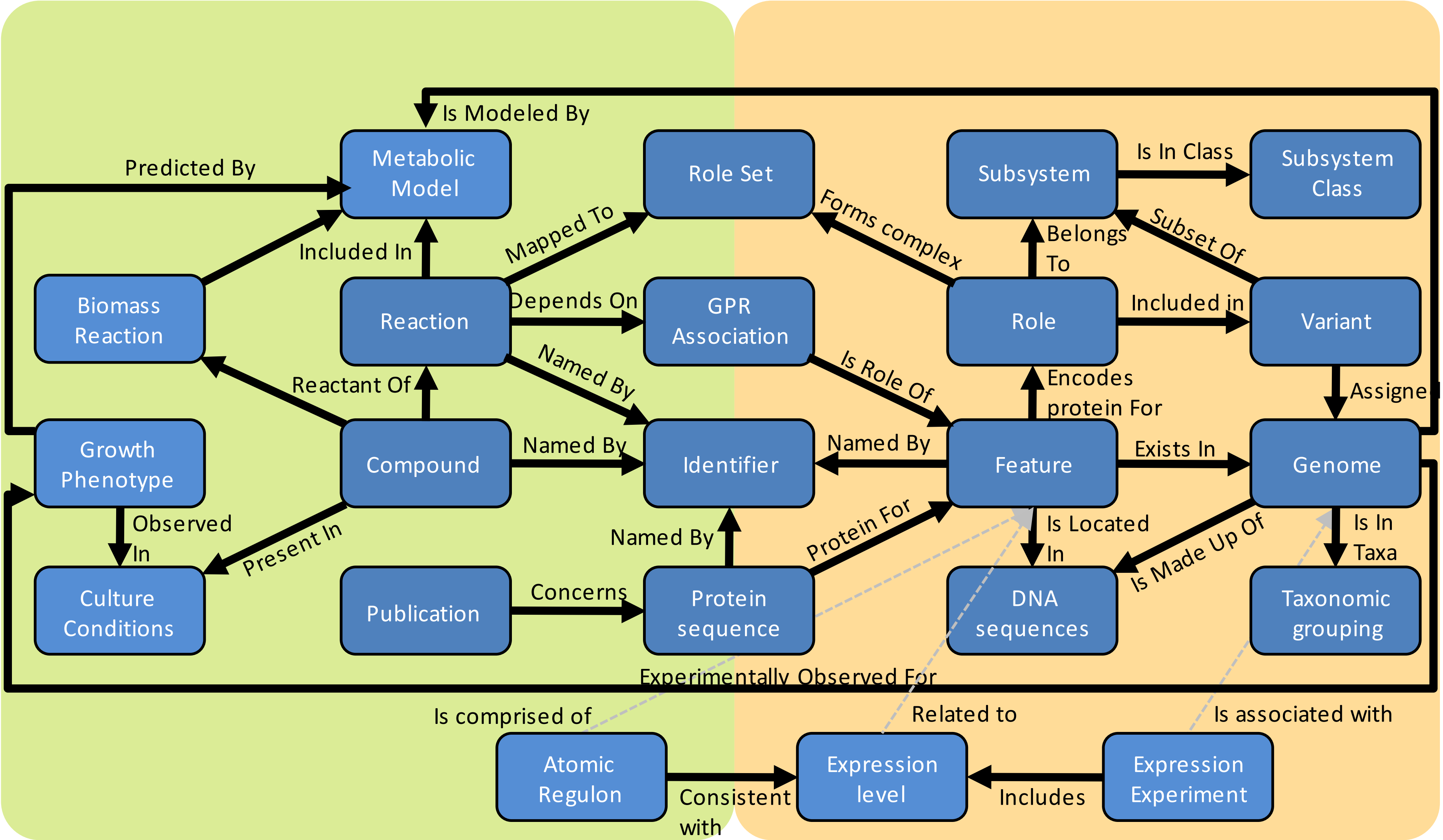
Cloud Virtualization
OpenStack Development and Scaling | KVM Production Endpoints

Initial KBase Services

- IRIS
- KBase Notebook
- Auxiliary Store Service
- Communities API
- Metagenomics Analysis Tools R
- QC Service
- Annotation Service
- Central Store
- Core Model Viewer
- Experiment data
- FBA Modeling
- Metabolic Map Viewer
- Microbes Model Builder
- Phispy
- Probabilistic Annotation
- Protein Info service
- Regulation Service
- Similarity Service
- Translation Service
- Workspace Service
- KB Model Seed
- Tree Service
- Assembly Service
- Authorization Service
- Network Service
- Genotype Phenotype Service
- Genotyping Service
- Ontology Service Plant
- Expression Service
- Authentication and Authorization Client
- Cluster Service
- ERDB service
- File Type Service
- ID Service
- Registry
- Type Compiler

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

KBase Central Data Model



Is Modeled By

Predicted By

Metabolic Model

Role Set

Subsystem

Is In Class

Subsystem Class

Included In

Mapped To

Forms complex

Belongs To

Subset Of

Biomass Reaction

Reaction

Depends On

GPR Association

Role

Included in

Variant

Reactant Of

Named By

Is Role Of

Encodes protein For

Assigne

Growth Phenotype

Compound

Named By

Identifier

Named By

Feature

Exists In

Genome

Observed In

Present In

Named By

Protein For

Is Located In

Is Made Up Of

Is In Taxa

Culture Conditions

Publication

Concerns

Protein sequence

DNA sequences

Taxonomic grouping

Experimentally Observed For

Is comprised of

Related to

Is associated with

Atomic Regulon

Consistent with

Expression level

Includes

Expression Experiment

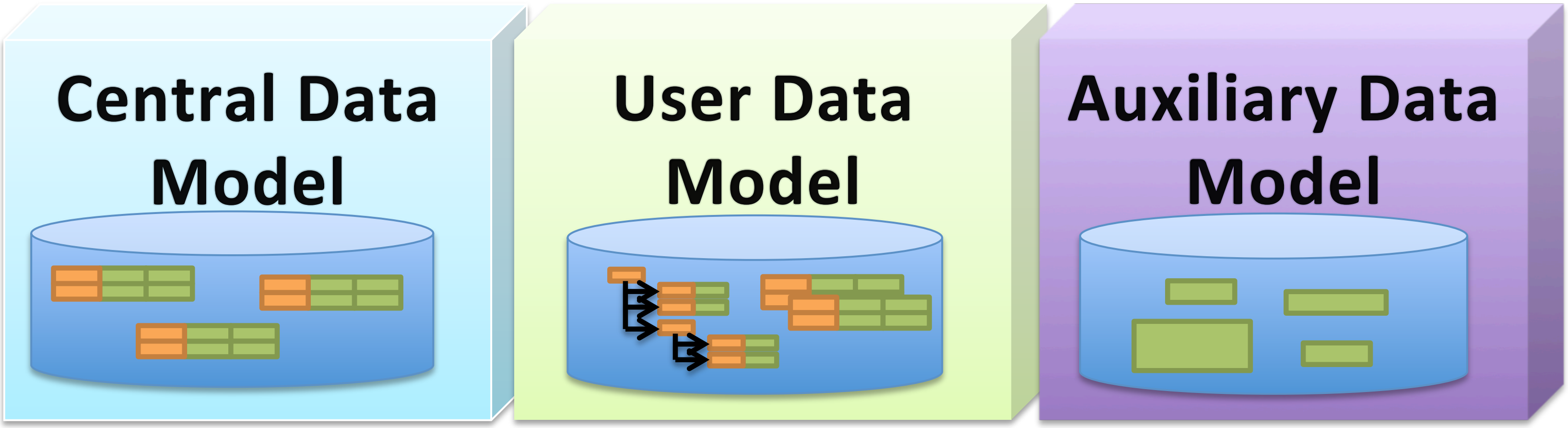
DOE Office of Science
Office of Biological and Environmental Research

U.S. DEPARTMENT OF
ENERGY

The KBase Data Stores

Data Stores: Storing a diverse range of biological data requires multiple approaches:

- Highly structured data in relational databases
- Frequently changing user data
- Large bulk data

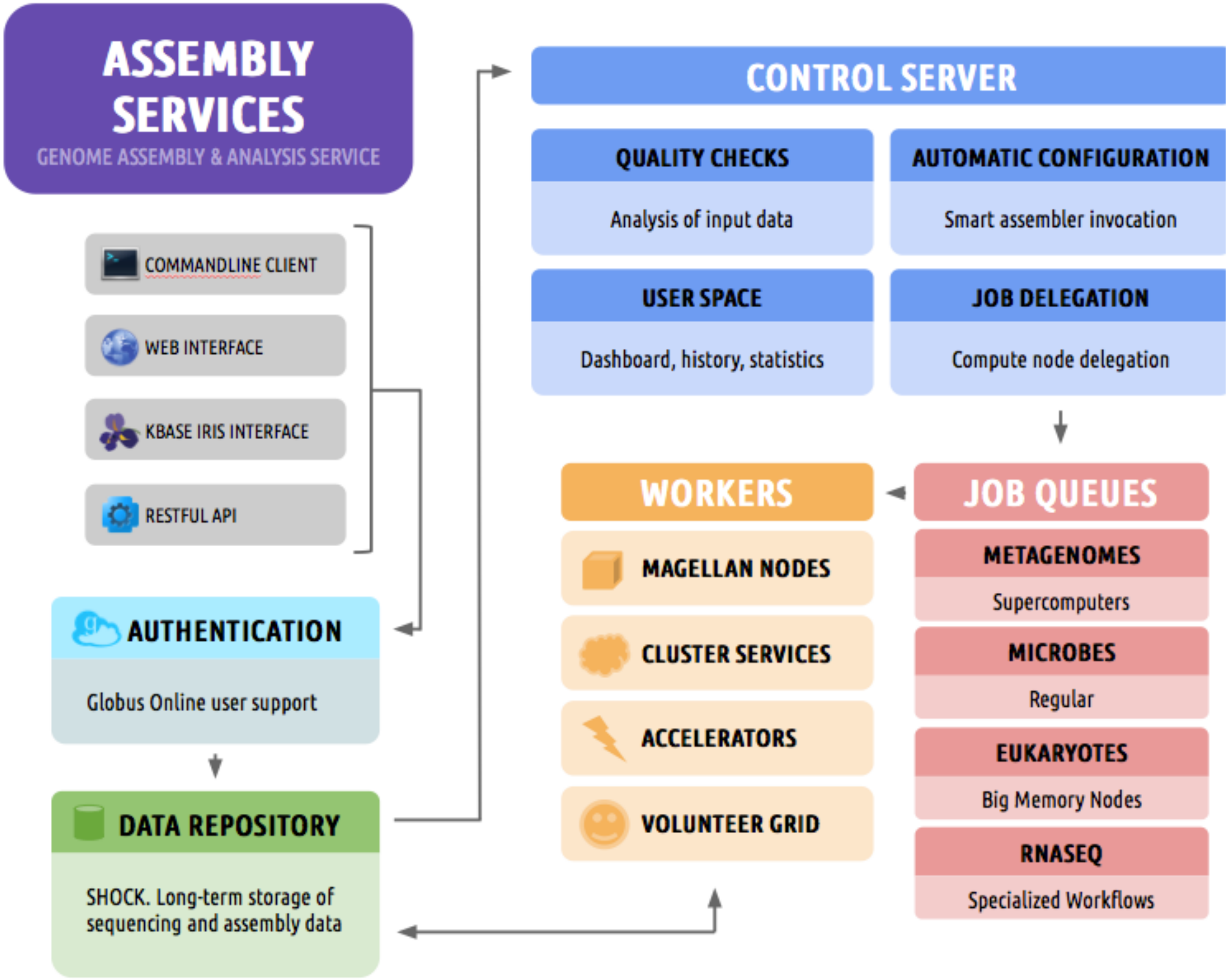


Petabytes of Raw Data

Flexible Storage For Workspaces

Structured Storage For Curated Data

The KBase Assembly Service



ASSEMBLY SERVICES

GENOME ASSEMBLY & ANALYSIS SERVICE

- COMMANDLINE CLIENT
- WEB INTERFACE
- KBASE IRIS INTERFACE
- RESTFUL API

AUTHENTICATION

Globus Online user support

DATA REPOSITORY

SHOCK. Long-term storage of sequencing and assembly data

CONTROL SERVER

QUALITY CHECKS

Analysis of input data

AUTOMATIC CONFIGURATION

Smart assembler invocation

USER SPACE

Dashboard, history, statistics

JOB DELEGATION

Compute node delegation

WORKERS

- MAGELLAN NODES
- CLUSTER SERVICES
- ACCELERATORS
- VOLUNTEER GRID

JOB QUEUES

METAGENOMES

Supercomputers

MICROBES

Regular

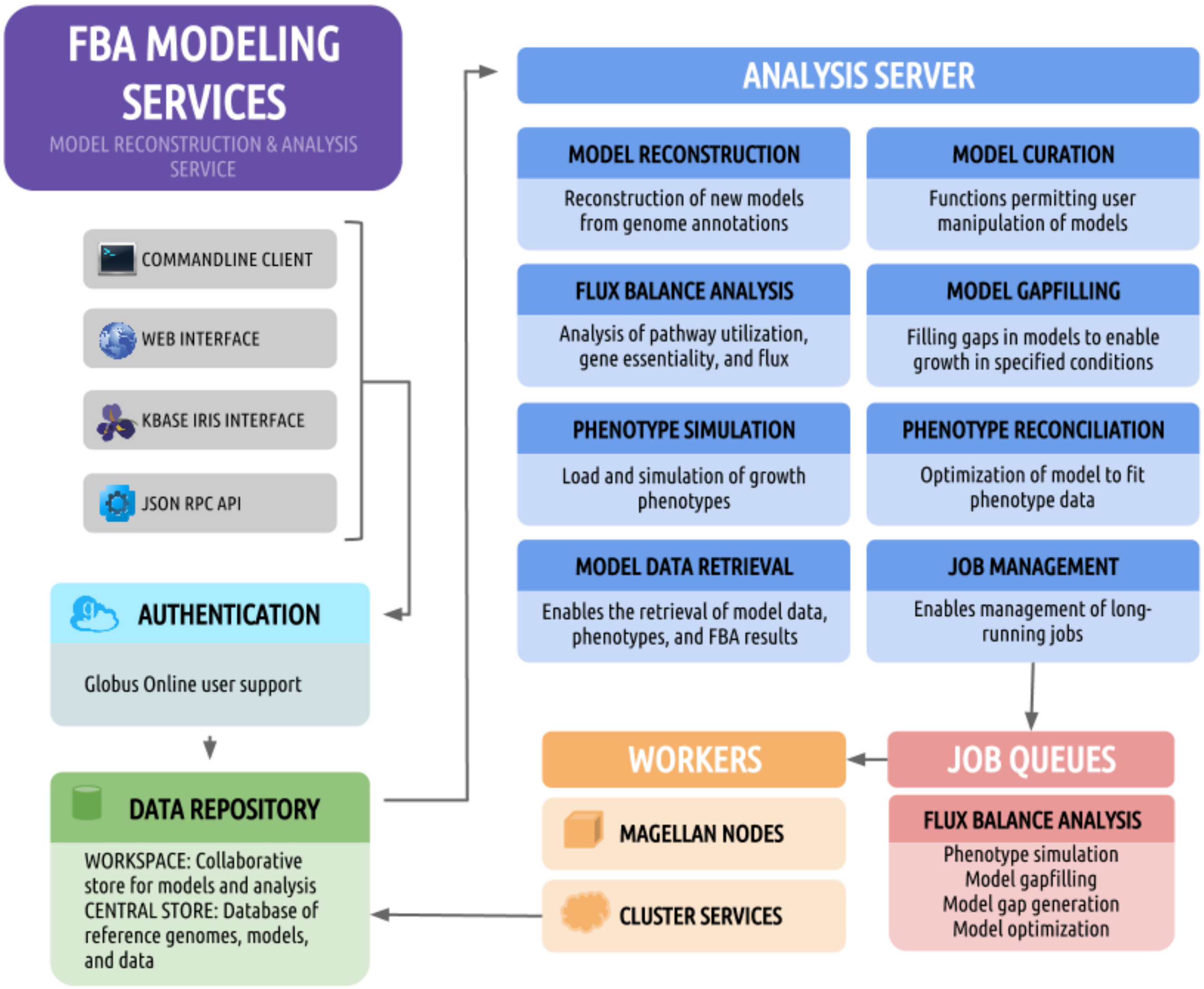
EUKARYOTES

Big Memory Nodes

RNASEQ

Specialized Workflows

Flux Balance Analysis Modeling Services



FBA MODELING SERVICES
MODEL RECONSTRUCTION & ANALYSIS SERVICE

- COMMANDLINE CLIENT
- WEB INTERFACE
- KBASE IRIS INTERFACE
- JSON RPC API

AUTHENTICATION
Globus Online user support

DATA REPOSITORY
WORKSPACE: Collaborative store for models and analysis
CENTRAL STORE: Database of reference genomes, models, and data

ANALYSIS SERVER

MODEL RECONSTRUCTION Reconstruction of new models from genome annotations	**MODEL CURATION** Functions permitting user manipulation of models
FLUX BALANCE ANALYSIS Analysis of pathway utilization, gene essentiality, and flux	**MODEL GAPFILLING** Filling gaps in models to enable growth in specified conditions
PHENOTYPE SIMULATION Load and simulation of growth phenotypes	**PHENOTYPE RECONCILIATION** Optimization of model to fit phenotype data
MODEL DATA RETRIEVAL Enables the retrieval of model data, phenotypes, and FBA results	**JOB MANAGEMENT** Enables management of long-running jobs

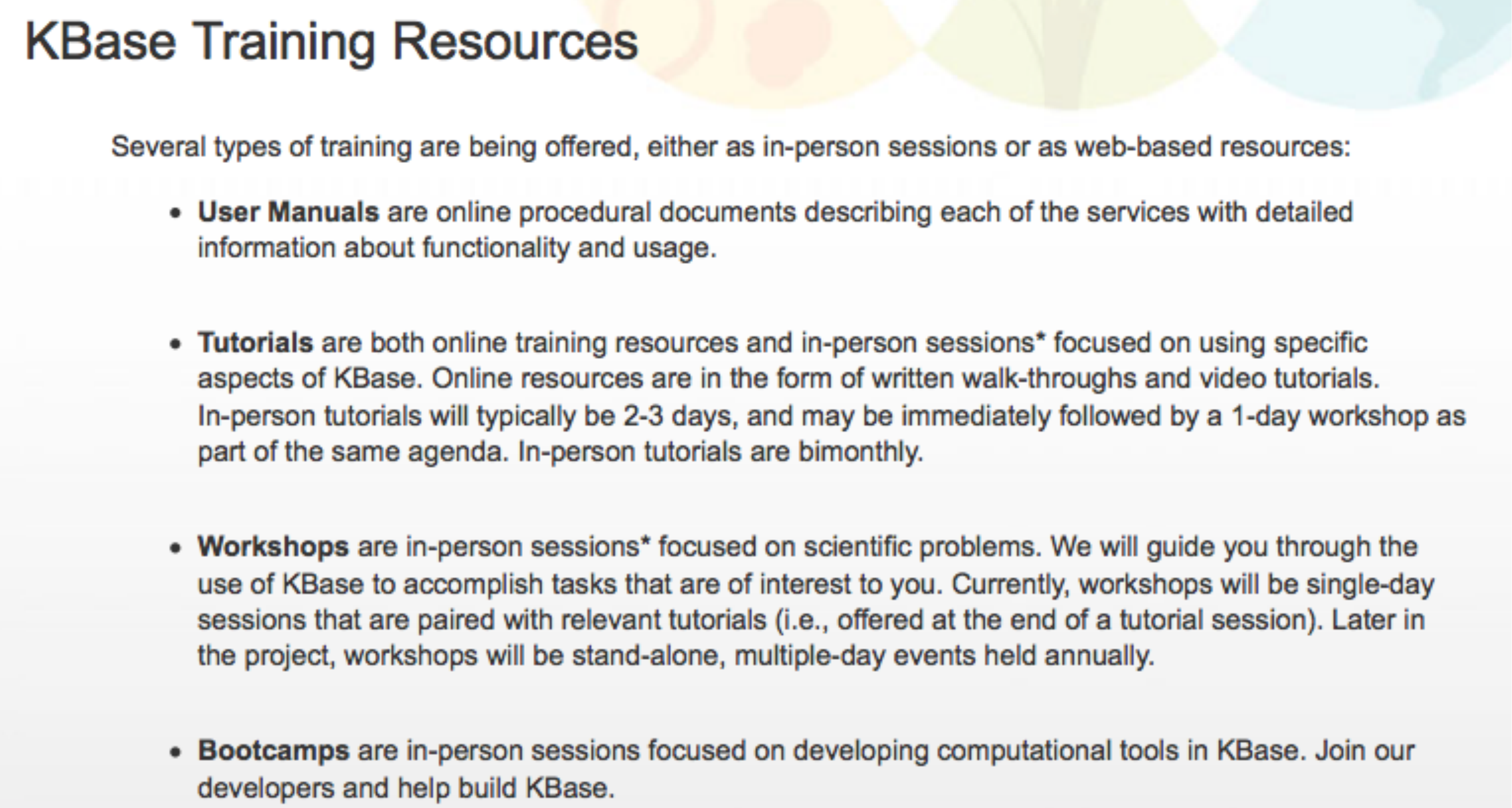
WORKERS
- MAGELLAN NODES
- CLUSTER SERVICES

JOB QUEUES
FLUX BALANCE ANALYSIS
Phenotype simulation
Model gapfilling
Model gap generation
Model optimization

Helping users and developers manage complexity

Website and Outreach

Training



KBase Training Resources

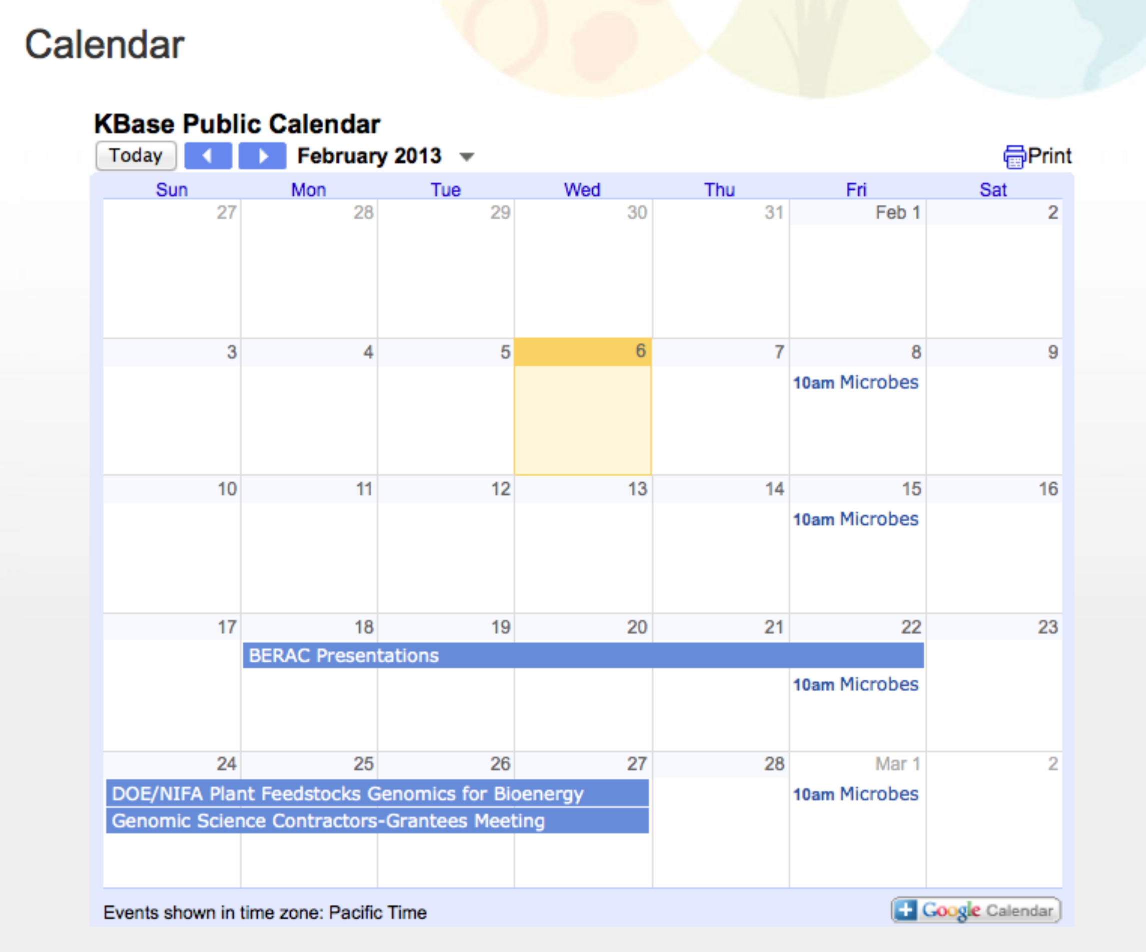
Several types of training are being offered, either as in-person sessions or as web-based resources:

- **User Manuals** are online procedural documents describing each of the services with detailed information about functionality and usage.
- **Tutorials** are both online training resources and in-person sessions* focused on using specific aspects of KBase. Online resources are in the form of written walk-throughs and video tutorials. In-person tutorials will typically be 2-3 days, and may be immediately followed by a 1-day workshop as part of the same agenda. In-person tutorials are bimonthly.
- **Workshops** are in-person sessions* focused on scientific problems. We will guide you through the use of KBase to accomplish tasks that are of interest to you. Currently, workshops will be single-day sessions that are paired with relevant tutorials (i.e., offered at the end of a tutorial session). Later in the project, workshops will be stand-alone, multiple-day events held annually.
- **Bootcamps** are in-person sessions focused on developing computational tools in KBase. Join our developers and help build KBase.

plus … Monthly webinars

When and Where

Scheduled KBase Events can be found at http://kbase.us



Calendar

KBase Public Calendar

Sun	Mon	Tue	Wed	Thu	Fri	Sat
27	28	29	30	31	Feb 1	2
3	4	5	6	7	8 10am Microbes	9
10	11	12	13	14	15 10am Microbes	16
17	18 BERAC Presentations	19	20	21	22 10am Microbes	23
24 DOE/NIFA Plant Feedstocks Genomics for Bioenergy Genomic Science Contractors-Grantees Meeting	25	26	27	28	Mar 1 10am Microbes	2

Events shown in time zone: Pacific Time

Upcoming Training Sessions

Bootcamps:

February 12-13, 2013: Full
March 6-7, 2013: Open
April 24-25, 2013: Open

Tutorials:

February 11-13, 2013: Full
March 11-13, 2013: Full
May 6-8, 2013: Open, topics include IRIS for Microbes, IRIS for Plants, and matR for Communities
July 15-17, 2013: Open, topics TBD

Workshops:

Requires concurrent registration in a Tutorial Session.
Feb 14, 2013: Full
March 14, 2013: Full
May 9, 2013: Open, topics TBD
July 18, 2013: Open, topics TBD

Contact us at outreach@kbase.us to register for a session.

Online Resources

- User manuals and tutorials
- Video tutorials (coming soon)
- Developer documentation
- Calendar of events
- FAQs
- Press

help@kbase.us

The Department of Energy Systems Biology Knowledgebase (KBase) is an emerging software and data environment designed to enable researchers to collaboratively generate, test and share new hypotheses about gene and protein functions, perform large-scale analyses on a scalable computing infrastructure, and model interactions in microbes, plants, and their communities. KBase provides an open, extensible framework for secure sharing of data, tools, and scientific conclusions in predictive and systems biology.

KBase includes

- 5695 prokaryotic genomes
- 175 eukaryotic genomes
- 4985 models
- 12 services

Search the database:

What can KBase do?

- Efficiently annotate new microbial genomes and infer metabolic and regulatory networks.
- Transform network inferences into metabolic models and map missing reactions to genes using novel data reconciliation tools.
- Test microbial ecological hypotheses through taxonomic and functional analysis of quality-assessed metagenomic data
- Discover genetic variations within plant populations and map these to complex organismal traits.

Latest News

KBase Developer's Minor Release
Posted by nlharris Jan 18, 2013

KBase at International Plant and Animal Genome XXI
Posted by salazar Jan 09, 2013

KBase Team at Argonne for November Build
Posted by salazar Nov 30, 2012

View news

Upcoming Events

2013-02-18
BERAC Presentations

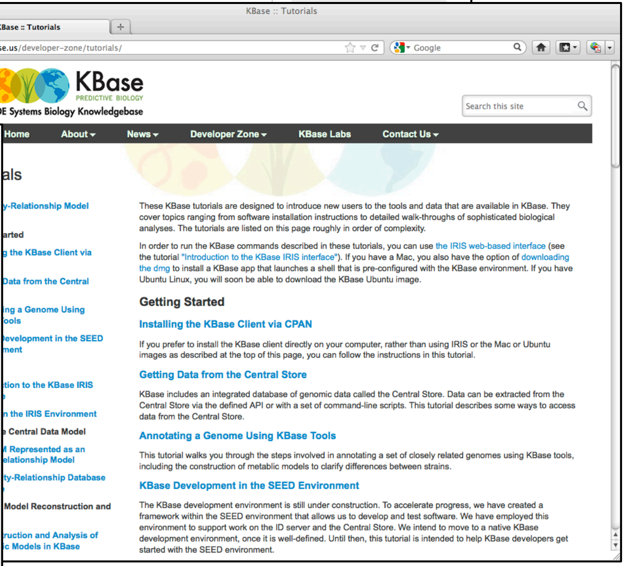
2013-02-22
Microbes Webinar

2013-02-24
DOE/NIFA Plant Feedstocks Genomics for Bioenergy

2013-02-24
Genomic Science Contractors-Grantees Meeting

View calendar

Sign up for a KBase acco



als

y-Relationship Model
arted
g the KBase Client via
Data from the Central
ing a Genome Using
ools
evelopment in the SEED
ment
tion to the KBase IRIS
n the IRIS Environment
e Central Data Model
l Represented as an
elationship Model
ty-Relationship Database
Model Reconstruction and
ruction and Analysis of
c Models in KBase

These KBase tutorials are designed to introduce new users to the tools and data that are available in KBase. They cover topics ranging from software installation instructions to detailed walk-throughs of sophisticated biological analyses. The tutorials are listed on this page roughly in order of complexity.

In order to run the KBase commands described in these tutorials, you can use the IRIS web-based interface (see the tutorial "Introduction to the KBase IRIS interface"). If you have a Mac, you also have the option of downloading the dmg to install a KBase app that launches a shell that is pre-configured with the KBase environment. If you have Ubuntu Linux, you will soon be able to download the KBase Ubuntu image.

Getting Started

Installing the KBase Client via CPAN

If you prefer to install the KBase client directly on your computer, rather than using IRIS or the Mac or Ubuntu images as described at the top of this page, you can follow the instructions in this tutorial.

Getting Data from the Central Store

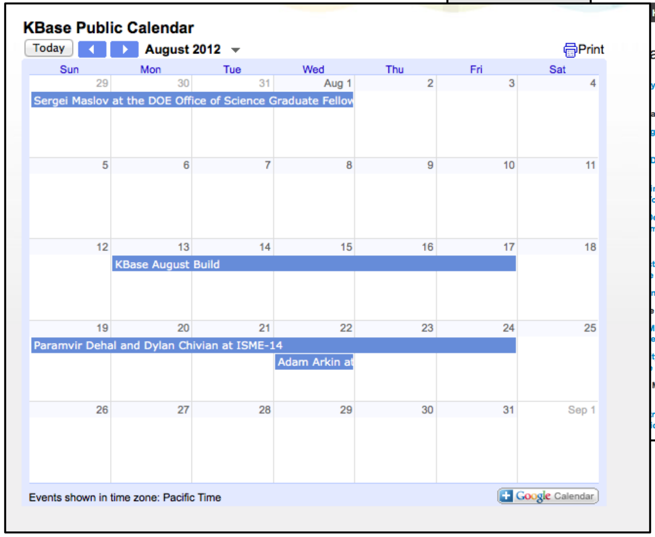
KBase includes an integrated database of genomic data called the Central Store. Data can be extracted from the Central Store via the defined API or with a set of command-line scripts. This tutorial describes some ways to access data from the Central Store.

Annotating a Genome Using KBase Tools

This tutorial walks you through the steps involved in annotating a set of closely related genomes using KBase tools, including the construction of metablic models to clarify differences between strains.

KBase Development in the SEED Environment

The KBase development environment is still under construction. To accelerate progress, we have created a framework within the SEED environment that allows us to develop and test software. We have employed this environment to support work on the ID server and the Central Store. We intend to move to a native KBase development environment, once it is well-defined. Until then, this tutorial is intended to help KBase developers get started with the SEED environment.


KBase Public Calendar

Sun	Mon	Tue	Wed	Thu	Fri	Sat
29 Sergei Maslov at the DOE Office of Science Graduate Fellow	30	31	Aug 1	2	3	4
5	6	7	8	9	10	11
12	13 KBase August Build	14	15	16	17	18
19 Paramvir Dehal and Dylan Chivian at ISME-14	20	21	22	23 Adam Arkin at	24	25
26	27	28	29	30	31	Sep 1

Events shown in time zone: Pacific Time

New Site for Easy Navigation

The Department of Energy Systems Biology Knowledgebase (KBase)

is an emerging software and data environment designed to enable researchers to collaboratively generate, test and share new hypotheses about gene and protein functions, perform large-scale analyses on a scalable computing infrastructure, and model interactions in microbes, plants, and their communities. KBase provides an open, extensible framework for secure sharing of data, tools, and scientific conclusions in predictive and systems biology.

Try KBase Now
Use a web-based command-line interface—no installation necessary

Download the Tools
Install and run KBase command-line tools on your computer

Visit KBase Labs
Sneak a peek at KBase applications in development

KBase includes

- 5695 prokaryotic genomes
- 175 eukaryotic genomes
- 4985 models
- 12 services

Search the database:
Search Advanced search

What can KBase do?

- Efficiently annotate new microbial genomes and infer metabolic and regulatory networks.
- Transform network inferences into metabolic models and map missing reactions to genes using novel data reconciliation tools.
- Test microbial ecological hypotheses through taxonomic and functional analysis of quality-assessed metagenomic data
- Discover genetic variations within plant populations and map these to complex organismal traits.

Glimpse the future

Latest News

KBase Developer's Minor Release
Posted by nlharris Jan 18, 2013

KBase at International Plant and Animal Genome XXI
Posted by salazar Jan 09, 2013

KBase Team at Argonne for November Build
Posted by salazar Nov 30, 2012

View news

Upcoming Events

2013-02-18
BERAC Presentations

2013-02-22
Microbes Webinar

2013-02-24
DOE/NIFA Plant Feedstocks Genomics for Bioenergy

2013-02-24
Genomic Science Contractors-Grantees Meeting

View calendar

Sign up for a KBase account

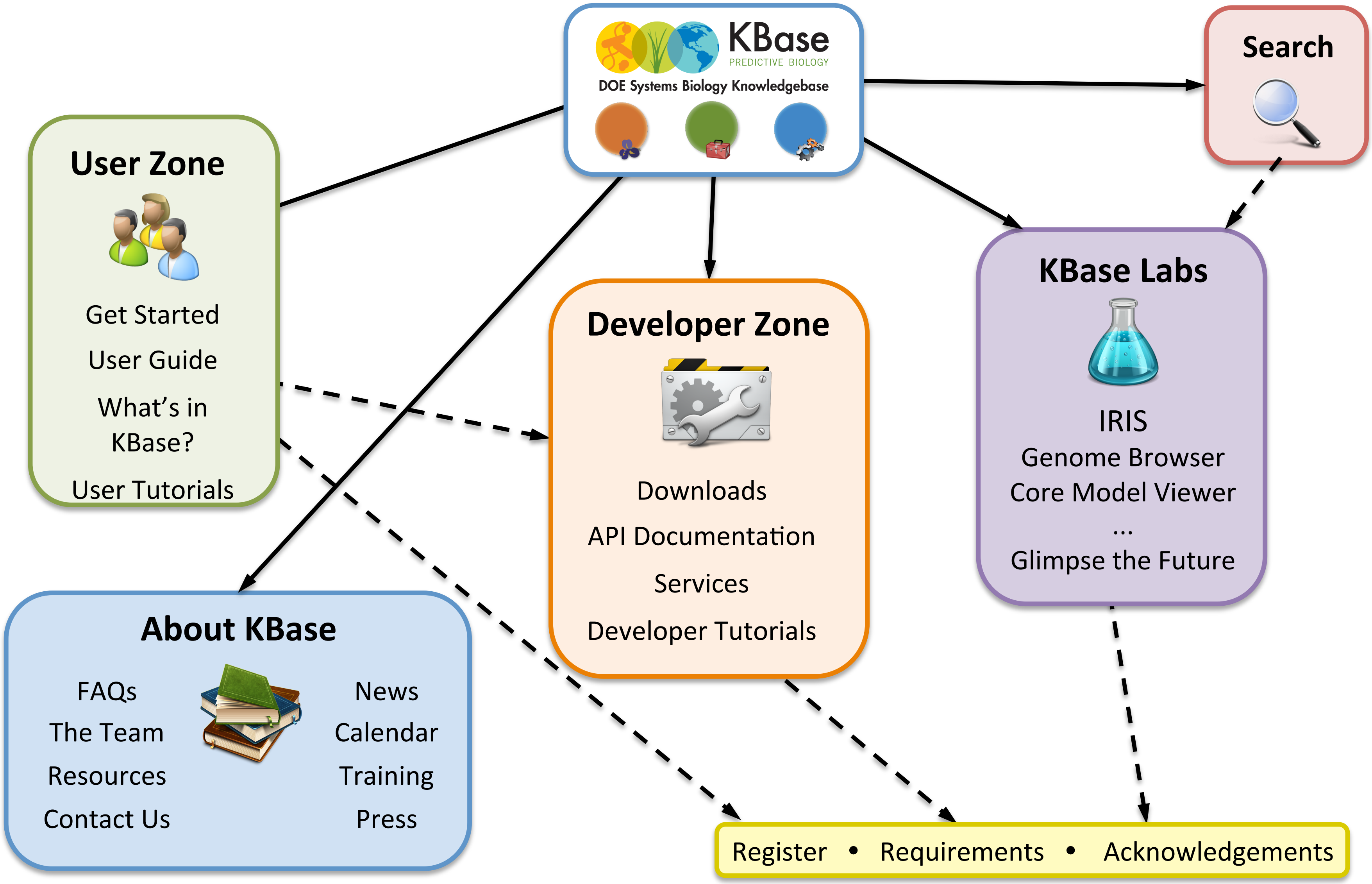
KBase is sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY Office of Science

Acknowledgements
Privacy and Security

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Sitemap for revamped kbase.us website



Search

User Zone

Get Started
User Guide
What's in KBase?
User Tutorials

Developer Zone

Downloads
API Documentation
Services
Developer Tutorials

KBase Labs

IRIS
Genome Browser
Core Model Viewer
...
Glimpse the Future

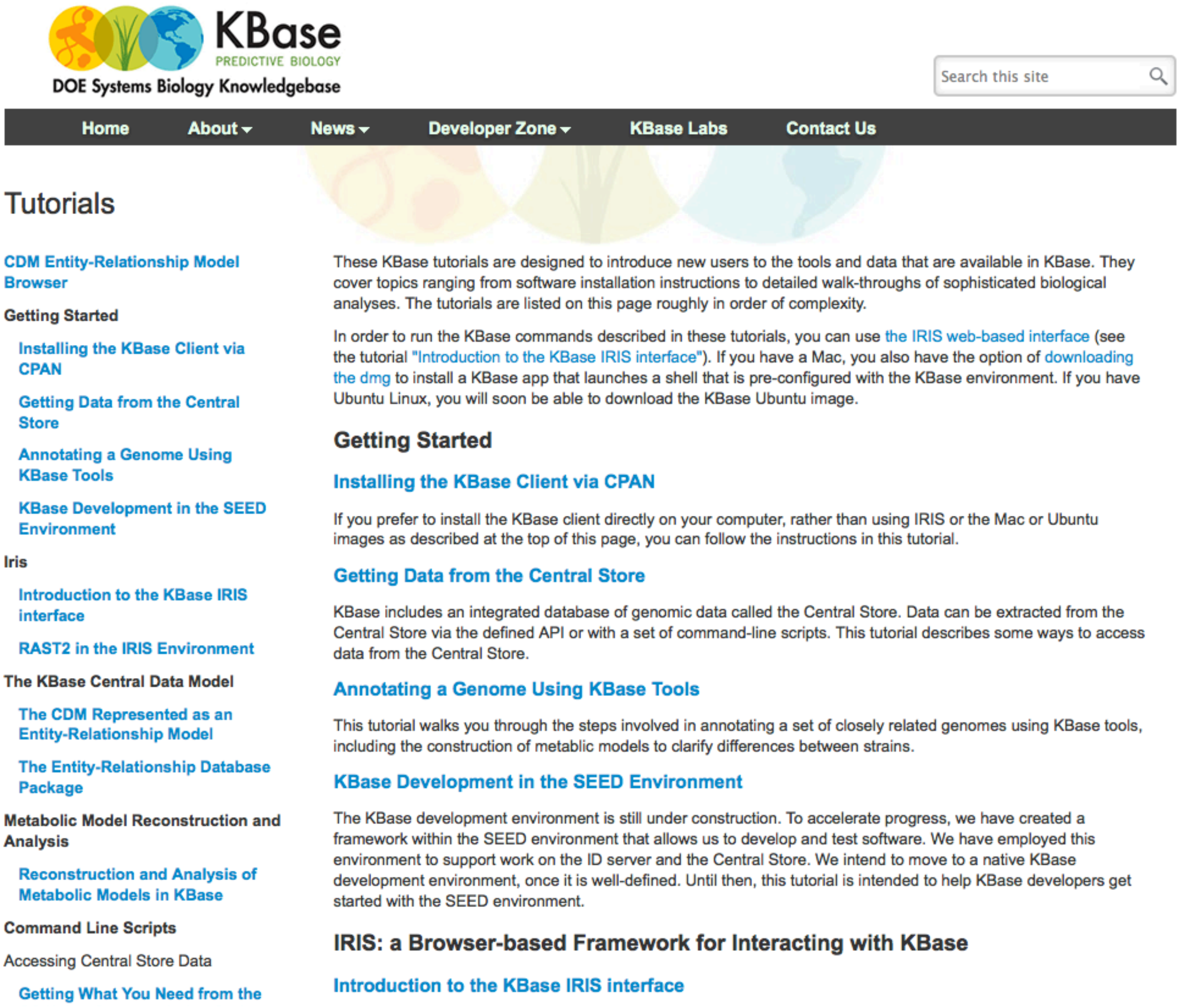
About KBase

FAQs
The Team
Resources
Contact Us

News
Calendar
Training
Press

Register • Requirements • Acknowledgements

Extensive Tutorials and Documentation



Tutorials

- CDM Entity-Relationship Model Browser
- Getting Started
  - Installing the KBase Client via CPAN
  - Getting Data from the Central Store
  - Annotating a Genome Using KBase Tools
  - KBase Development in the SEED Environment
- Iris
  - Introduction to the KBase IRIS interface
  - RAST2 in the IRIS Environment
- The KBase Central Data Model
  - The CDM Represented as an Entity-Relationship Model
  - The Entity-Relationship Database Package
- Metabolic Model Reconstruction and Analysis
  - Reconstruction and Analysis of Metabolic Models in KBase
- Command Line Scripts
- Accessing Central Store Data
  - Getting What You Need from the

These KBase tutorials are designed to introduce new users to the tools and data that are available in KBase. They cover topics ranging from software installation instructions to detailed walk-throughs of sophisticated biological analyses. The tutorials are listed on this page roughly in order of complexity.

In order to run the KBase commands described in these tutorials, you can use the IRIS web-based interface (see the tutorial "Introduction to the KBase IRIS interface"). If you have a Mac, you also have the option of downloading the dmg to install a KBase app that launches a shell that is pre-configured with the KBase environment. If you have Ubuntu Linux, you will soon be able to download the KBase Ubuntu image.

Getting Started

Installing the KBase Client via CPAN

If you prefer to install the KBase client directly on your computer, rather than using IRIS or the Mac or Ubuntu images as described at the top of this page, you can follow the instructions in this tutorial.

Getting Data from the Central Store

KBase includes an integrated database of genomic data called the Central Store. Data can be extracted from the Central Store via the defined API or with a set of command-line scripts. This tutorial describes some ways to access data from the Central Store.

Annotating a Genome Using KBase Tools

This tutorial walks you through the steps involved in annotating a set of closely related genomes using KBase tools, including the construction of metablic models to clarify differences between strains.

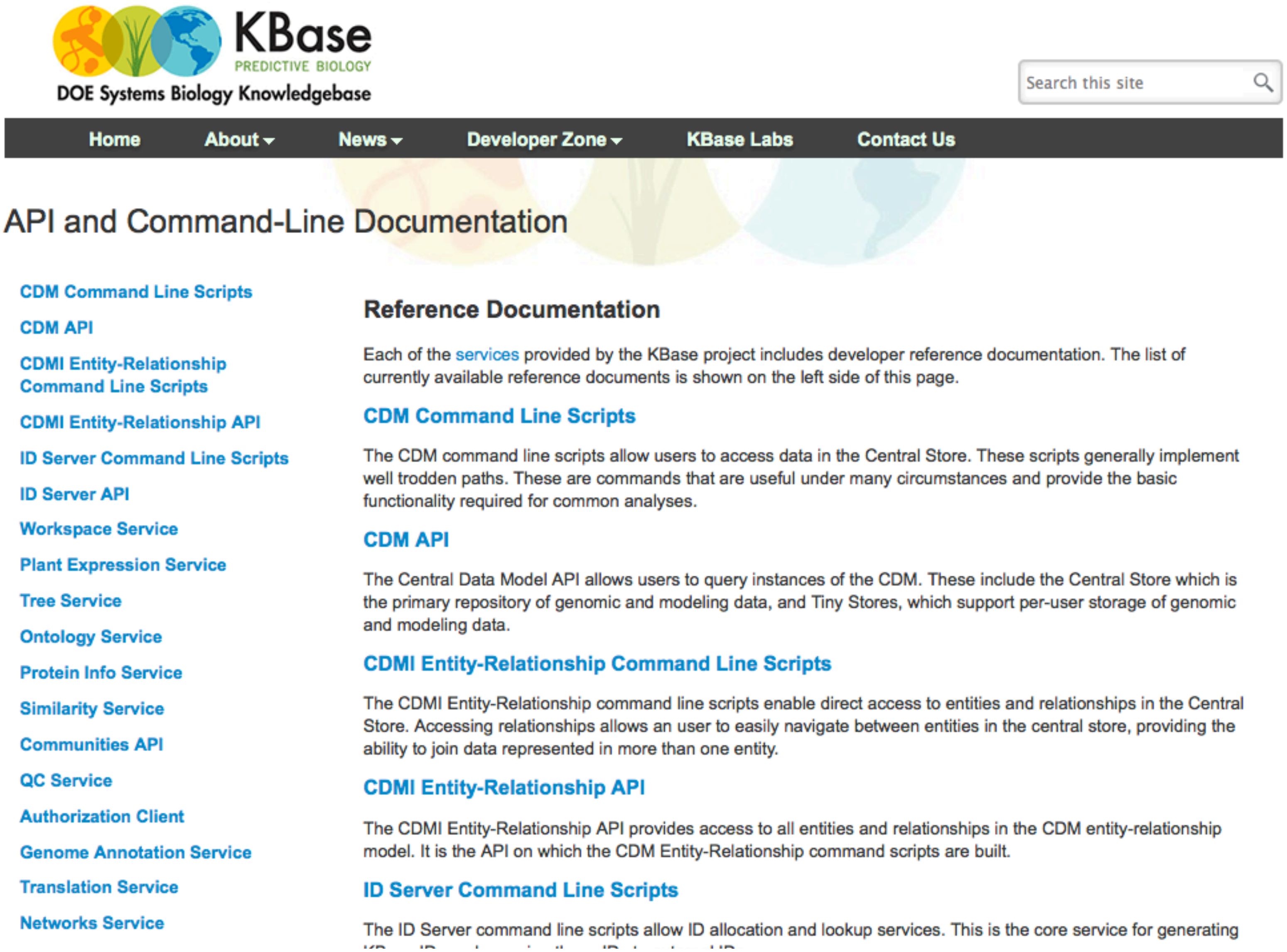
KBase Development in the SEED Environment

The KBase development environment is still under construction. To accelerate progress, we have created a framework within the SEED environment that allows us to develop and test software. We have employed this environment to support work on the ID server and the Central Store. We intend to move to a native KBase development environment, once it is well-defined. Until then, this tutorial is intended to help KBase developers get started with the SEED environment.

IRIS: a Browser-based Framework for Interacting with KBase

Introduction to the KBase IRIS interface

Extensive Tutorials and Documentation



API and Command-Line Documentation

- CDM Command Line Scripts
- CDM API
- CDMI Entity-Relationship Command Line Scripts
- CDMI Entity-Relationship API
- ID Server Command Line Scripts
- ID Server API
- Workspace Service
- Plant Expression Service
- Tree Service
- Ontology Service
- Protein Info Service
- Similarity Service
- Communities API
- QC Service
- Authorization Client
- Genome Annotation Service
- Translation Service
- Networks Service

Reference Documentation

Each of the services provided by the KBase project includes developer reference documentation. The list of currently available reference documents is shown on the left side of this page.

CDM Command Line Scripts

The CDM command line scripts allow users to access data in the Central Store. These scripts generally implement well trodden paths. These are commands that are useful under many circumstances and provide the basic functionality required for common analyses.

CDM API

The Central Data Model API allows users to query instances of the CDM. These include the Central Store which is the primary repository of genomic and modeling data, and Tiny Stores, which support per-user storage of genomic and modeling data.

CDMI Entity-Relationship Command Line Scripts

The CDMI Entity-Relationship command line scripts enable direct access to entities and relationships in the Central Store. Accessing relationships allows an user to easily navigate between entities in the central store, providing the ability to join data represented in more than one entity.

CDMI Entity-Relationship API

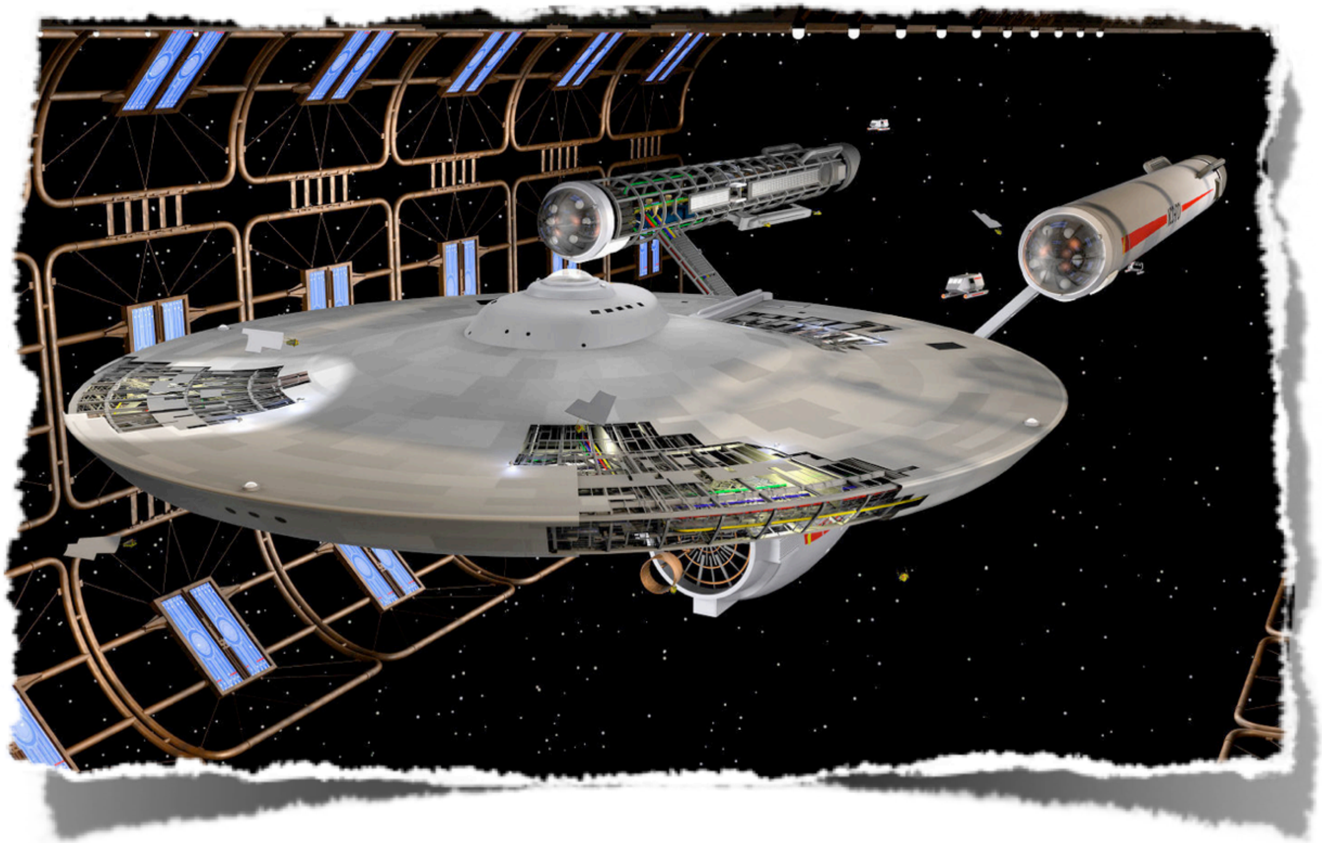
The CDMI Entity-Relationship API provides access to all entities and relationships in the CDM entity-relationship model. It is the API on which the CDM Entity-Relationship command scripts are built.

ID Server Command Line Scripts

The ID Server command line scripts allow ID allocation and lookup services. This is the core service for generating

MOAs and Collaborations

- Coordination and shared milestones with the BRCs
- Infrastructure collaborations with JGI, EMSL, NERC.
- Scientific collaborations with ENIGMA, PMI, and other SFAs
- MOAs with NSF NSF for iPLANT and pending for USDA
- Preparation for collaboration and support for future DOE calls
- Needs survey, bootcamps, and scientific discussions with >70 community scientists.



What next?

Building a starship in deep space?

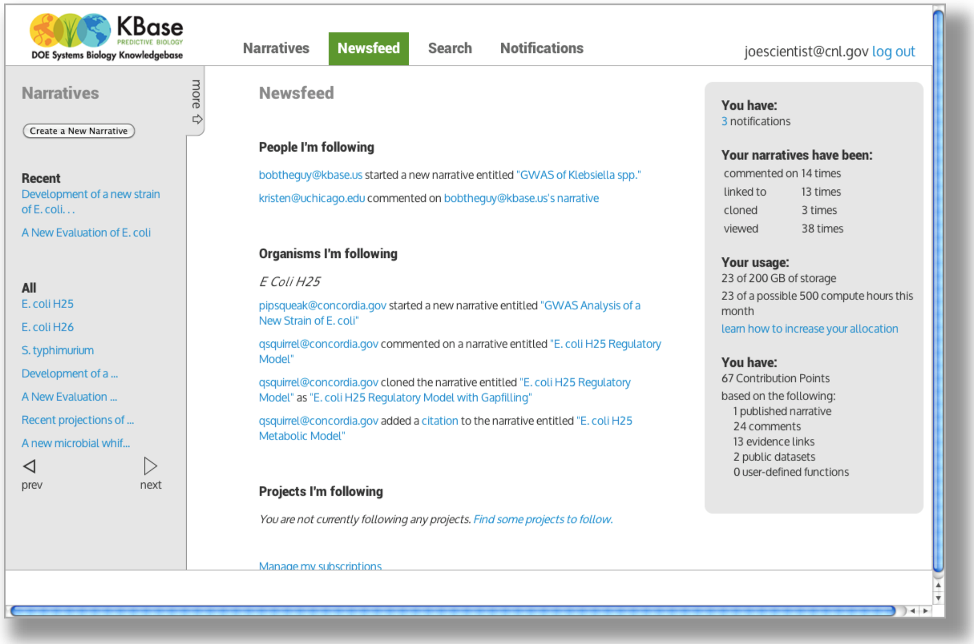
Into the future

- The data model must evolve to support the modeling mission
- A framework for turning bioinformatics algorithms' output into models needs further development
- **Theory for integrative, cross-scale predictive biology under development**
- Much better ontologies for nearly everything
- Building the social tools.
- Need to improve data import, quality assessment and metadata
- Efficient incorporation of new third party algorithms and support with scalable computing.
- Growing a strong external development community while maintaining quality, stability and vision.
- Launching the KBase Foundation to ease licensing and growth of KBase user participation.

A more concrete view of the "Narrative" interface

Narrative Desiderata

- **Users can share knowledge at at multiple levels of granularity**
  - Cut and paste parts of narratives to reuse workflows
  - Cross-citation and branching of narratives
- **Project management**
  - Creation of teams allows management of projects
  - Electronic lab notebook for computational researchers
  - Projects can track progress
- **Possible publication model**
  - Finalized Narratives are reviewed, assigned DOI numbers, and accepted in an appropriate journal
  - Citation metrics
- **Metrics of Research Efficiency**
  - Times to completion of narratives
  - Times from hypothesis to confirming data/narratives
- **User, data, and algorithm ratings** by how many times the products appear in or are cited by narratives.
  - Can be aggregated by user, team, location, agency
  - People networks can be inferred by looking at team, citation, and comment structures.



Narratives

Recent
Development of a new strain of E. coli. . .
A New Evaluation of E. coli

All
E. coli H25
E. coli H26
S. typhimurium
Development of a ...
A New Evaluation ...
Recent projections of ...
A new microbial whif...

Newsfeed

People I'm following

bobtheguy@kbase.us started a new narrative entitled "GWAS of Klebsiella spp."

kristen@uchicago.edu commented on bobtheguy@kbase.us's narrative

Organisms I'm following

E Coli H25

pipsqueak@concordia.gov started a new narrative entitled "GWAS Analysis of a New Strain of E. coli"

qsquirrel@concordia.gov commented on a narrative entitled "E. coli H25 Regulatory Model"

qsquirrel@concordia.gov cloned the narrative entitled "E. coli H25 Regulatory Model" as "E. coli H25 Regulatory Model with Gapfilling"

qsquirrel@concordia.gov added a citation to the narrative entitled "E. coli H25 Metabolic Model"

Projects I'm following

You are not currently following any projects. Find some projects to follow.

Manage my subscriptions

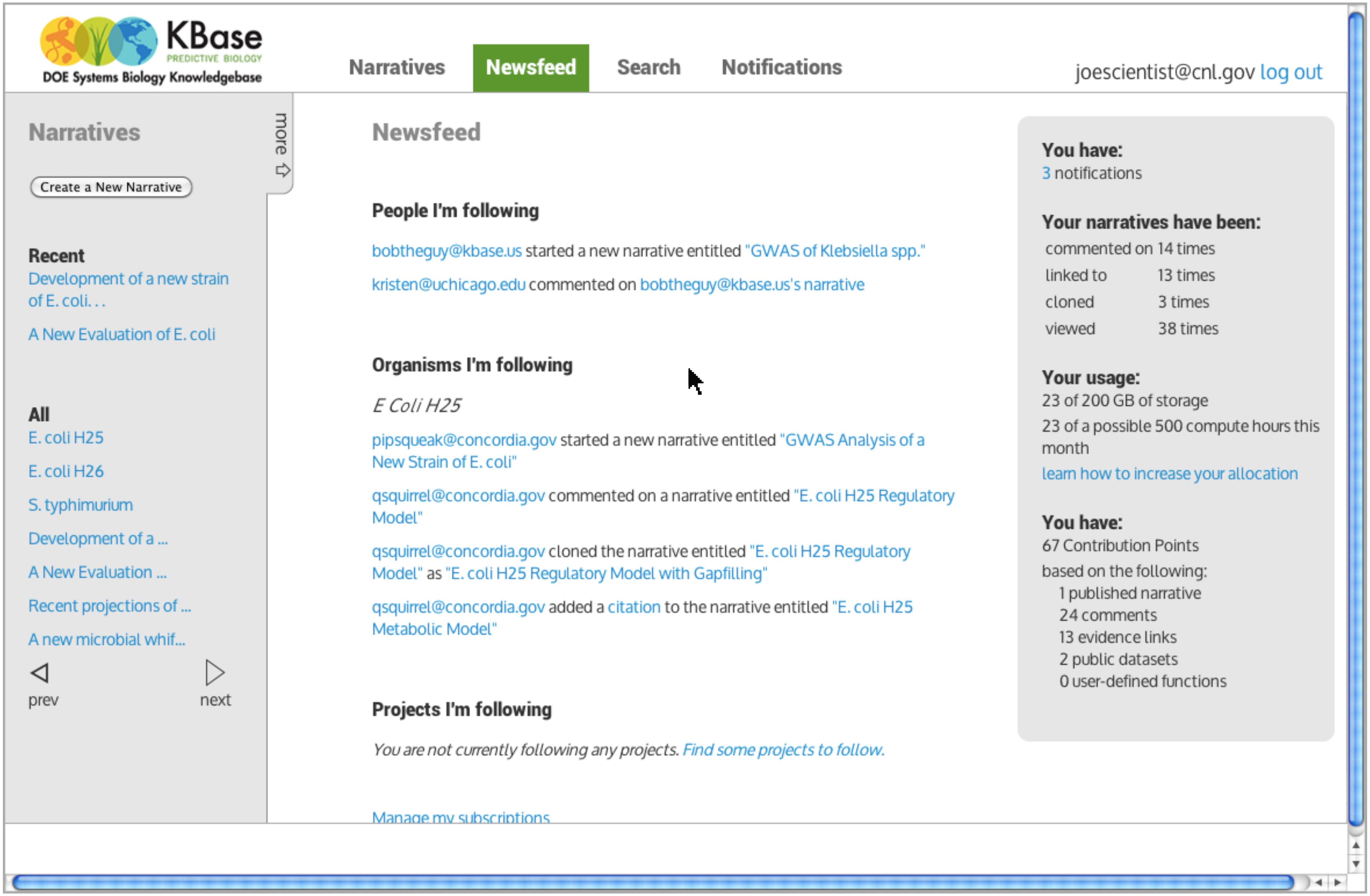
You have:
3 notifications

Your narratives have been:

commented on	14 times
linked to	13 times
cloned	3 times
viewed	38 times

Your usage:
23 of 200 GB of storage
23 of a possible 500 compute hours this month
learn how to increase your allocation

You have:
67 Contribution Points
based on the following:
  1 published narrative
  24 comments
  13 evidence links
  2 public datasets
  0 user-defined functions



joescientist@cnl.gov log out

Narratives

Recent

Development of a new strain of E. coli. . .

A New Evaluation of E. coli

All

- E. coli H25
- E. coli H26
- S. typhimurium
- Development of a ...
- A New Evaluation ...
- Recent projections of ...
- A new microbial whif...

Newsfeed

People I'm following

bobtheguy@kbase.us started a new narrative entitled "GWAS of Klebsiella spp."

kristen@uchicago.edu commented on bobtheguy@kbase.us's narrative

Organisms I'm following

E Coli H25

pipsqueak@concordia.gov started a new narrative entitled "GWAS Analysis of a New Strain of E. coli"

qsquirrel@concordia.gov commented on a narrative entitled "E. coli H25 Regulatory Model"

qsquirrel@concordia.gov cloned the narrative entitled "E. coli H25 Regulatory Model" as "E. coli H25 Regulatory Model with Gapfilling"

qsquirrel@concordia.gov added a citation to the narrative entitled "E. coli H25 Metabolic Model"

Projects I'm following

You are not currently following any projects. Find some projects to follow.

Manage my subscriptions

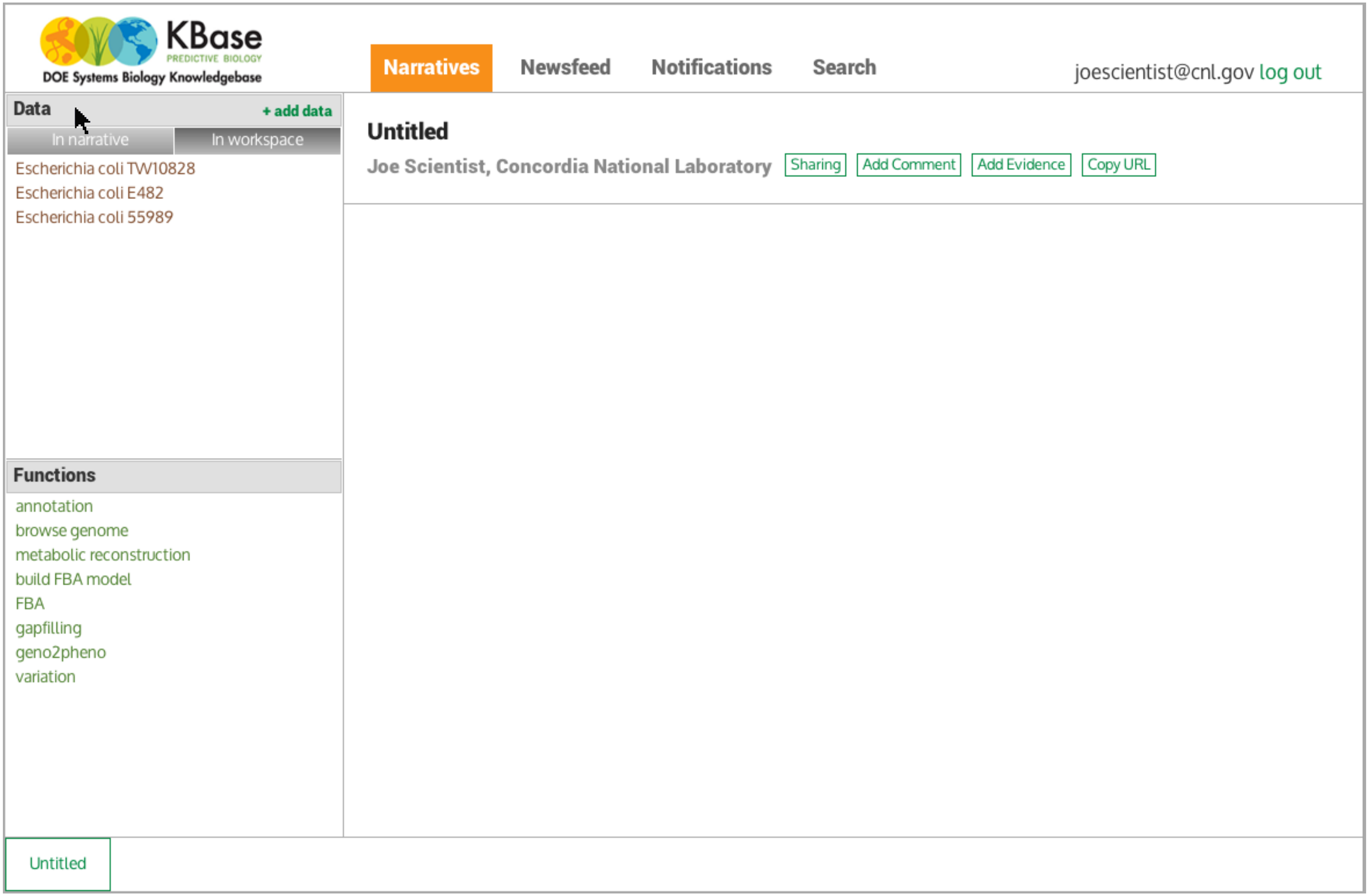
You have:
3 notifications

Your narratives have been:

commented on	14 times
linked to	13 times
cloned	3 times
viewed	38 times

Your usage:
23 of 200 GB of storage
23 of a possible 500 compute hours this month
learn how to increase your allocation

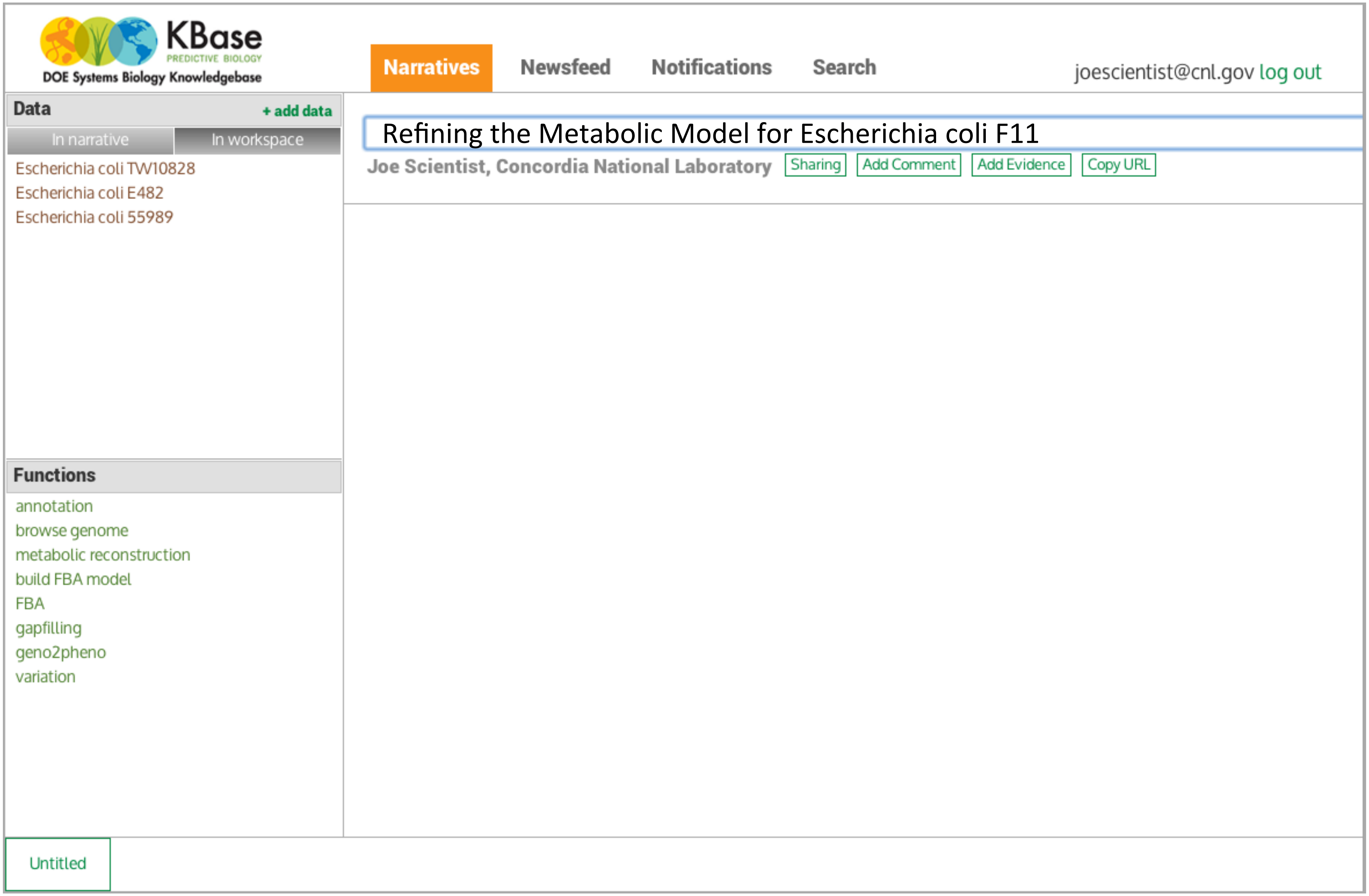
You have:
67 Contribution Points
based on the following:
- 1 published narrative
- 24 comments
- 13 evidence links
- 2 public datasets
- 0 user-defined functions



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989

Untitled

Joe Scientist, Concordia National Laboratory



Refining the Metabolic Model for Escherichia coli F11

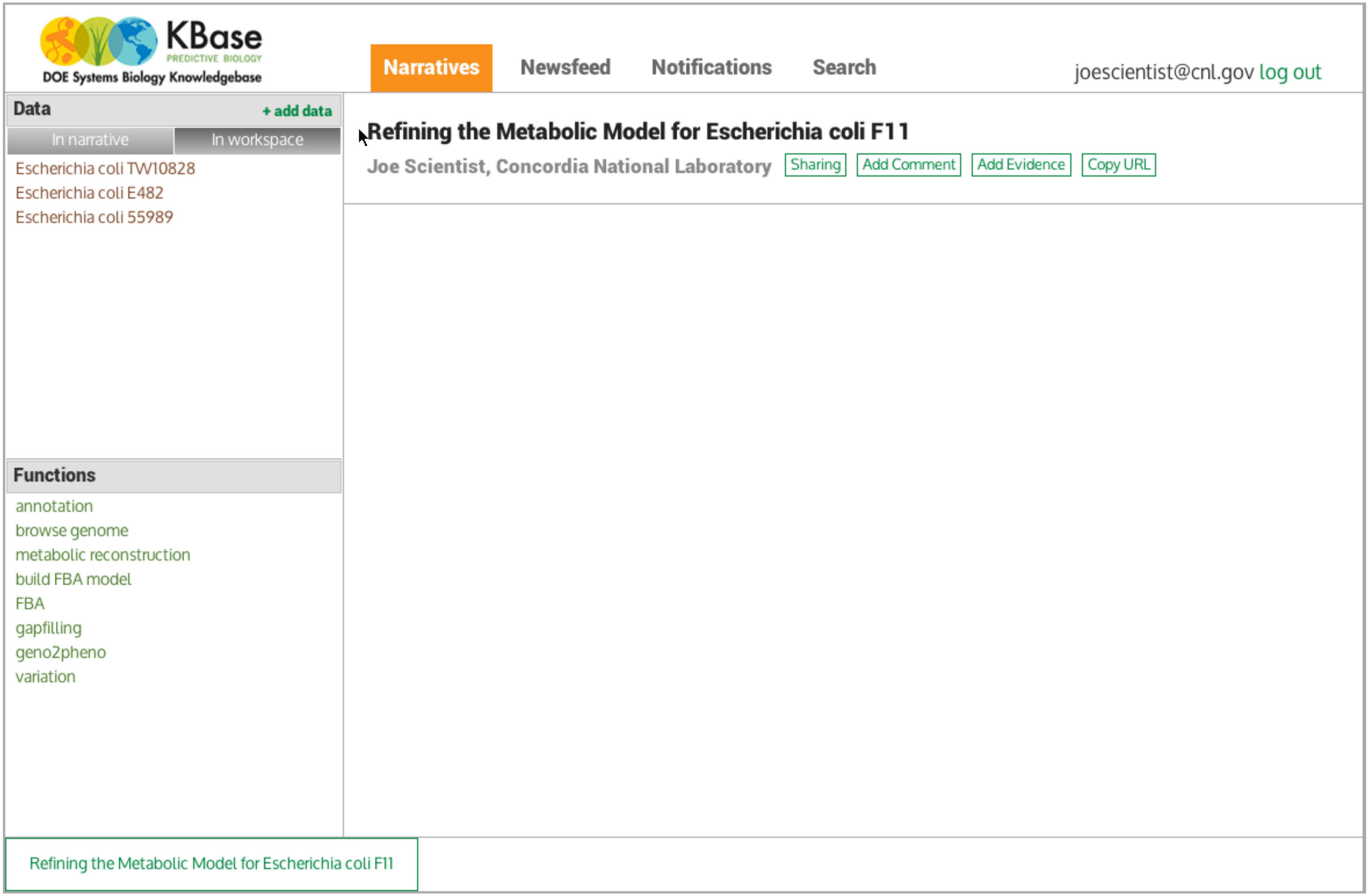
Joe Scientist, Concordia National Laboratory

Data

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989

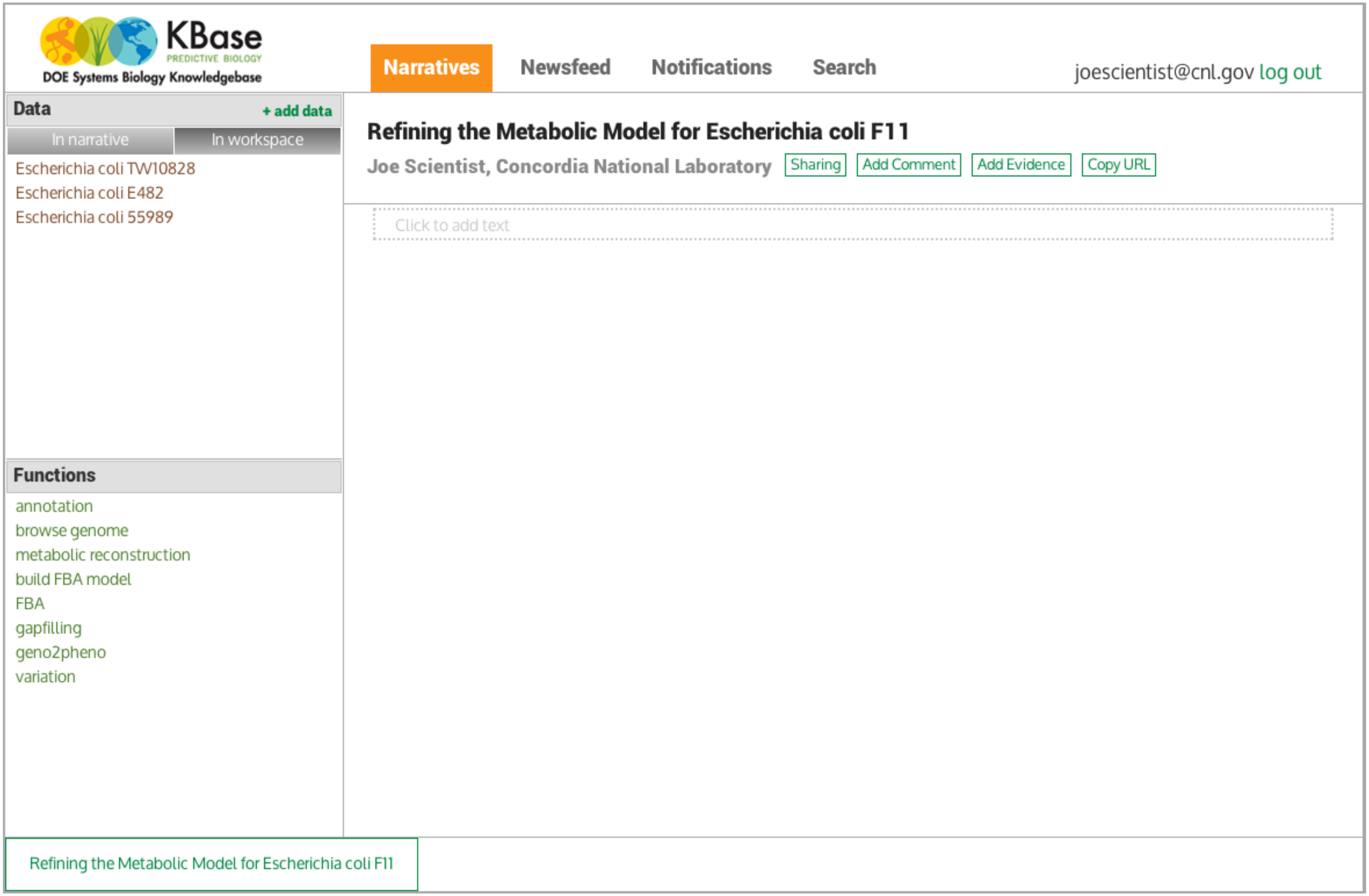
Functions

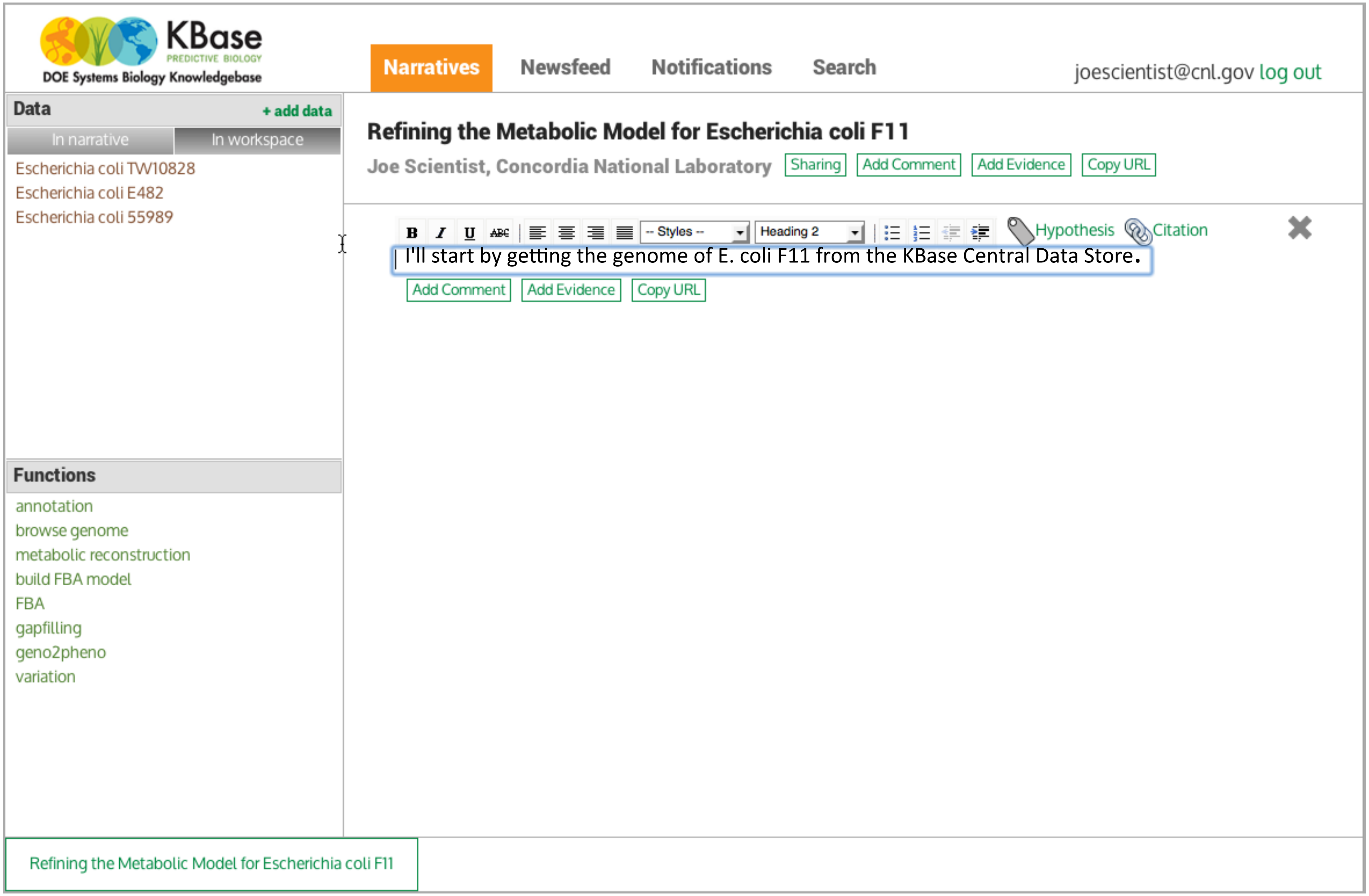
- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory


KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase
Narratives
Newsfeed
Notifications
Search
joescientist@cnl.gov log out
Data
+ add data
In narrative
In workspace
Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Functions
annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation
Refining the Metabolic Model for Escherichia coli F11
Joe Scientist, Concordia National Laboratory
Sharing
Add Comment
Add Evidence
Copy URL
Click to add text
Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

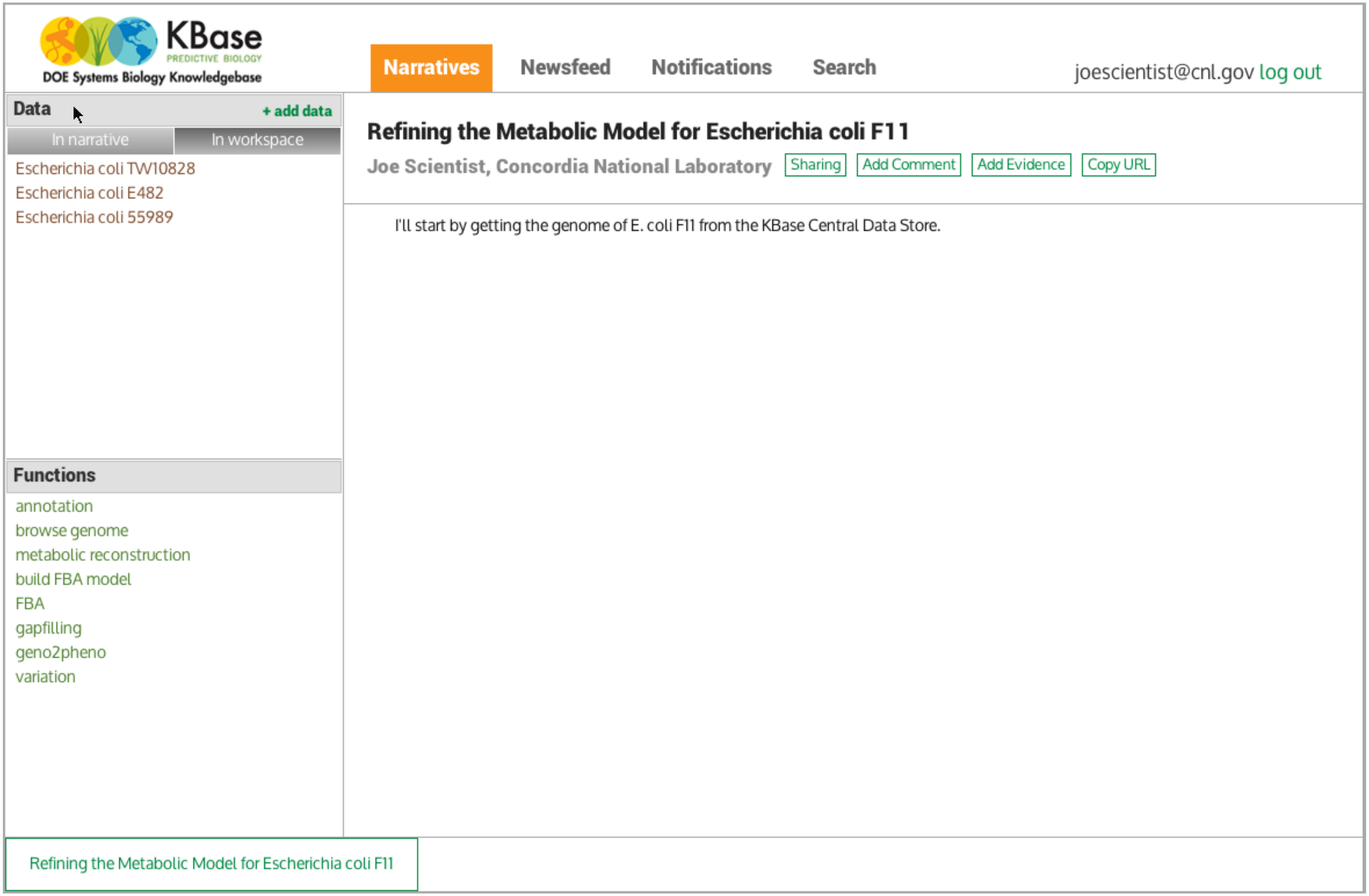
Data

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation

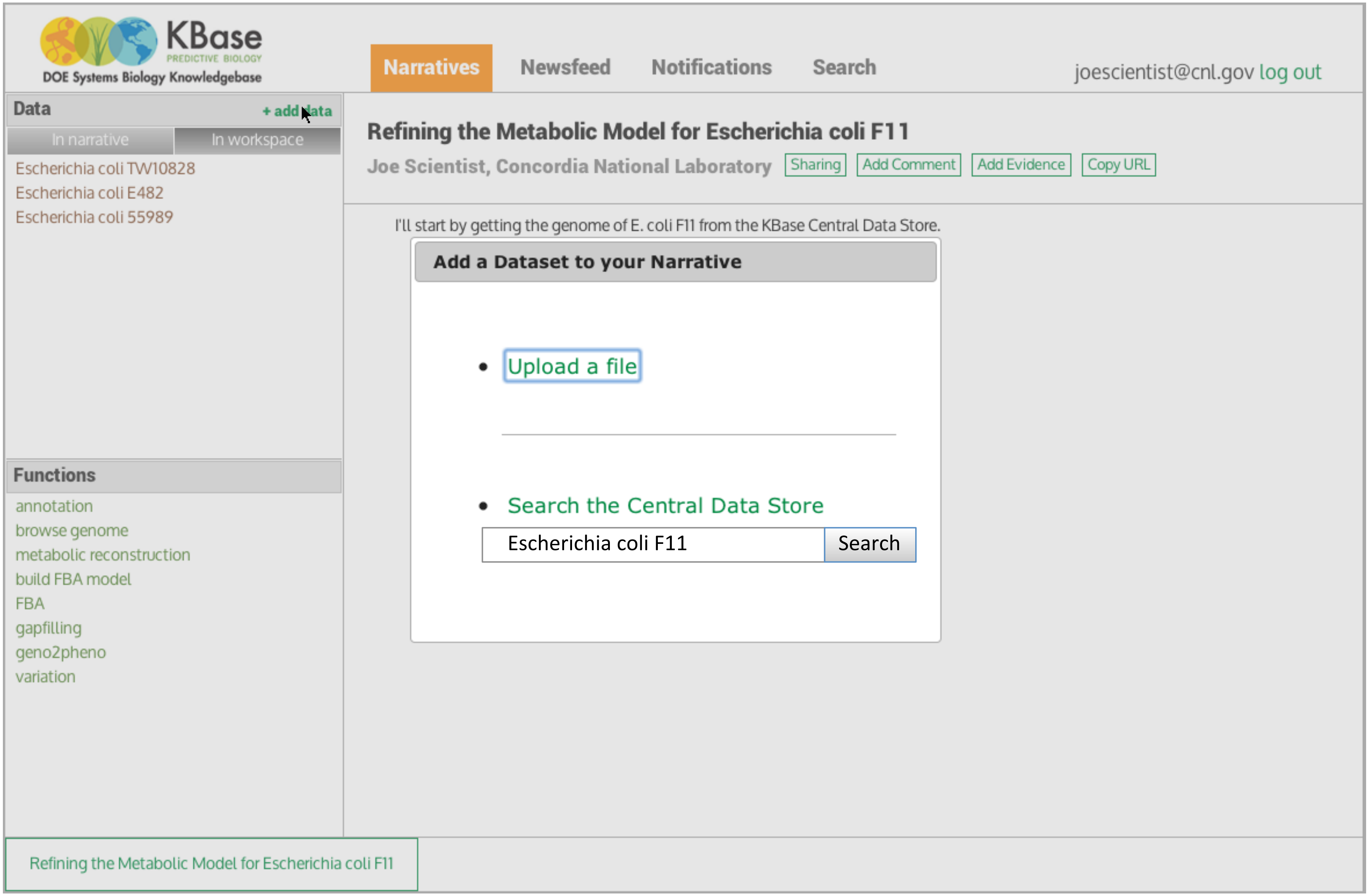
Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.



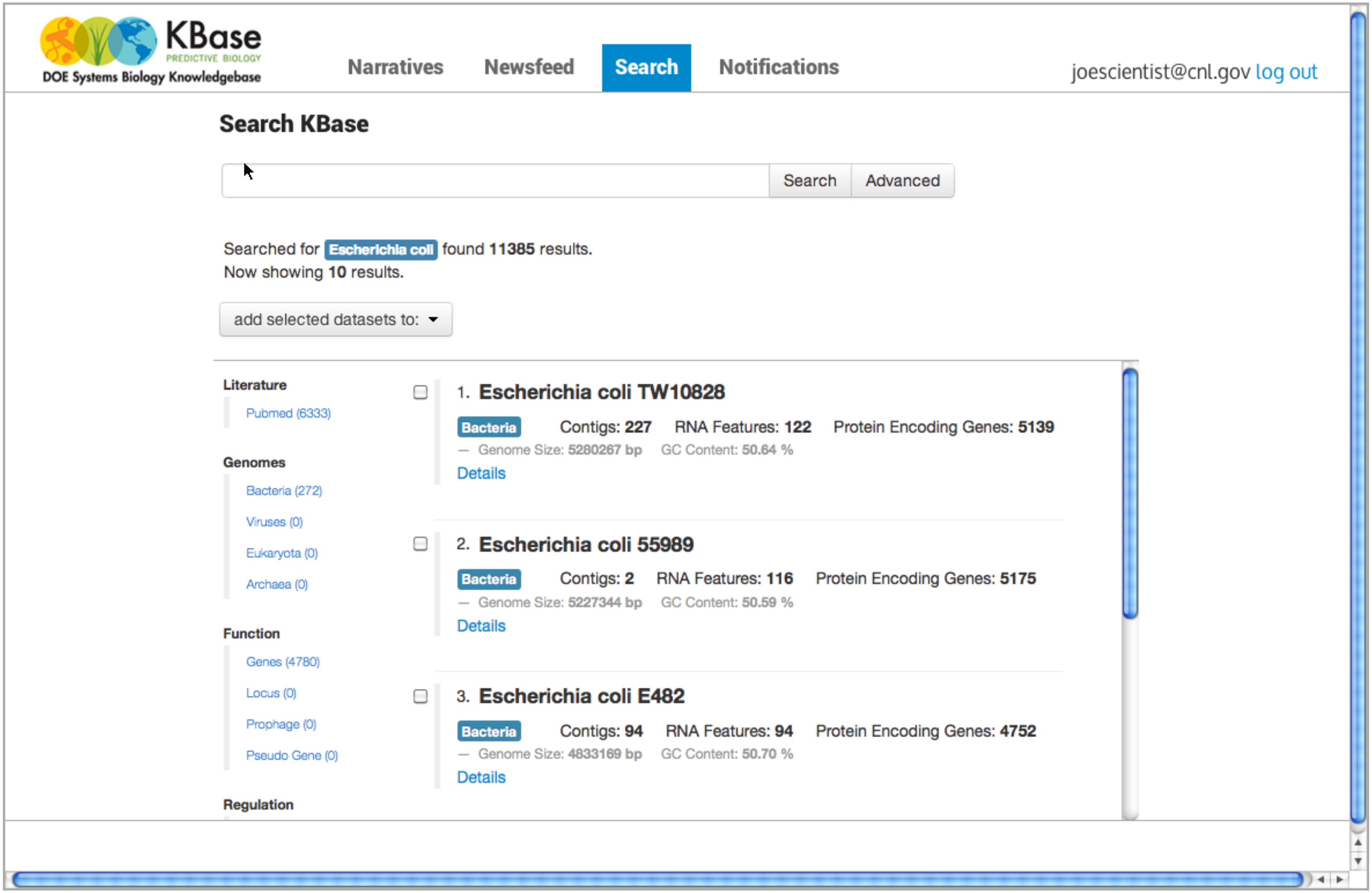
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Add a Dataset to your Narrative

Escherichia coli F11



Search KBase

Searched for **Escherichia coli** found **11385** results.
Now showing **10** results.

Literature
- Pubmed (6333)

Genomes
- Bacteria (272)
- Viruses (0)
- Eukaryota (0)
- Archaea (0)

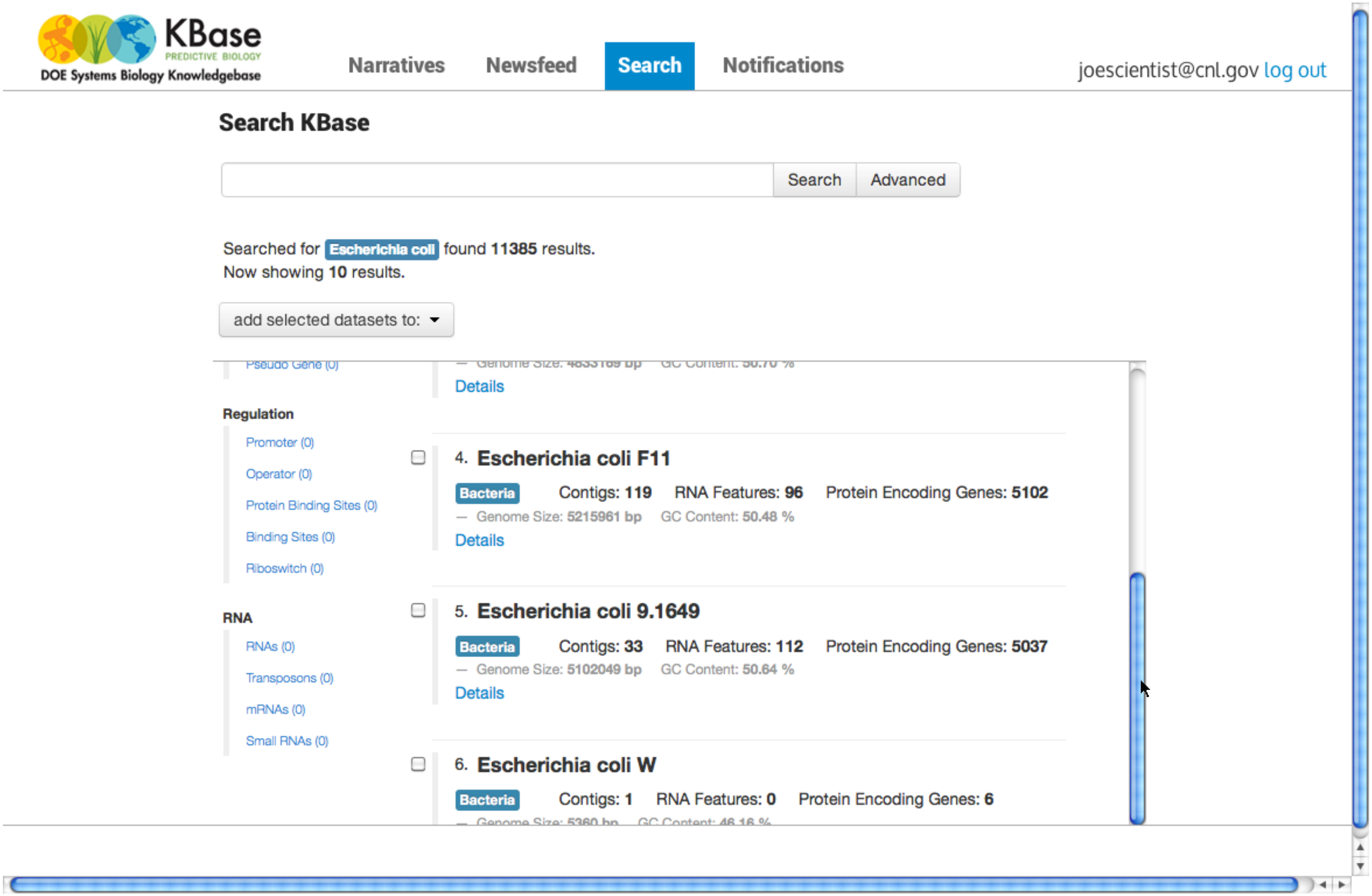
Function
- Genes (4780)
- Locus (0)
- Prophage (0)
- Pseudo Gene (0)

Regulation

☐ 1. **Escherichia coli TW10828**
Bacteria Contigs: **227** RNA Features: **122** Protein Encoding Genes: **5139**
— Genome Size: **5280267 bp** GC Content: **50.64 %**
Details

☐ 2. **Escherichia coli 55989**
Bacteria Contigs: **2** RNA Features: **116** Protein Encoding Genes: **5175**
— Genome Size: **5227344 bp** GC Content: **50.59 %**
Details

☐ 3. **Escherichia coli E482**
Bacteria Contigs: **94** RNA Features: **94** Protein Encoding Genes: **4752**
— Genome Size: **4833169 bp** GC Content: **50.70 %**
Details



Search KBase

Searched for **Escherichia coli** found **11385** results.
Now showing **10** results.

Regulation

- Promoter (0)
- Operator (0)
- Protein Binding Sites (0)
- Binding Sites (0)
- Riboswitch (0)

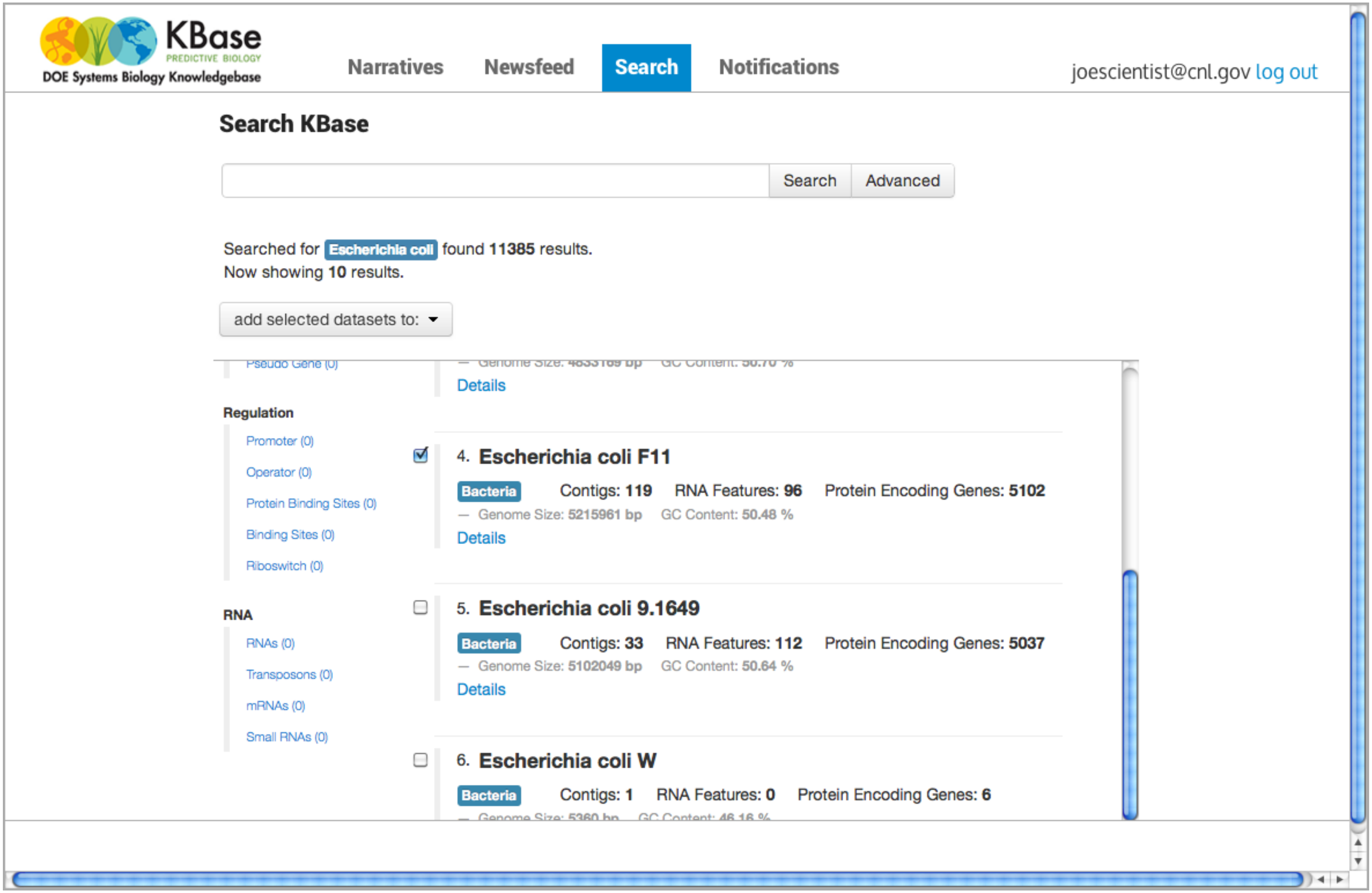
RNA

- RNAs (0)
- Transposons (0)
- mRNAs (0)
- Small RNAs (0)

☐ 4. **Escherichia coli F11**
Bacteria Contigs: **119** RNA Features: **96** Protein Encoding Genes: **5102**
— Genome Size: **5215961 bp** GC Content: **50.48 %**
Details

☐ 5. **Escherichia coli 9.1649**
Bacteria Contigs: **33** RNA Features: **112** Protein Encoding Genes: **5037**
— Genome Size: **5102049 bp** GC Content: **50.64 %**
Details

☐ 6. **Escherichia coli W**
Bacteria Contigs: **1** RNA Features: **0** Protein Encoding Genes: **6**



Search KBase

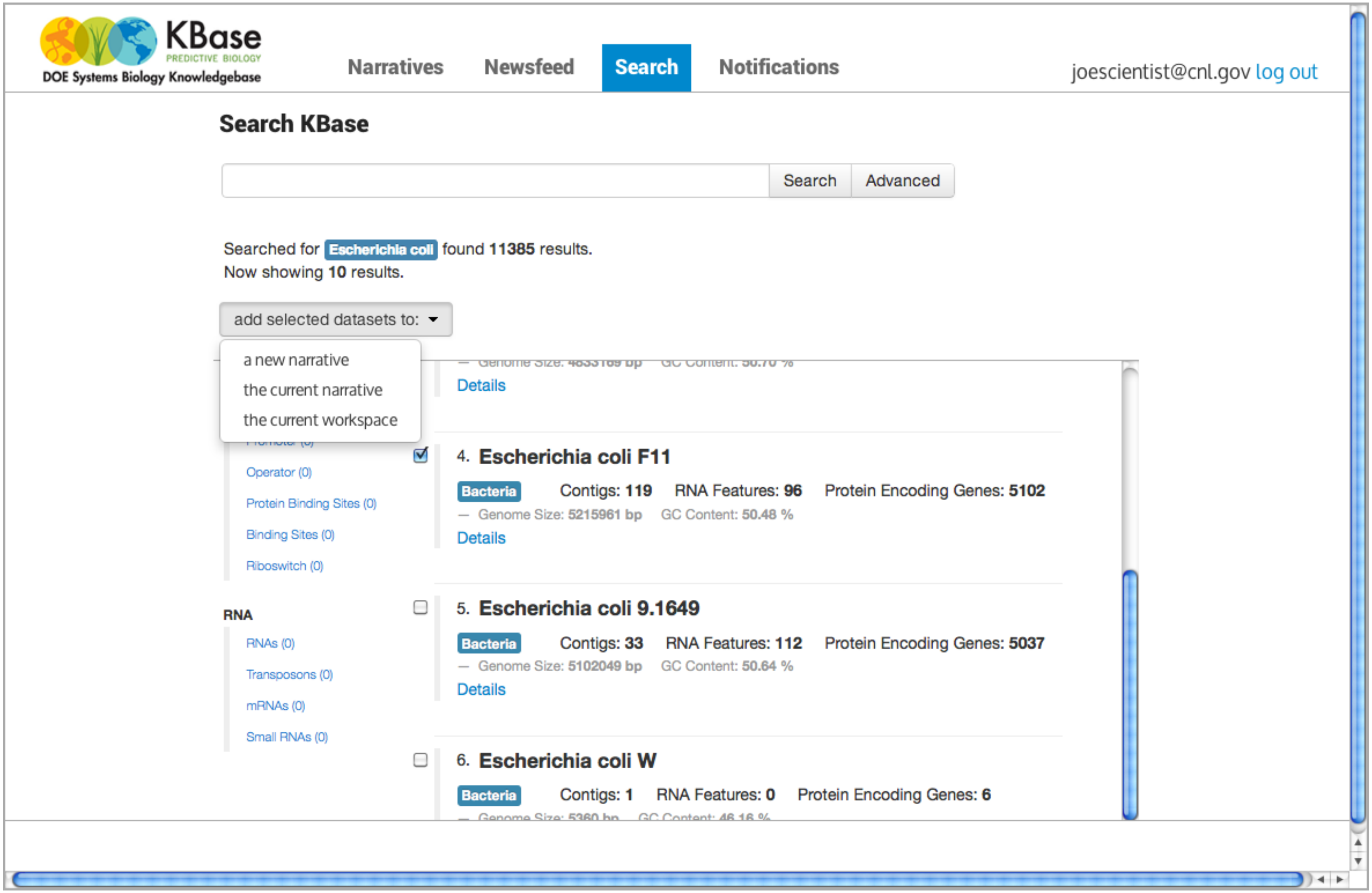
Searched for **Escherichia coli** found **11385** results.
Now showing **10** results.

— Genome Size: **4833169 bp** GC Content: **50.70 %**
Details

☑ 4. **Escherichia coli F11**
Bacteria Contigs: **119** RNA Features: **96** Protein Encoding Genes: **5102**
— Genome Size: **5215961 bp** GC Content: **50.48 %**
Details

☐ 5. **Escherichia coli 9.1649**
Bacteria Contigs: **33** RNA Features: **112** Protein Encoding Genes: **5037**
— Genome Size: **5102049 bp** GC Content: **50.64 %**
Details

☐ 6. **Escherichia coli W**
Bacteria Contigs: **1** RNA Features: **0** Protein Encoding Genes: **6**
— Genome Size: **5360 bp** GC Content: **46.16 %**



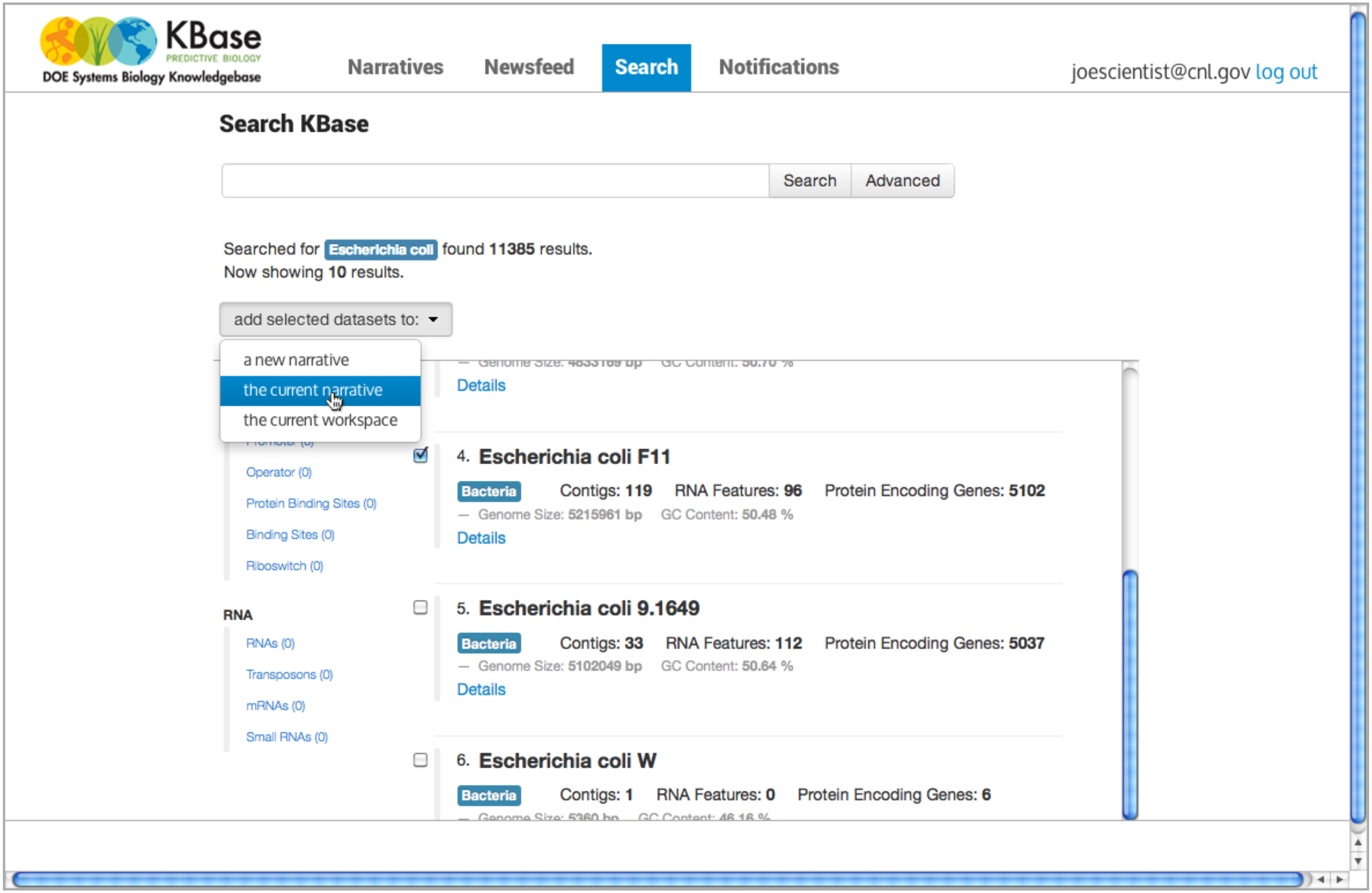
Search KBase

Searched for **Escherichia coli** found **11385** results.
Now showing **10** results.

☑ 4. **Escherichia coli F11**
Bacteria Contigs: **119** RNA Features: **96** Protein Encoding Genes: **5102**
— Genome Size: **5215961 bp** GC Content: **50.48 %**

☐ 5. **Escherichia coli 9.1649**
Bacteria Contigs: **33** RNA Features: **112** Protein Encoding Genes: **5037**
— Genome Size: **5102049 bp** GC Content: **50.64 %**

☐ 6. **Escherichia coli W**
Bacteria Contigs: **1** RNA Features: **0** Protein Encoding Genes: **6**



Search KBase

Searched for **Escherichia coli** found **11385** results.
Now showing **10** results.

Details

☑ 4. **Escherichia coli F11**

Bacteria Contigs: **119** RNA Features: **96** Protein Encoding Genes: **5102**

— Genome Size: **5215961 bp** GC Content: **50.48 %**

Details

☐ 5. **Escherichia coli 9.1649**

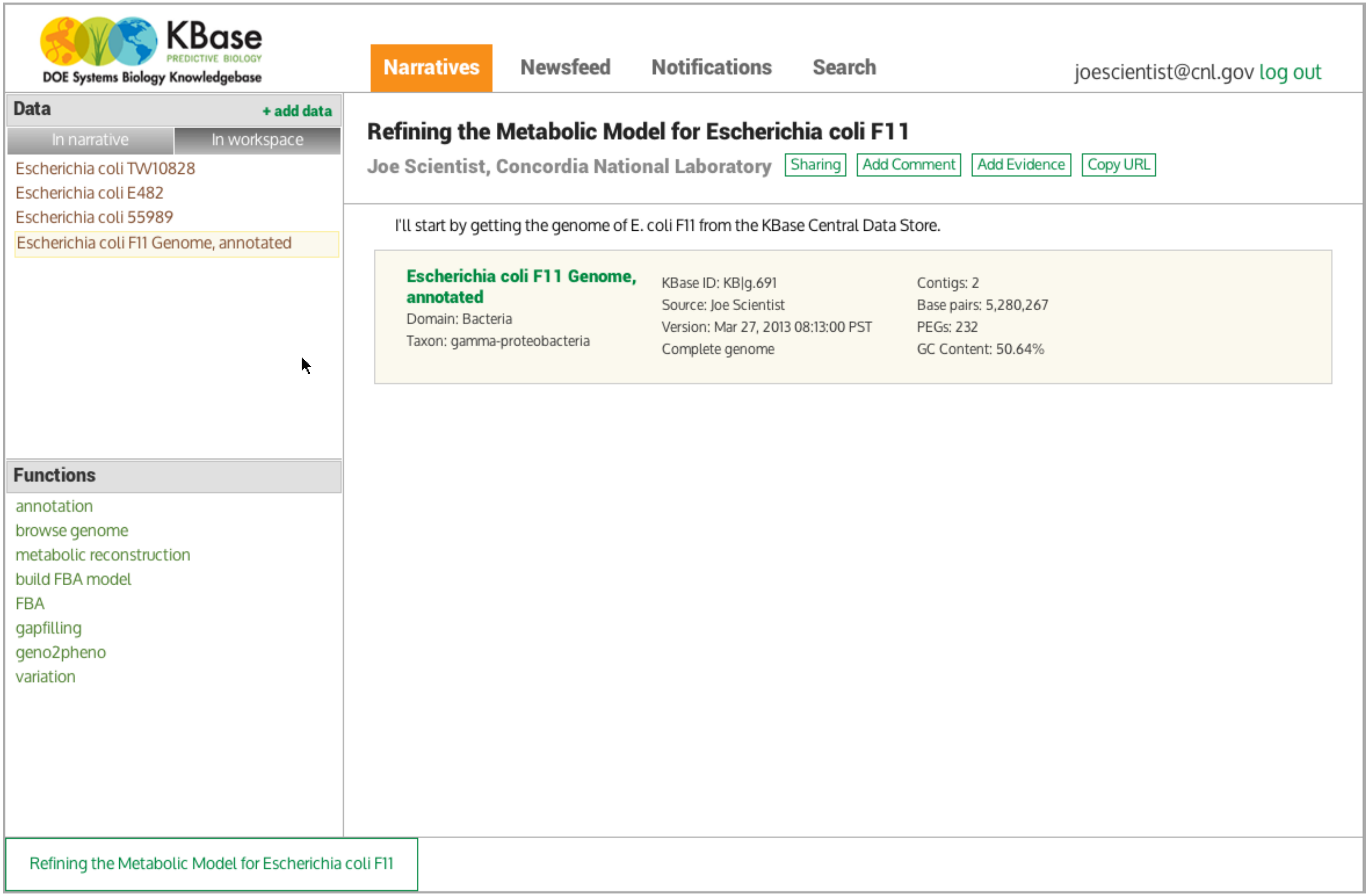
Bacteria Contigs: **33** RNA Features: **112** Protein Encoding Genes: **5037**

— Genome Size: **5102049 bp** GC Content: **50.64 %**

Details

☐ 6. **Escherichia coli W**

Bacteria Contigs: **1** RNA Features: **0** Protein Encoding Genes: **6**



Refining the Metabolic Model for Escherichia coli F11

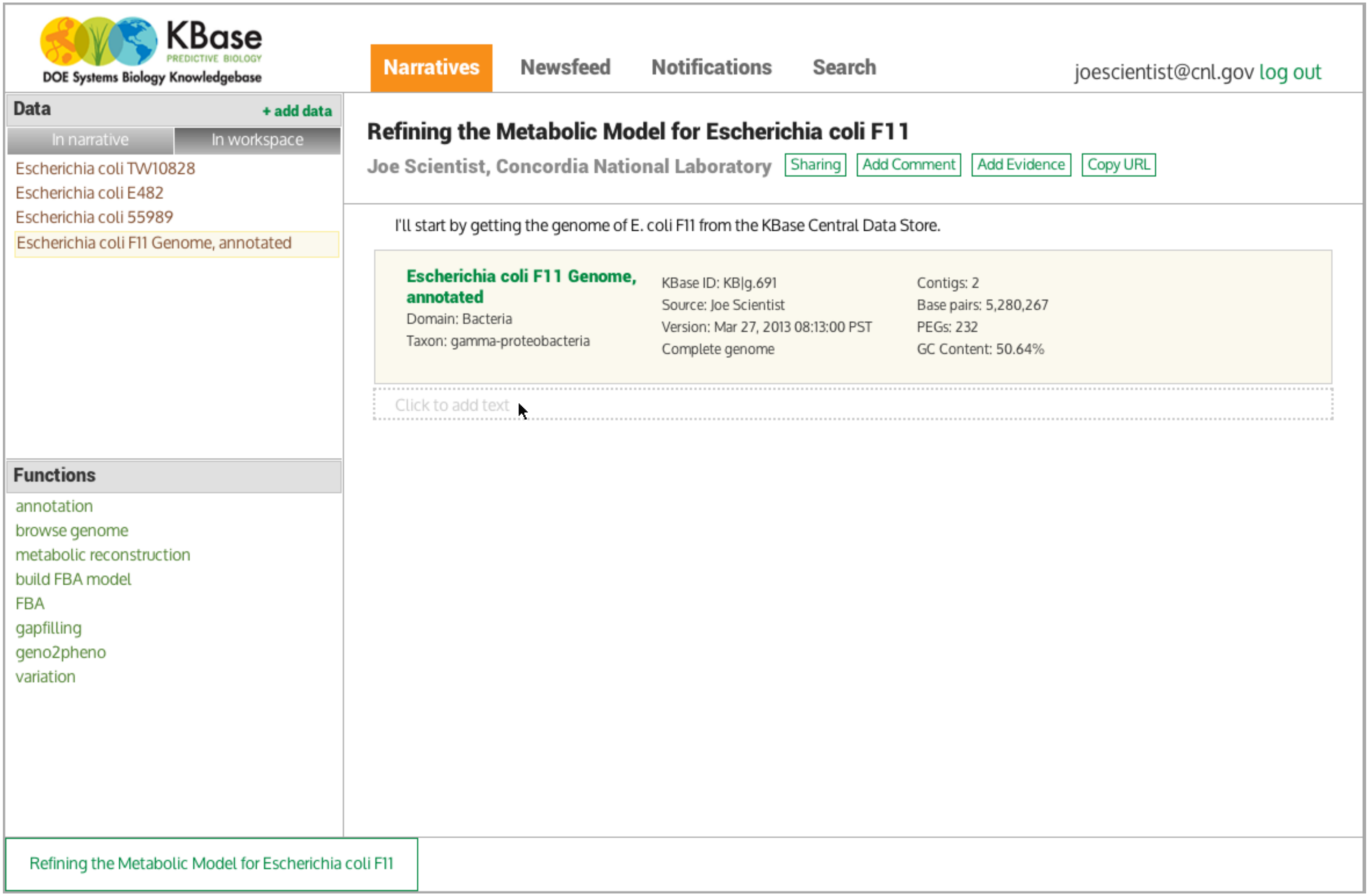
Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%



joescientist@cnl.gov

Data

Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated

Functions

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

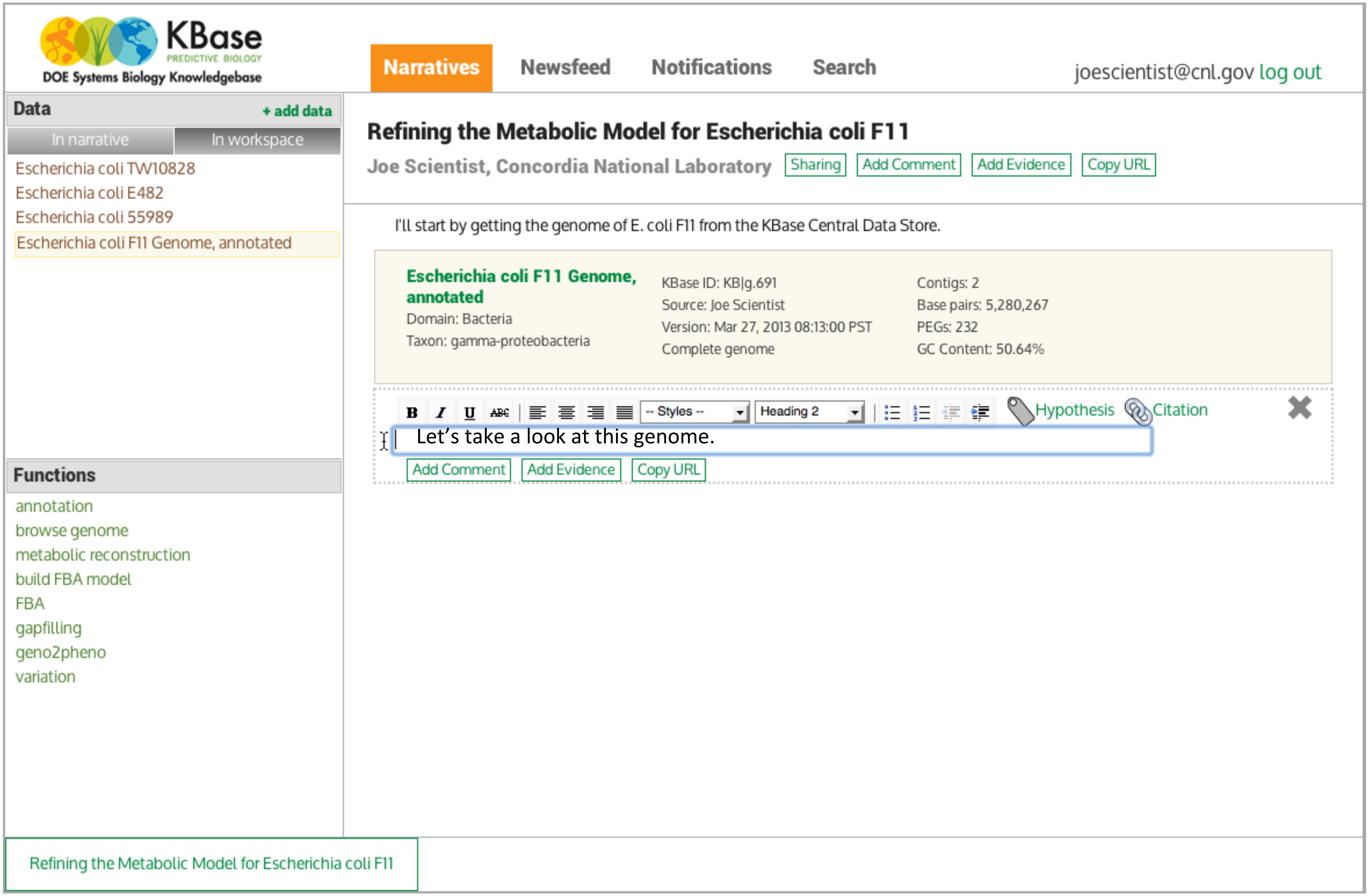
I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Refining the Metabolic Model for Escherichia coli F11



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

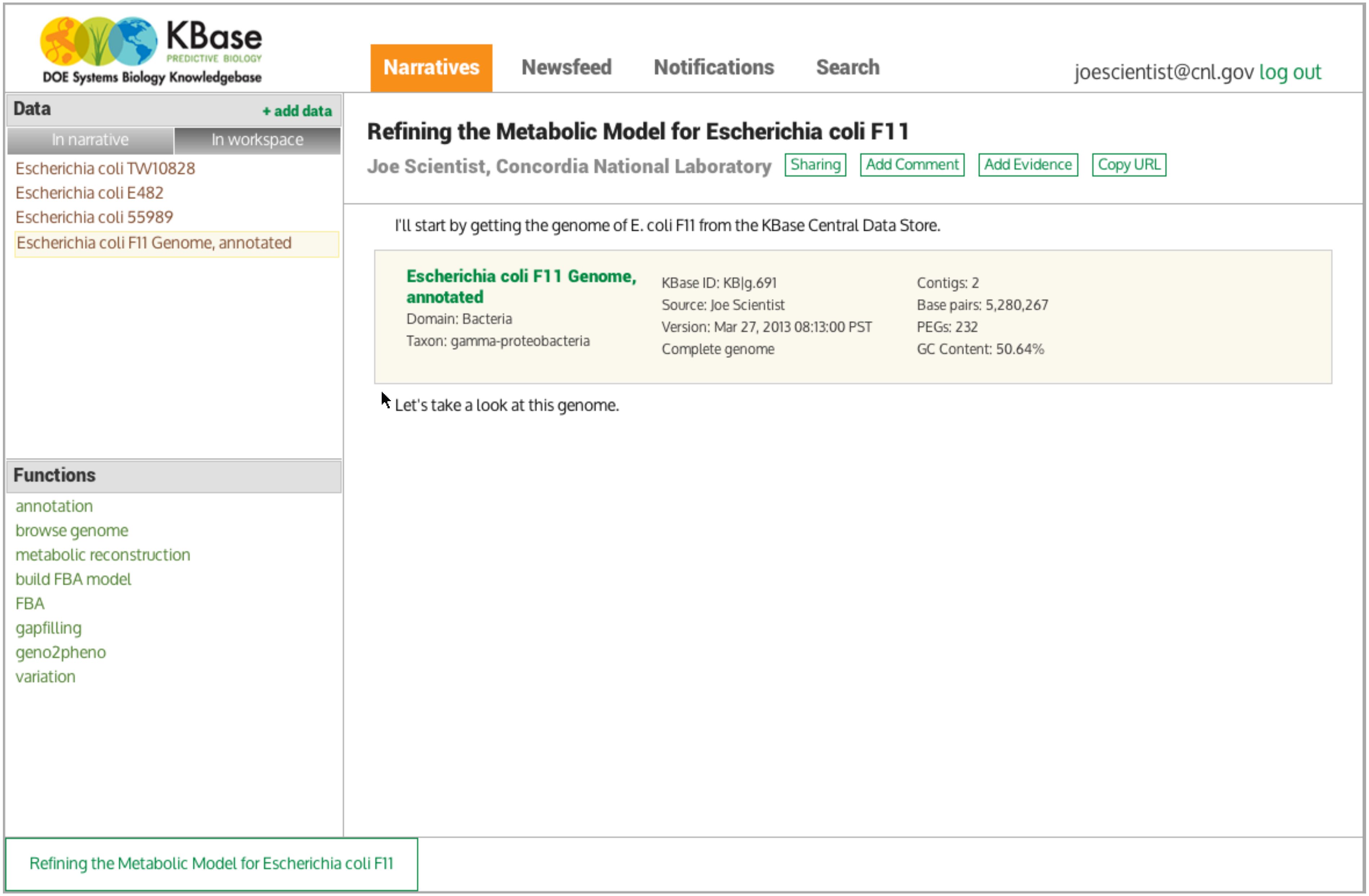
Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

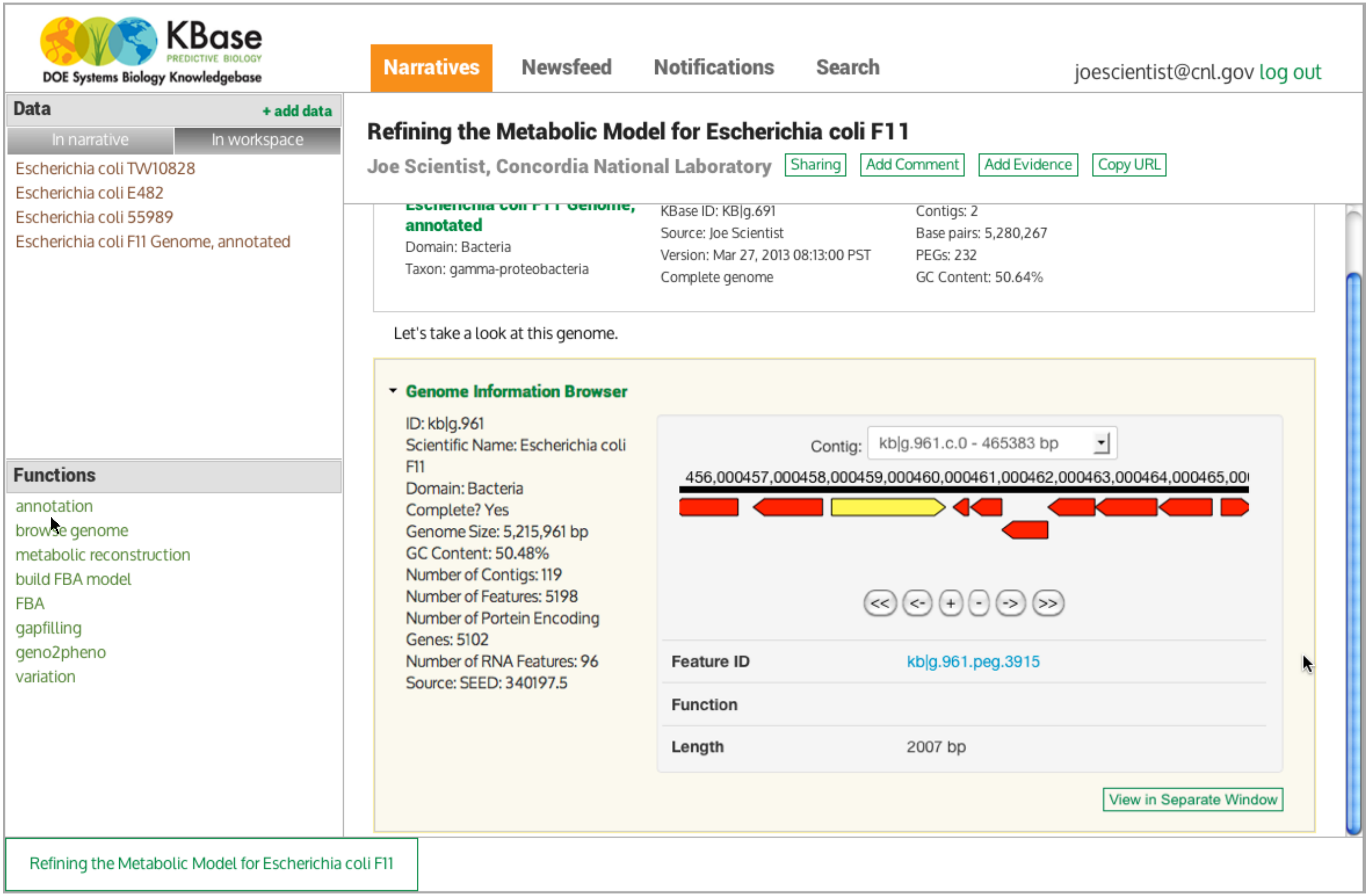
I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Genome Information Browser

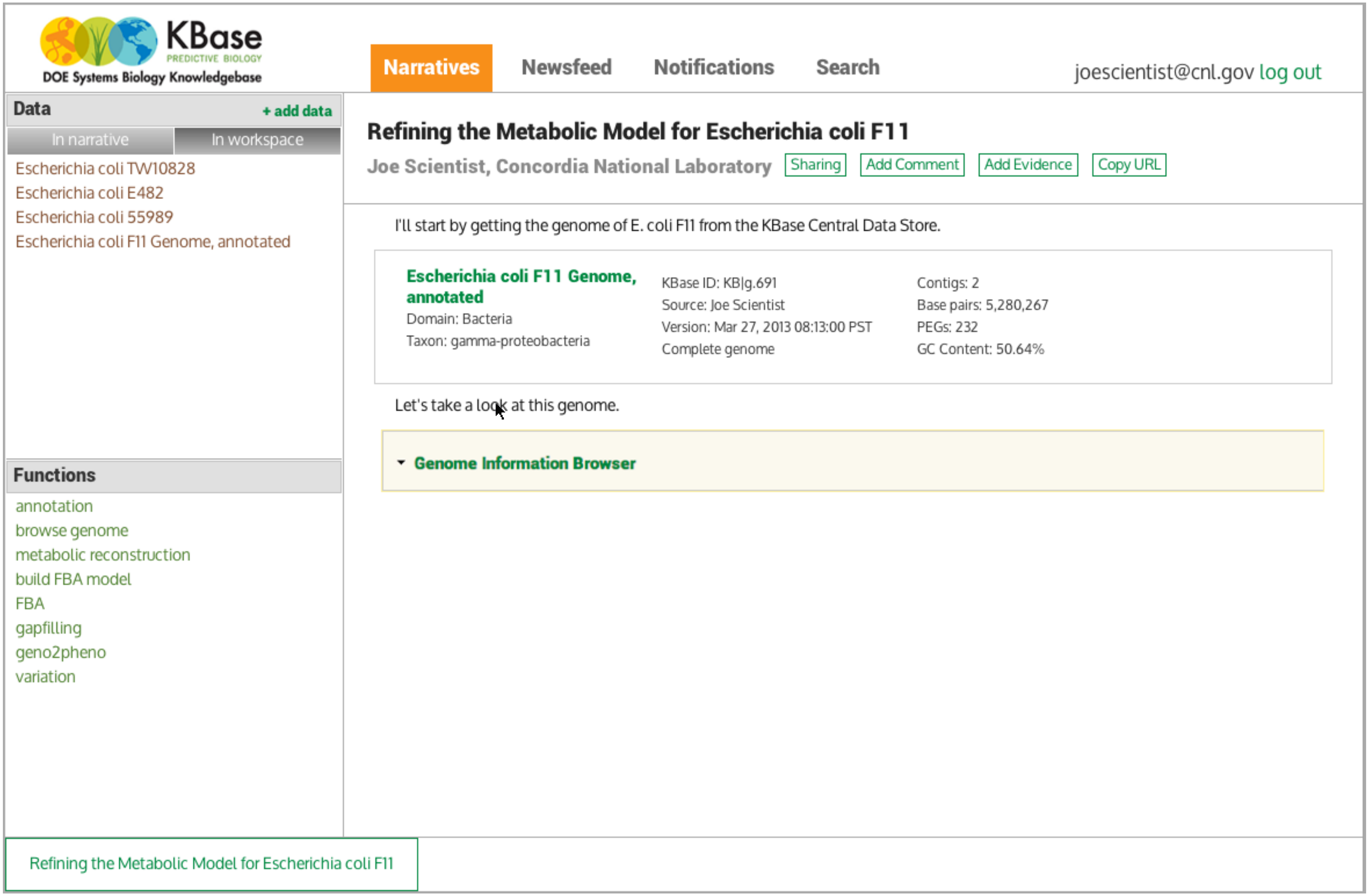
ID: kb|g.961
Scientific Name: Escherichia coli F11
Domain: Bacteria
Complete? Yes
Genome Size: 5,215,961 bp
GC Content: 50.48%
Number of Contigs: 119
Number of Features: 5198
Number of Portein Encoding Genes: 5102
Number of RNA Features: 96
Source: SEED: 340197.5

Contig: kb|g.961.c.0 - 465383 bp

456,000457,000458,000459,000460,000461,000462,000463,000464,000465,00

Feature ID	kb|g.961.peg.3915
Function	
Length	2007 bp

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

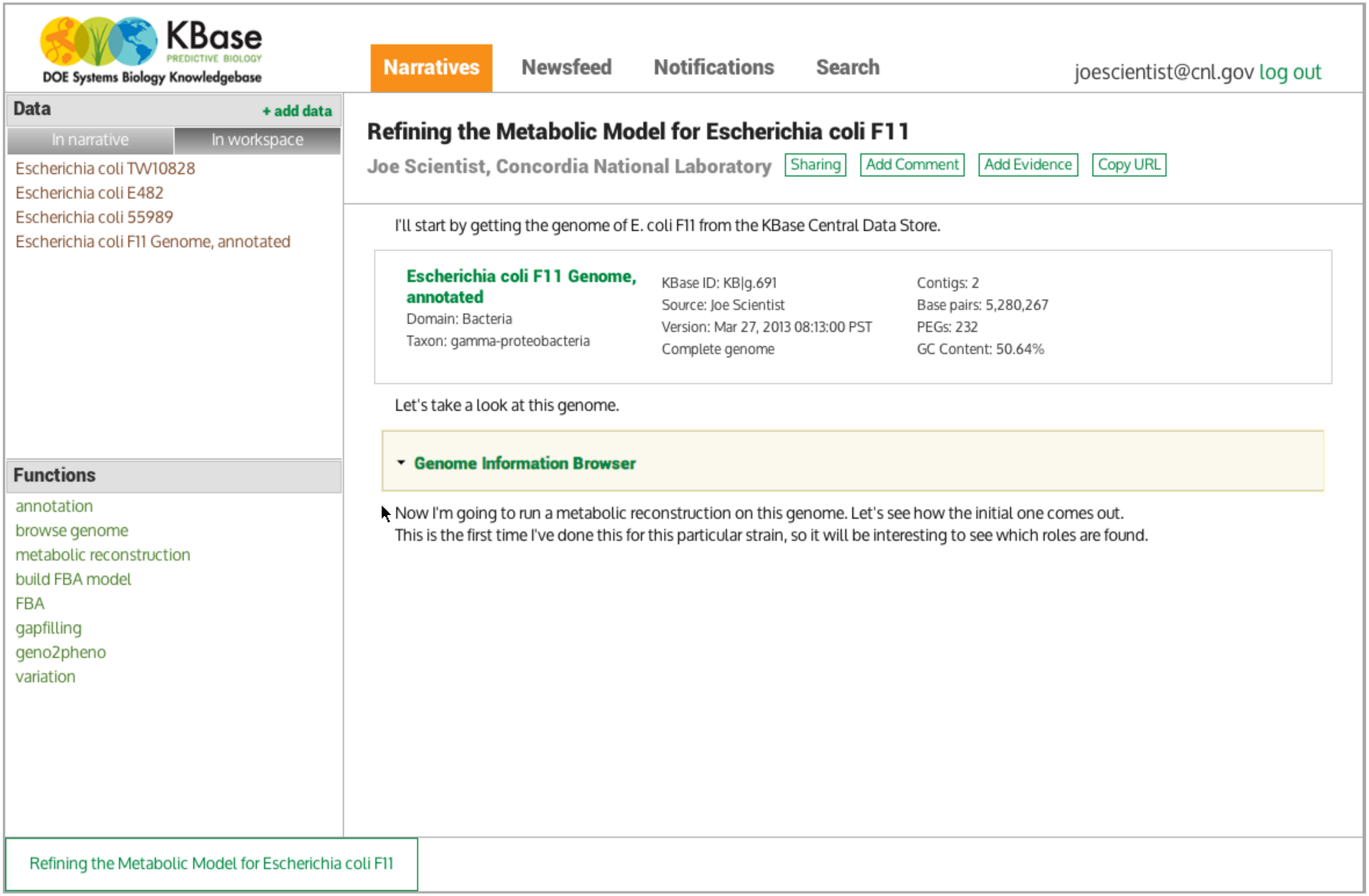
I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

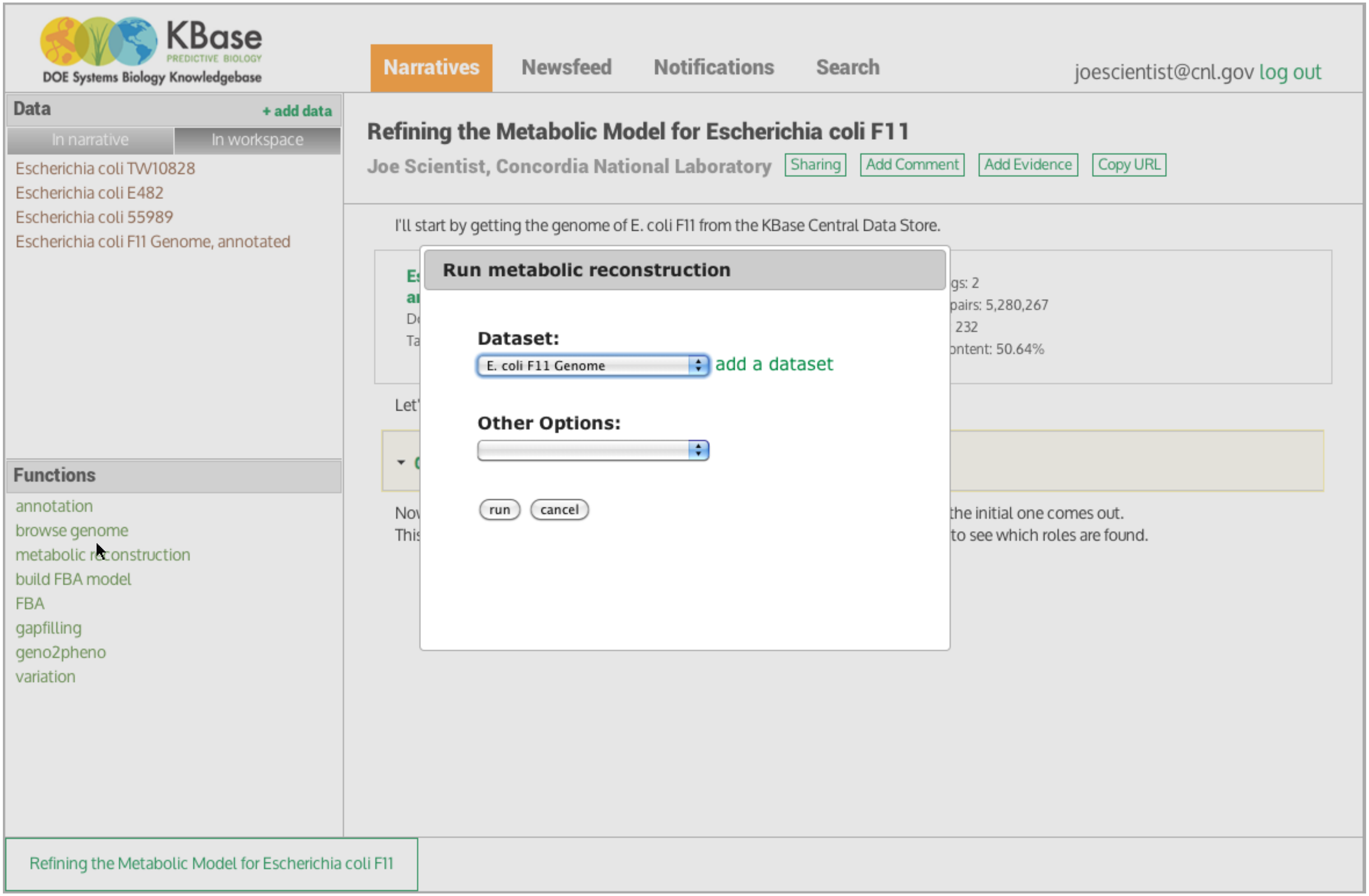
Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out. This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Es
an
De
Ta

gs: 2
pairs: 5,280,267
232
ontent: 50.64%

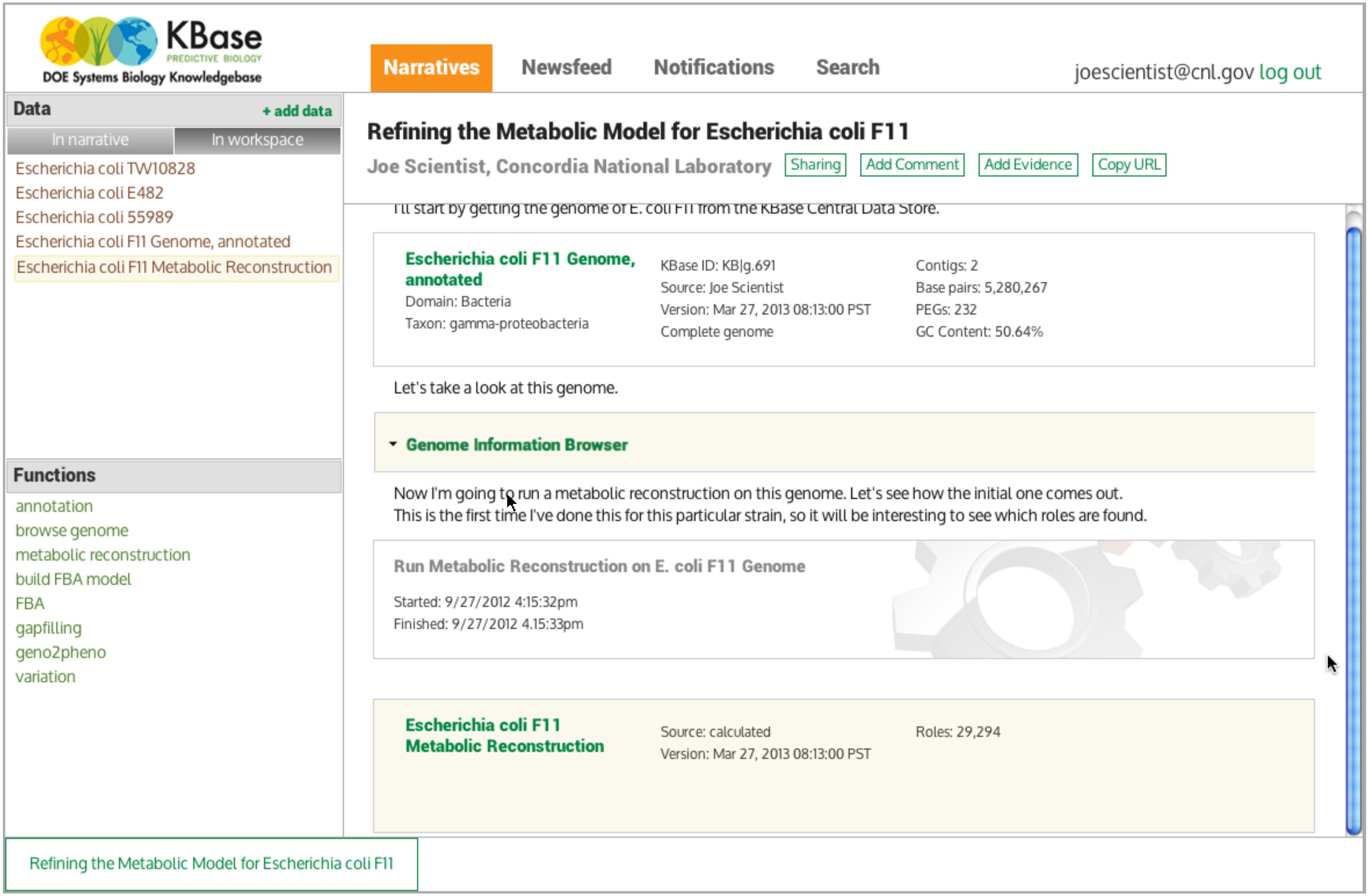
Dataset: E. coli F11 Genome

Other Options:

Let'

Nov … the initial one comes out.
This … to see which roles are found.

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Genome Information Browser

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out.
This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.

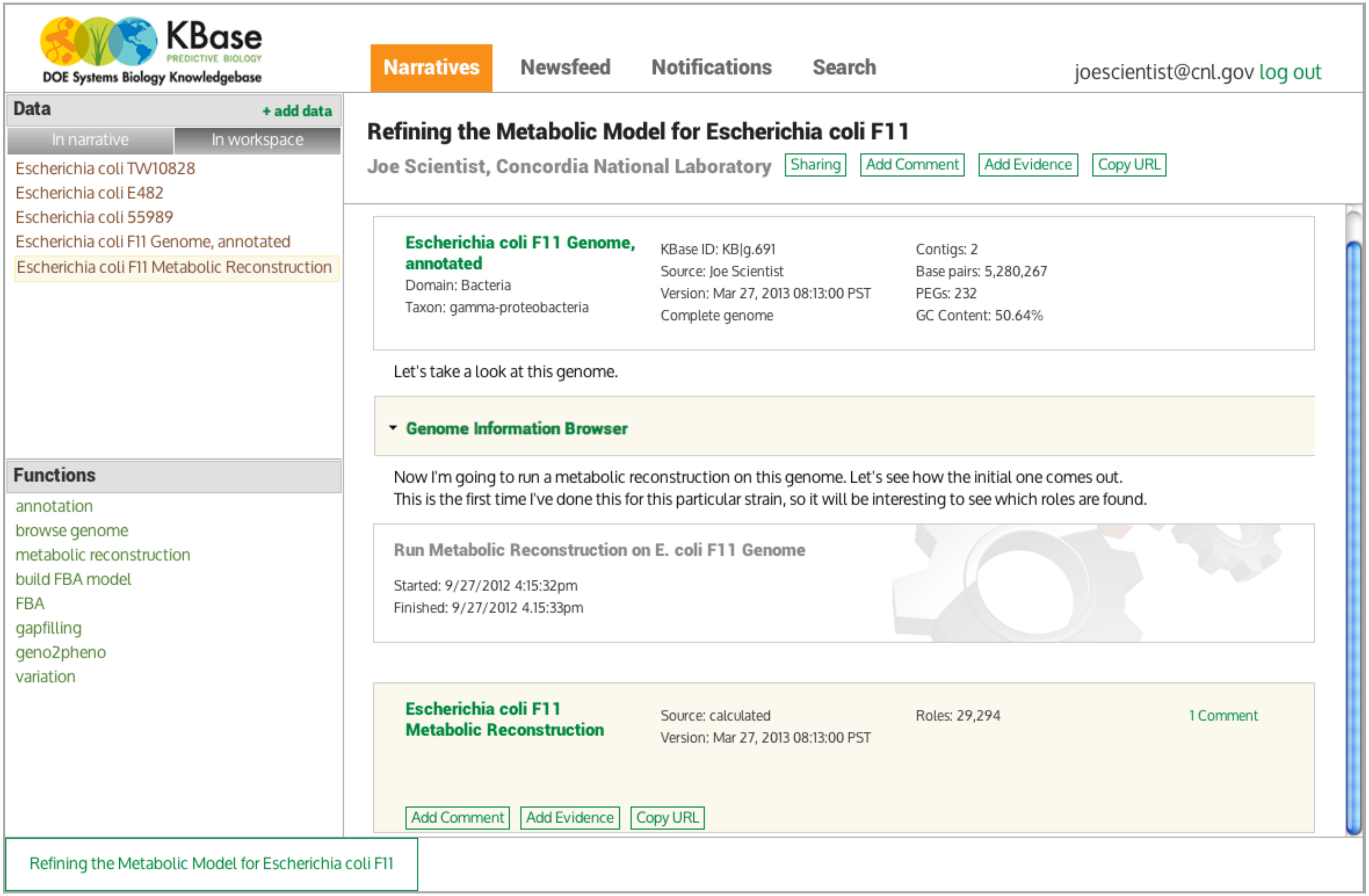
Run Metabolic Reconstruction on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 Metabolic Reconstruction

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Roles: 29,294



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out.
This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.

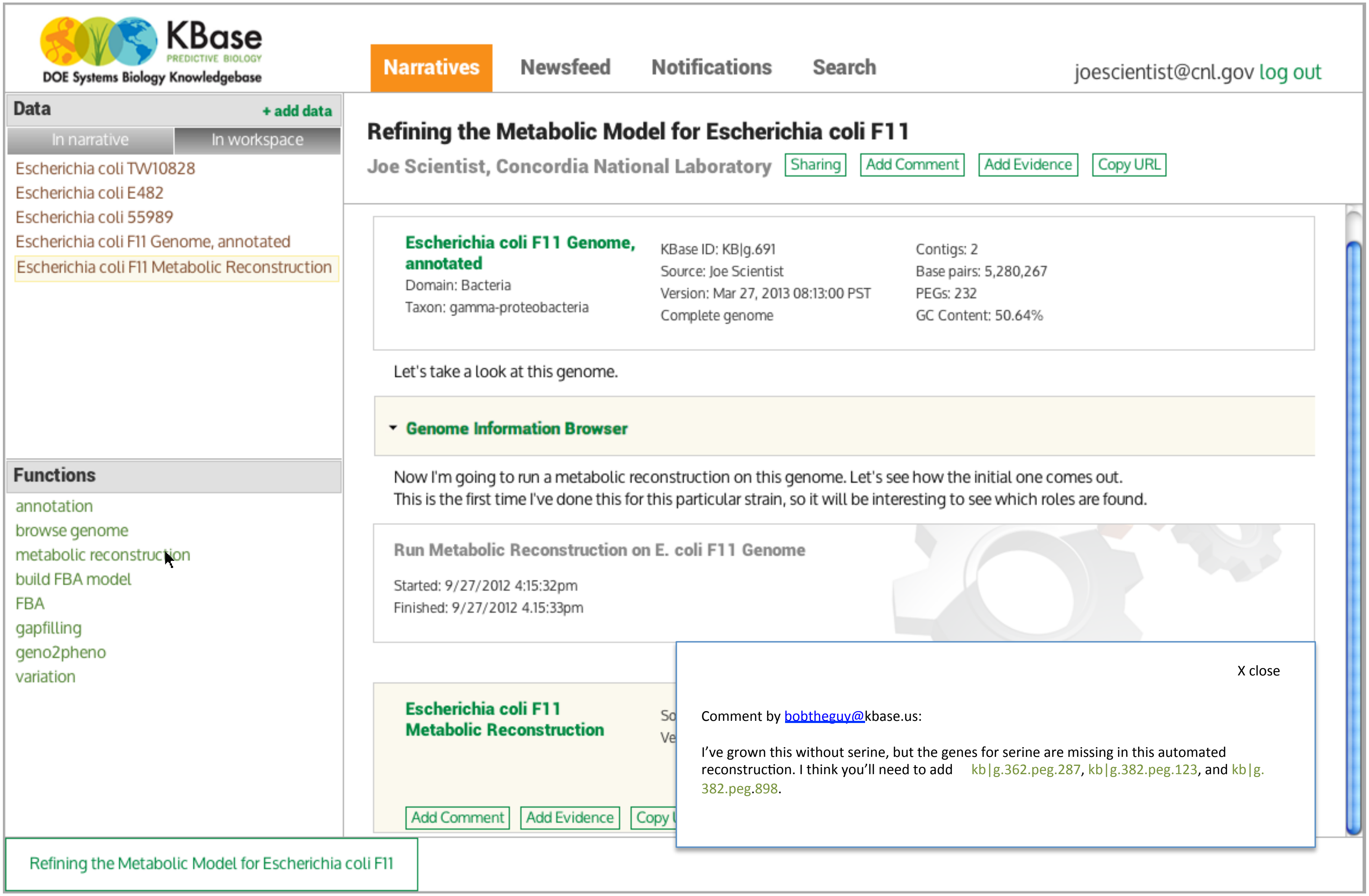
Run Metabolic Reconstruction on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 Metabolic Reconstruction

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Roles: 29,294



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out.
This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.

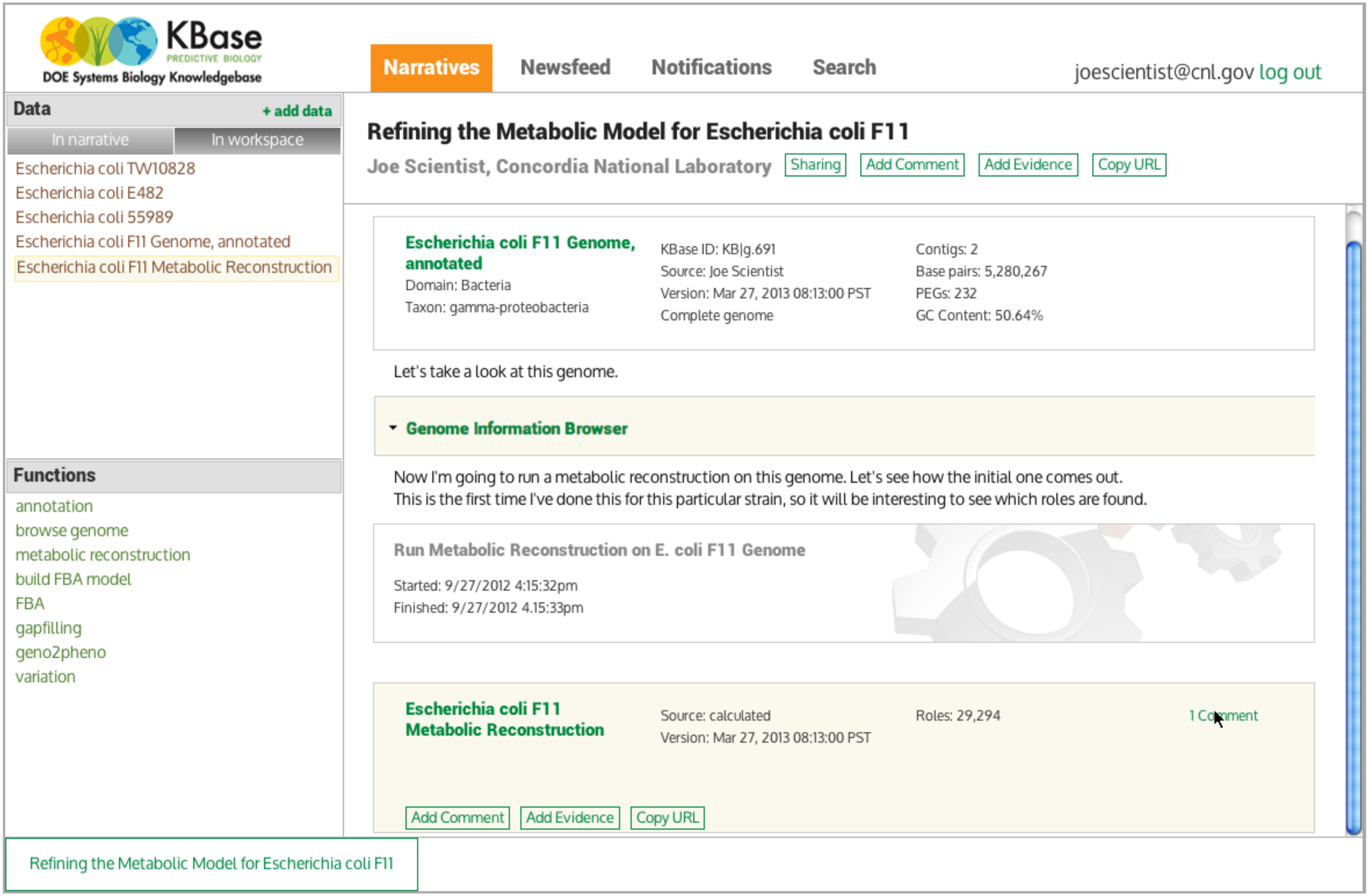
Run Metabolic Reconstruction on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 Metabolic Reconstruction

Comment by bobtheguy@kbase.us:

I've grown this without serine, but the genes for serine are missing in this automated reconstruction. I think you'll need to add kb|g.362.peg.287, kb|g.382.peg.123, and kb|g.382.peg.898.



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Escherichia coli F11 Genome, annotated
Domain: Bacteria
Taxon: gamma-proteobacteria

KBase ID: KB|g.691
Source: Joe Scientist
Version: Mar 27, 2013 08:13:00 PST
Complete genome

Contigs: 2
Base pairs: 5,280,267
PEGs: 232
GC Content: 50.64%

Let's take a look at this genome.

▾ **Genome Information Browser**

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out.
This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.

Run Metabolic Reconstruction on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

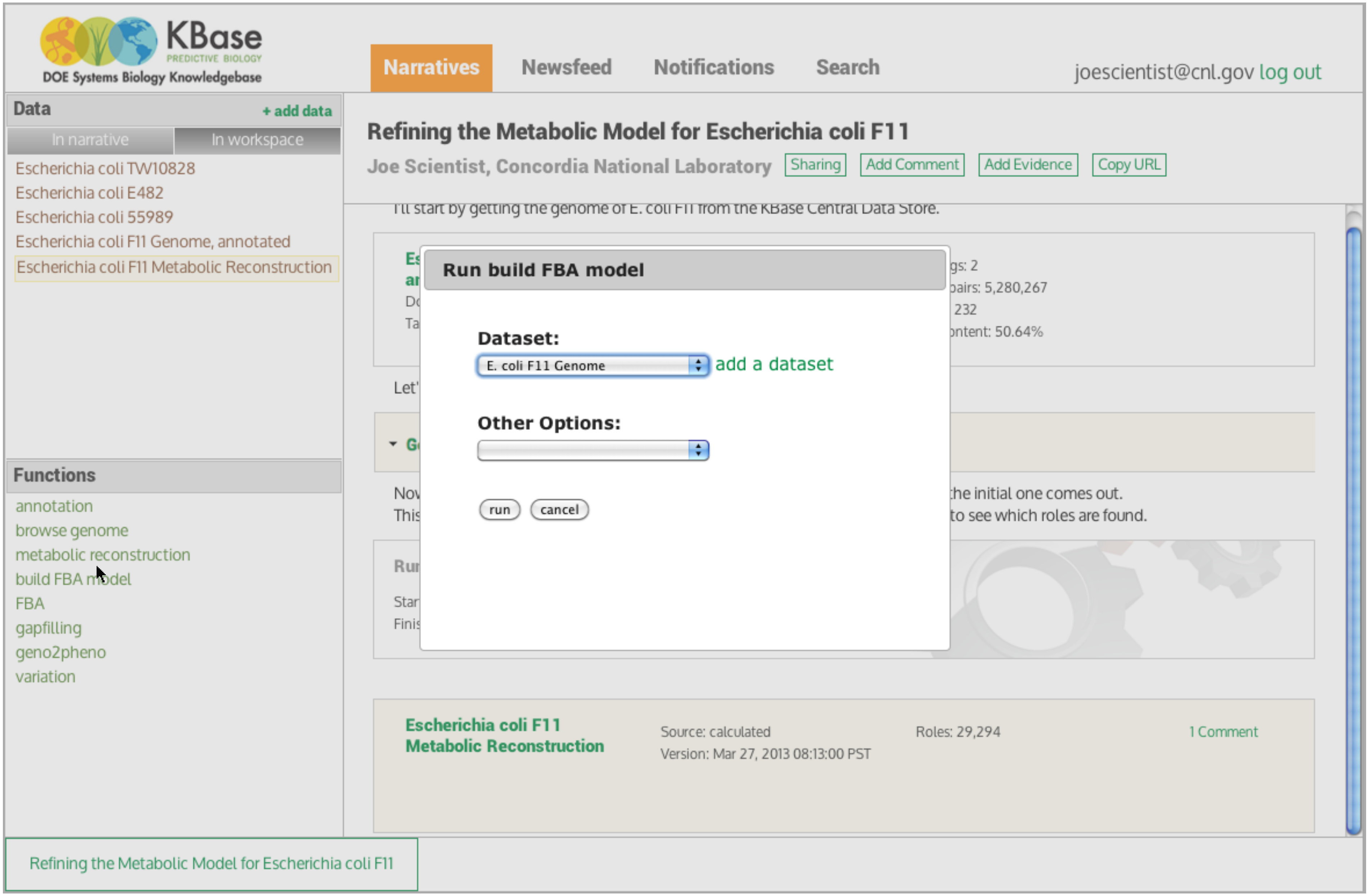
Escherichia coli F11 Metabolic Reconstruction

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Roles: 29,294

1 Comment

Refining the Metabolic Model for Escherichia coli F11



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.

Run build FBA model

Dataset: E. coli F11 Genome

Other Options:

gs: 2
pairs: 5,280,267
232
ontent: 50.64%

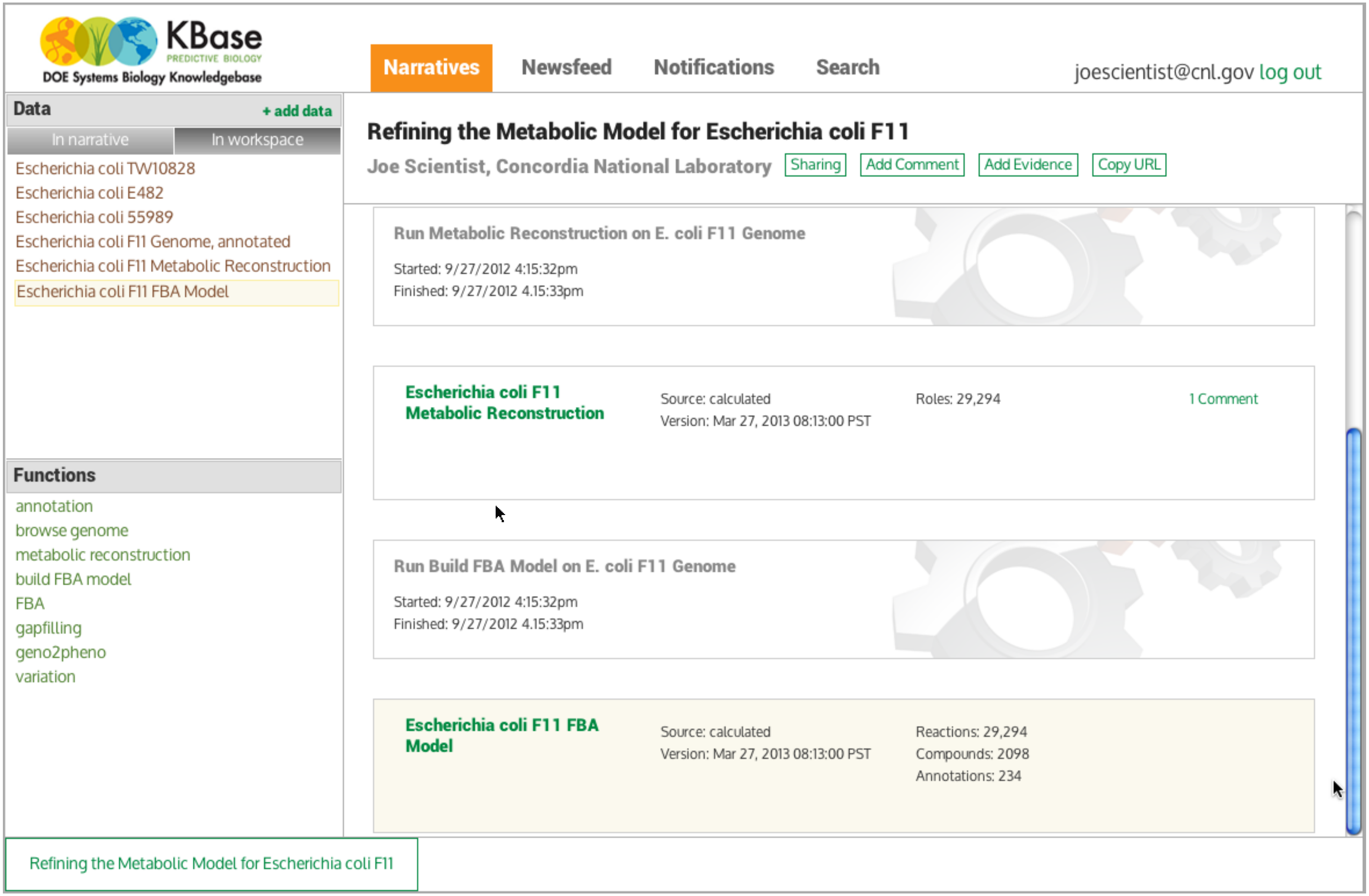
the initial one comes out.
to see which roles are found.

Escherichia coli F11 Metabolic Reconstruction

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Roles: 29,294

Refining the Metabolic Model for Escherichia coli F11



KBase PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Narratives Newsfeed Notifications Search

joescientist@cnl.gov log out

Data + add data

In narrative | In workspace

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory Sharing Add Comment Add Evidence Copy URL

Run Metabolic Reconstruction on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

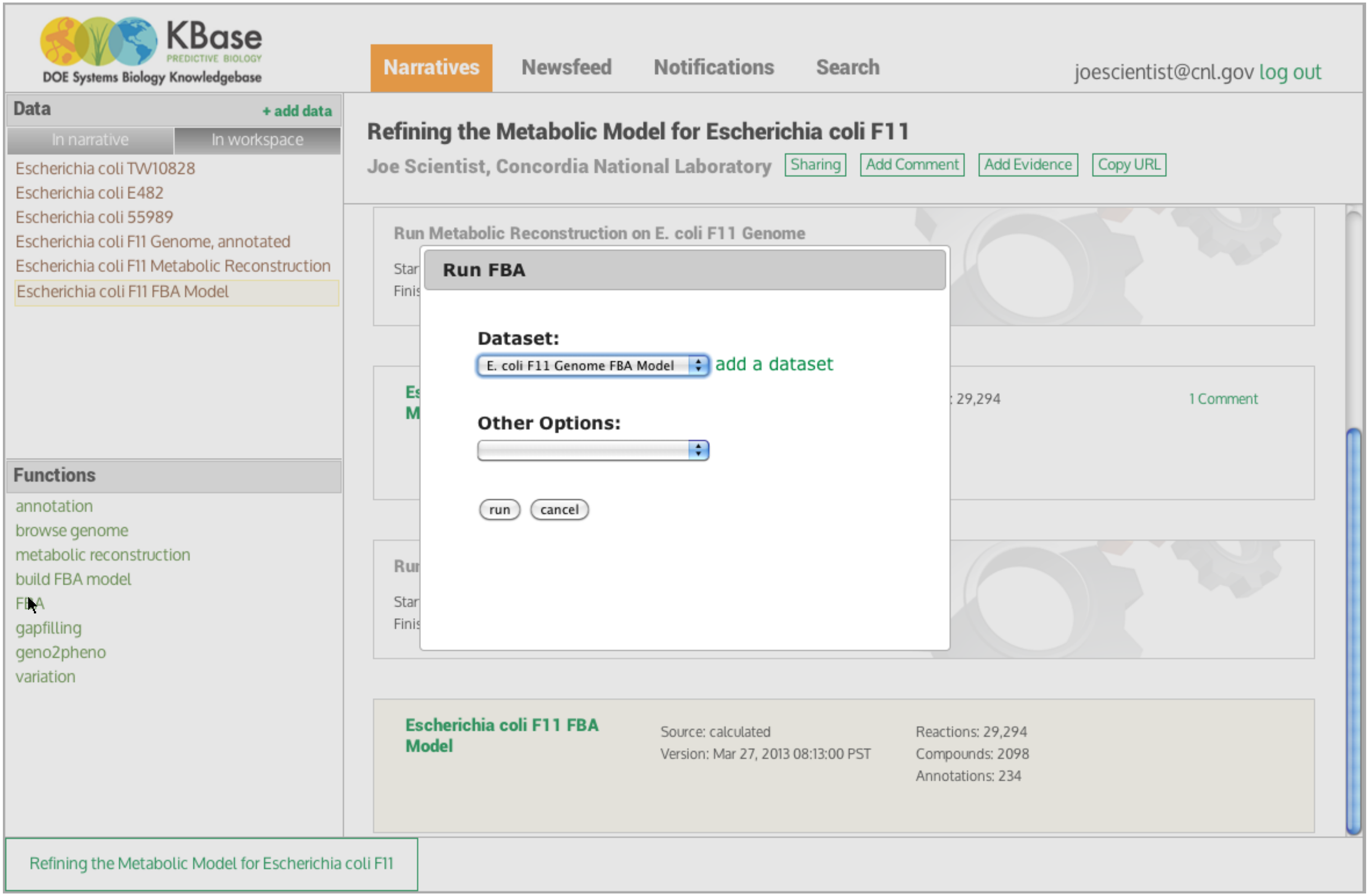
Escherichia coli F11 Metabolic Reconstruction
Source: calculated
Version: Mar 27, 2013 08:13:00 PST
Roles: 29,294
1 Comment

Run Build FBA Model on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 FBA Model
Source: calculated
Version: Mar 27, 2013 08:13:00 PST
Reactions: 29,294
Compounds: 2098
Annotations: 234

Refining the Metabolic Model for Escherichia coli F11



Data

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Run Metabolic Reconstruction on E. coli F11 Genome

Run FBA

Dataset: E. coli F11 Genome FBA Model

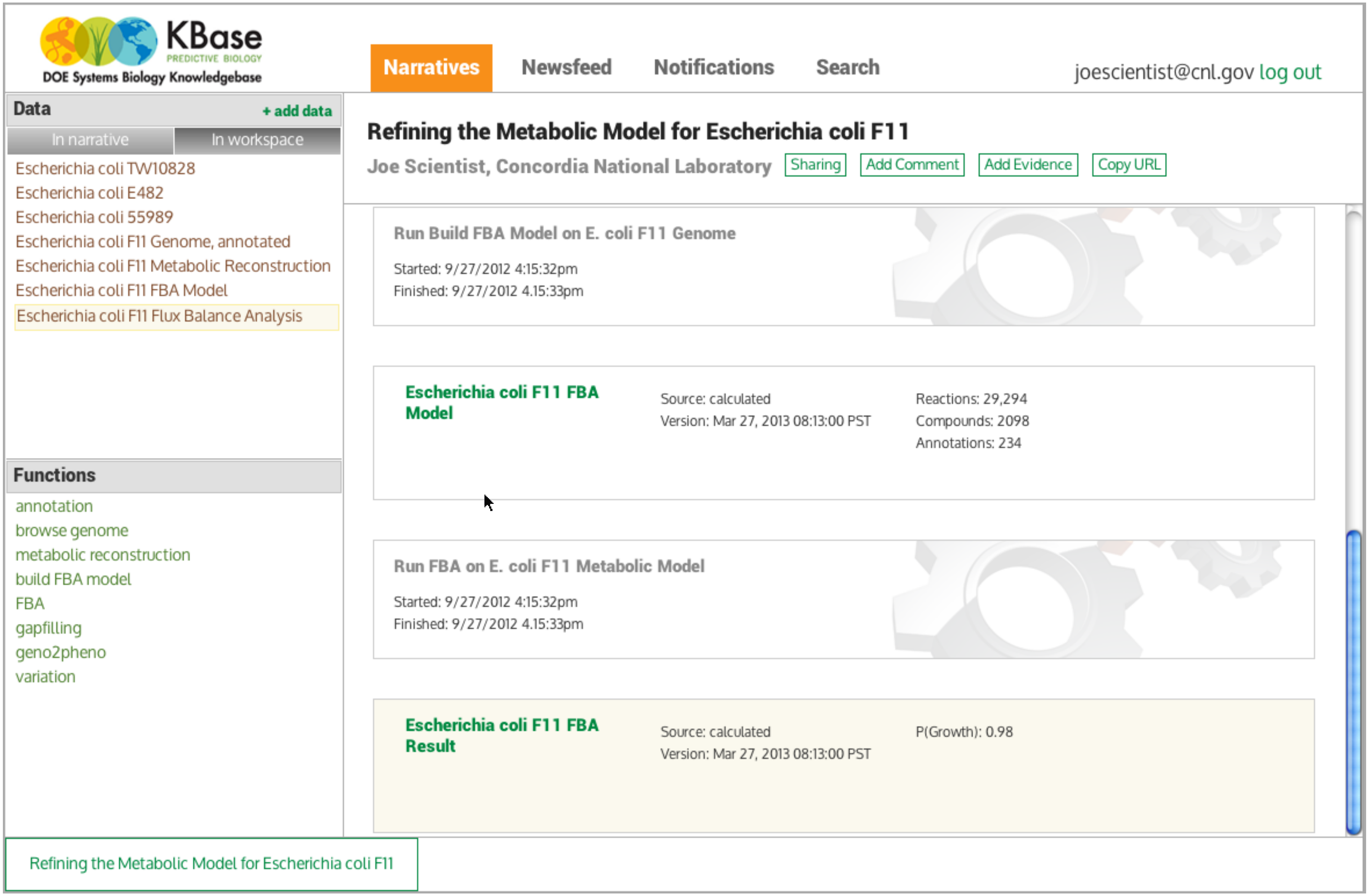
Other Options:

Escherichia coli F11 FBA Model

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Reactions: 29,294
Compounds: 2098
Annotations: 234

Refining the Metabolic Model for Escherichia coli F11



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction
Escherichia coli F11 FBA Model
Escherichia coli F11 Flux Balance Analysis

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Run Build FBA Model on E. coli F11 Genome

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 FBA Model

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

Reactions: 29,294
Compounds: 2098
Annotations: 234

Run FBA on E. coli F11 Metabolic Model

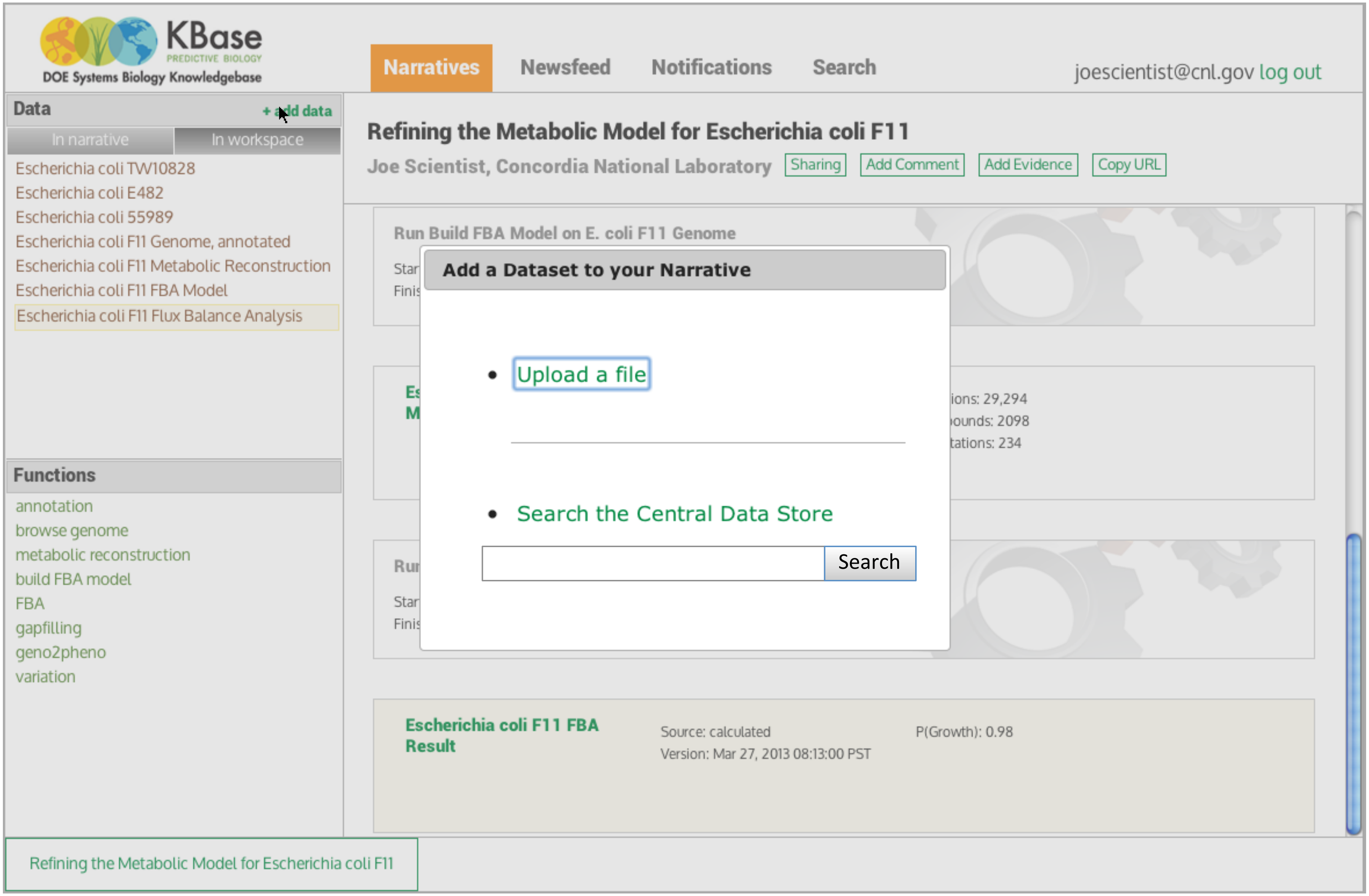
Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 FBA Result

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

P(Growth): 0.98

Refining the Metabolic Model for Escherichia coli F11



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction
Escherichia coli F11 FBA Model
Escherichia coli F11 Flux Balance Analysis

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Run Build FBA Model on E. coli F11 Genome

Star
Finis

Es
M

ions: 29,294
ounds: 2098
tations: 234

Rur

Star
Finis

Add a Dataset to your Narrative

- Upload a file

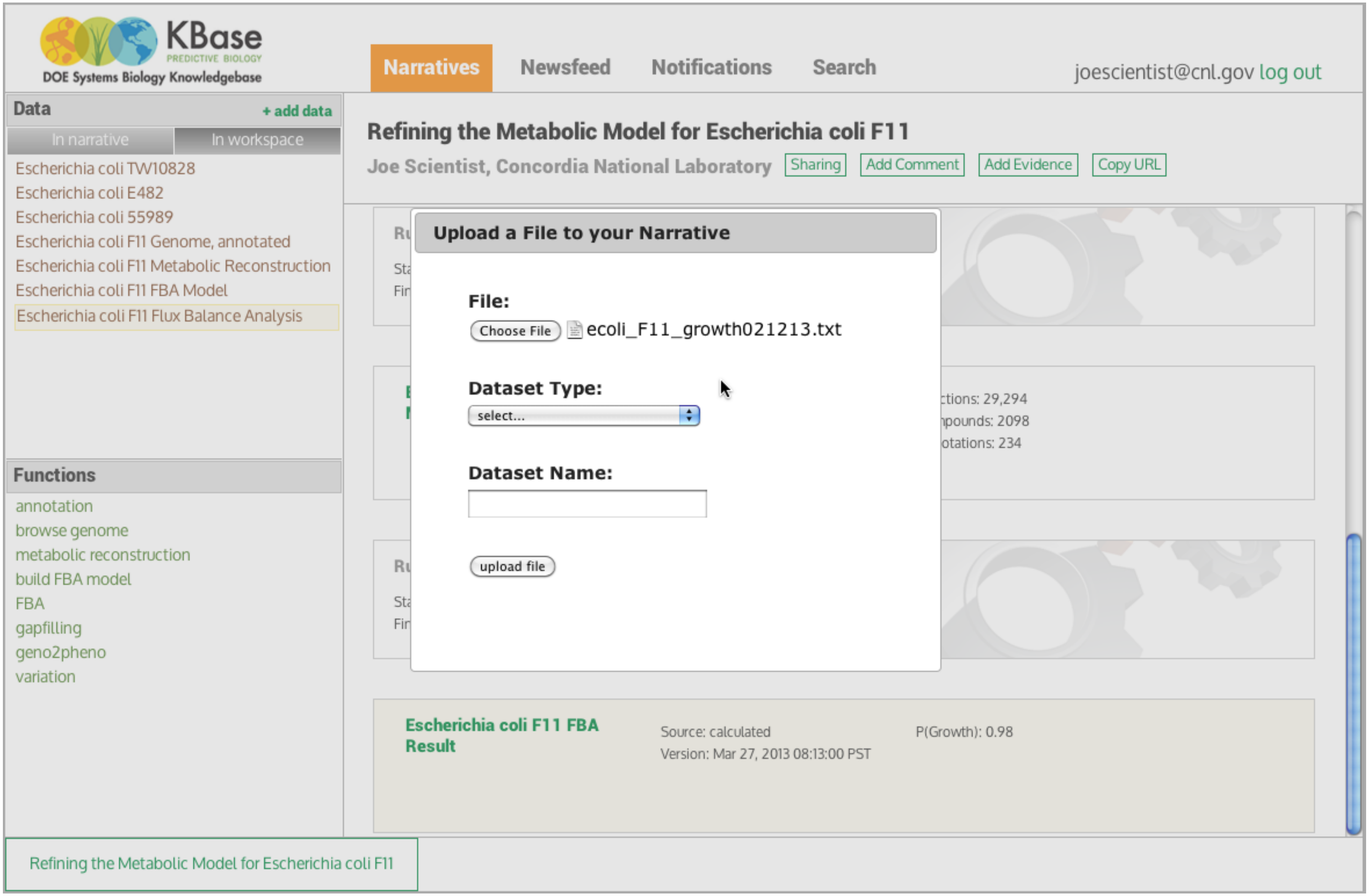
- Search the Central Data Store

Escherichia coli F11 FBA Result

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

P(Growth): 0.98

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Data:

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model
- Escherichia coli F11 Flux Balance Analysis

Functions:

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation

Upload a File to your Narrative

File: ecoli_F11_growth021213.txt

Dataset Type: select...

Dataset Name:

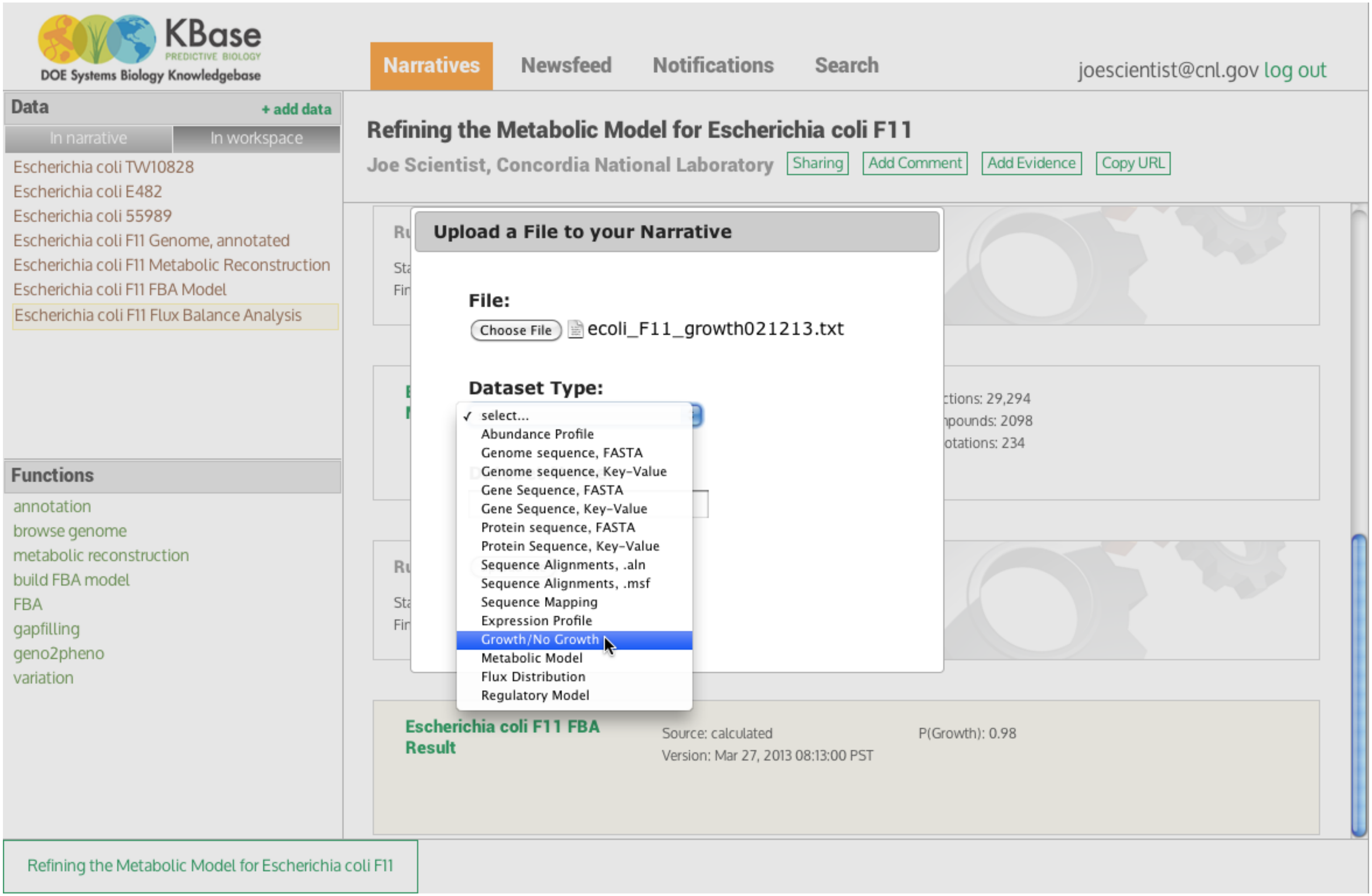
Escherichia coli F11 FBA Result

Source: calculated

Version: Mar 27, 2013 08:13:00 PST

P(Growth): 0.98

Refining the Metabolic Model for Escherichia coli F11



joescientist@cnl.gov

Data

Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction
Escherichia coli F11 FBA Model
Escherichia coli F11 Flux Balance Analysis

Functions

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Upload a File to your Narrative

File: ecoli_F11_growth021213.txt

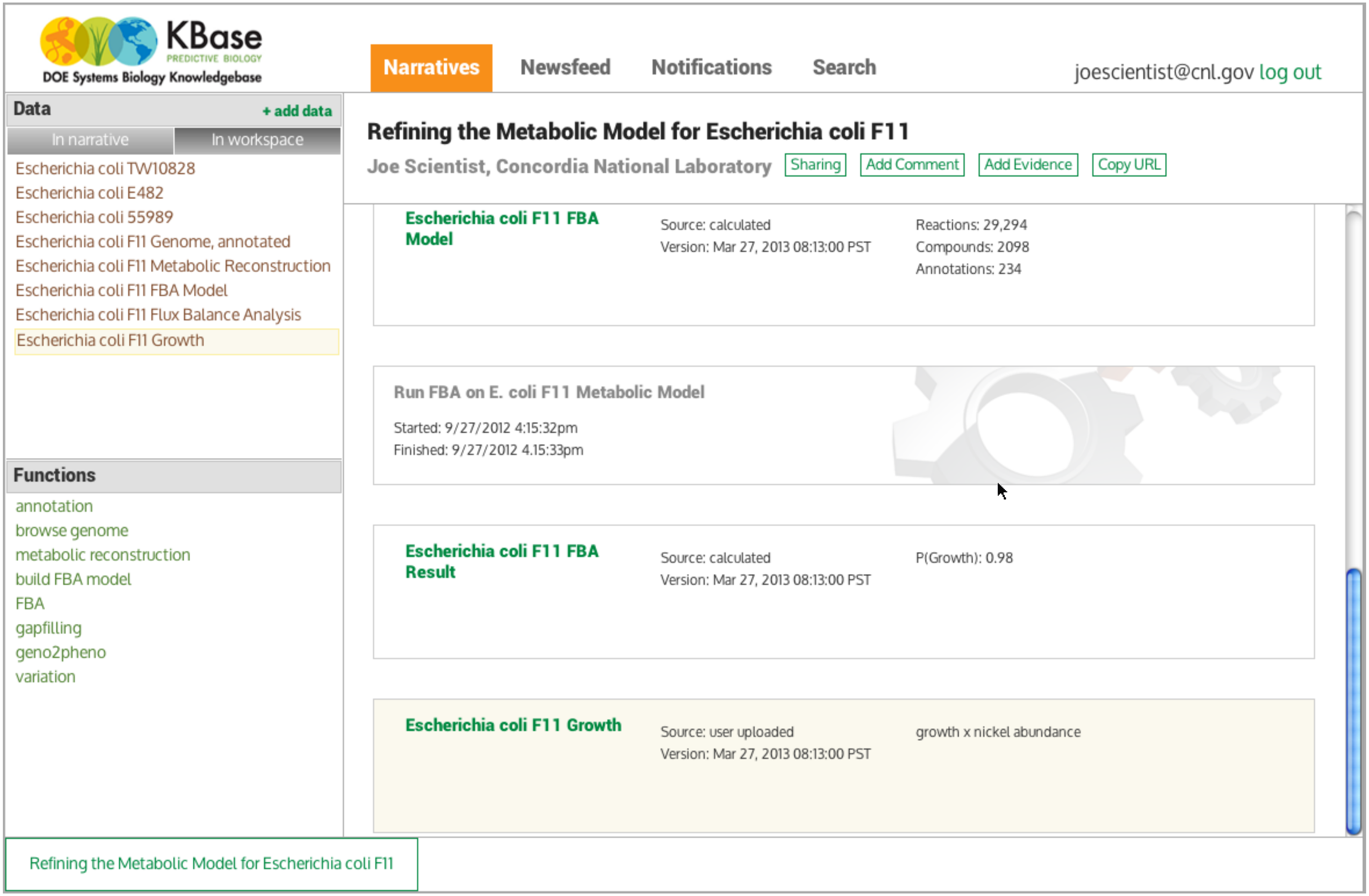
Dataset Type: Growth/No Growth

Escherichia coli F11 FBA Result

Source: calculated
Version: Mar 27, 2013 08:13:00 PST

P(Growth): 0.98

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Data

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model
- Escherichia coli F11 Flux Balance Analysis
- Escherichia coli F11 Growth

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation

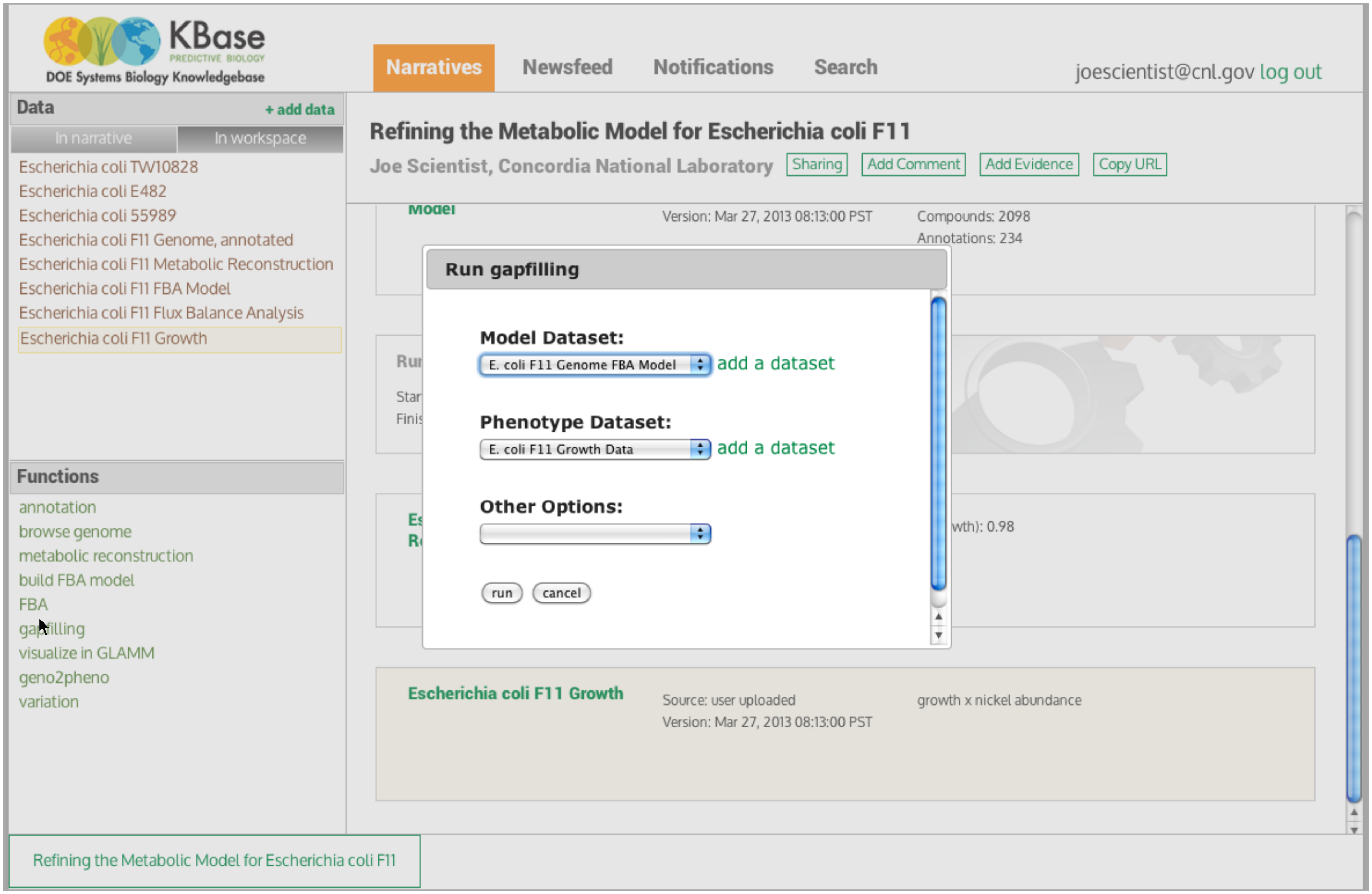
Escherichia coli F11 FBA Model	Source: calculated Version: Mar 27, 2013 08:13:00 PST	Reactions: 29,294 Compounds: 2098 Annotations: 234

Run FBA on E. coli F11 Metabolic Model

Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 FBA Result	Source: calculated Version: Mar 27, 2013 08:13:00 PST	P(Growth): 0.98
Escherichia coli F11 Growth	Source: user uploaded Version: Mar 27, 2013 08:13:00 PST	growth x nickel abundance

Refining the Metabolic Model for Escherichia coli F11



KBase PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Narratives Newsfeed Notifications Search joescientist@cnl.gov log out

Data + add data

In narrative | In workspace

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model
- Escherichia coli F11 Flux Balance Analysis
- Escherichia coli F11 Growth

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- visualize in GLAMM
- geno2pheno
- variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory Sharing Add Comment Add Evidence Copy URL

Model Version: Mar 27, 2013 08:13:00 PST Compounds: 2098 Annotations: 234

Run gapfilling

Model Dataset: E. coli F11 Genome FBA Model add a dataset

Phenotype Dataset: E. coli F11 Growth Data add a dataset

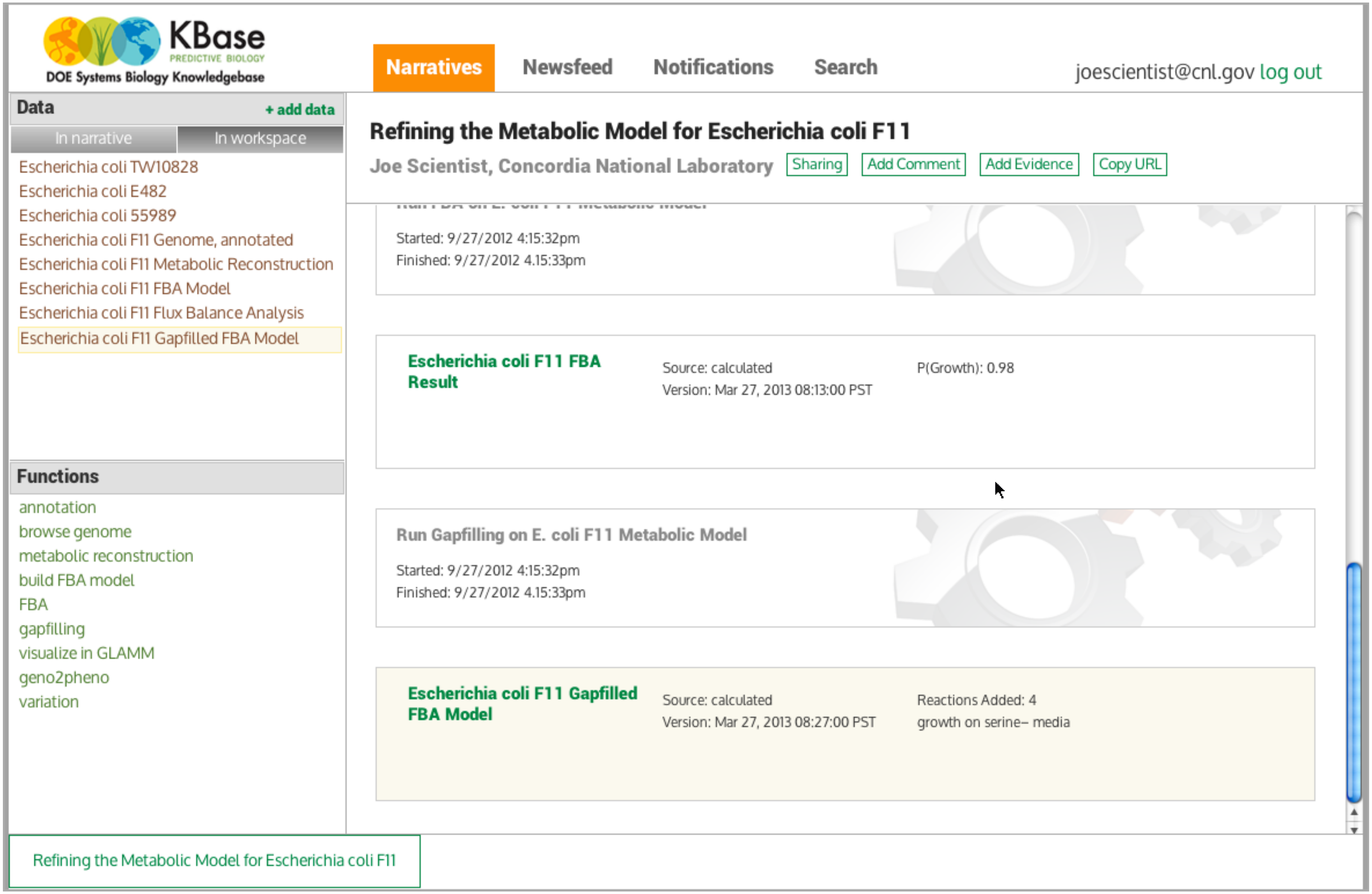
Other Options:

run cancel

wth): 0.98

Escherichia coli F11 Growth Source: user uploaded Version: Mar 27, 2013 08:13:00 PST growth x nickel abundance

Refining the Metabolic Model for Escherichia coli F11



Escherichia coli TW10828
Escherichia coli E482
Escherichia coli 55989
Escherichia coli F11 Genome, annotated
Escherichia coli F11 Metabolic Reconstruction
Escherichia coli F11 FBA Model
Escherichia coli F11 Flux Balance Analysis
Escherichia coli F11 Gapfilled FBA Model

annotation
browse genome
metabolic reconstruction
build FBA model
FBA
gapfilling
visualize in GLAMM
geno2pheno
variation

Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

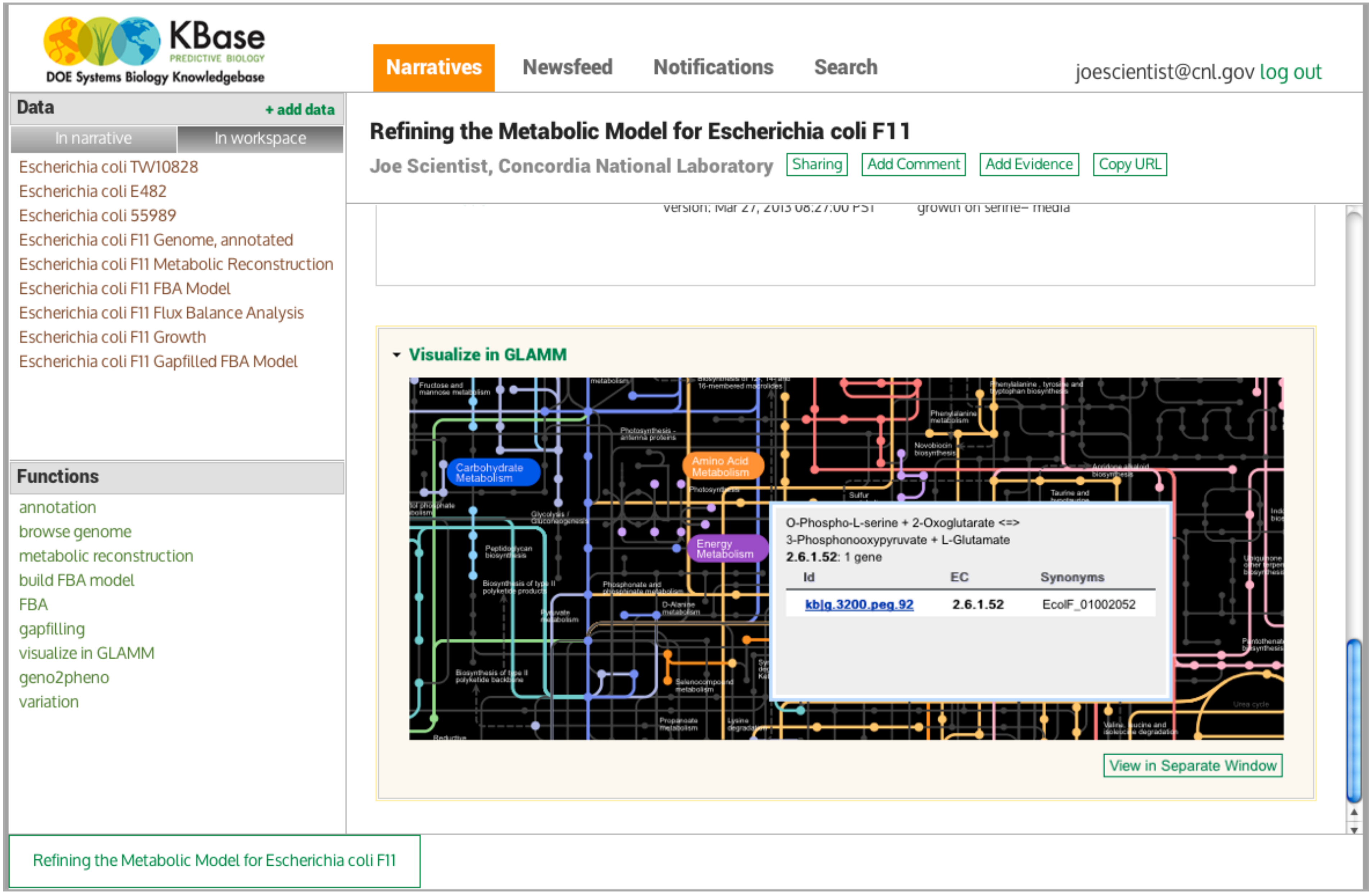
Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 FBA Result
Source: calculated
Version: Mar 27, 2013 08:13:00 PST
P(Growth): 0.98

Run Gapfilling on E. coli F11 Metabolic Model
Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4.15:33pm

Escherichia coli F11 Gapfilled FBA Model
Source: calculated
Version: Mar 27, 2013 08:27:00 PST
Reactions Added: 4
growth on serine– media

Refining the Metabolic Model for Escherichia coli F11



Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Data:

- Escherichia coli TW10828
- Escherichia coli E482
- Escherichia coli 55989
- Escherichia coli F11 Genome, annotated
- Escherichia coli F11 Metabolic Reconstruction
- Escherichia coli F11 FBA Model
- Escherichia coli F11 Flux Balance Analysis
- Escherichia coli F11 Growth
- Escherichia coli F11 Gapfilled FBA Model

Functions:

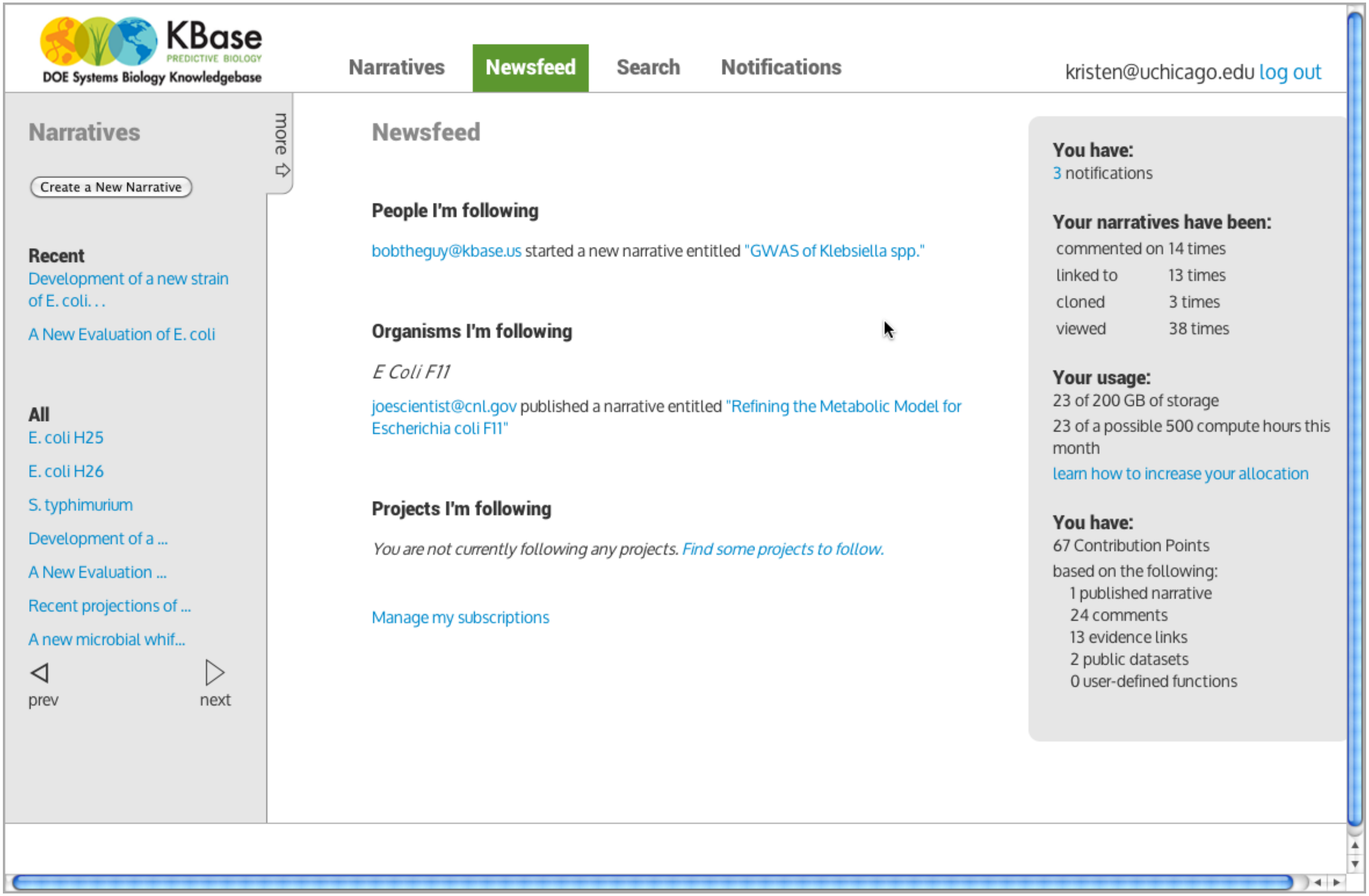
- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- visualize in GLAMM
- geno2pheno
- variation

Version: Mar 27, 2013 08:27:00 PST growth on serine– media

O-Phospho-L-serine + 2-Oxoglutarate <=>
3-Phosphonooxypyruvate + L-Glutamate
2.6.1.52: 1 gene

Id	EC	Synonyms
kb|g.3200.peg.92	2.6.1.52	EcolF_01002052

Refining the Metabolic Model for Escherichia coli F11



Narratives

Recent
Development of a new strain of E. coli. . .
A New Evaluation of E. coli

All
E. coli H25
E. coli H26
S. typhimurium
Development of a ...
A New Evaluation ...
Recent projections of ...
A new microbial whif...

Newsfeed

People I'm following

bobtheguy@kbase.us started a new narrative entitled "GWAS of Klebsiella spp."

Organisms I'm following

E Coli F11

joescientist@cnl.gov published a narrative entitled "Refining the Metabolic Model for Escherichia coli F11"

Projects I'm following

You are not currently following any projects. Find some projects to follow.

Manage my subscriptions

You have:
3 notifications

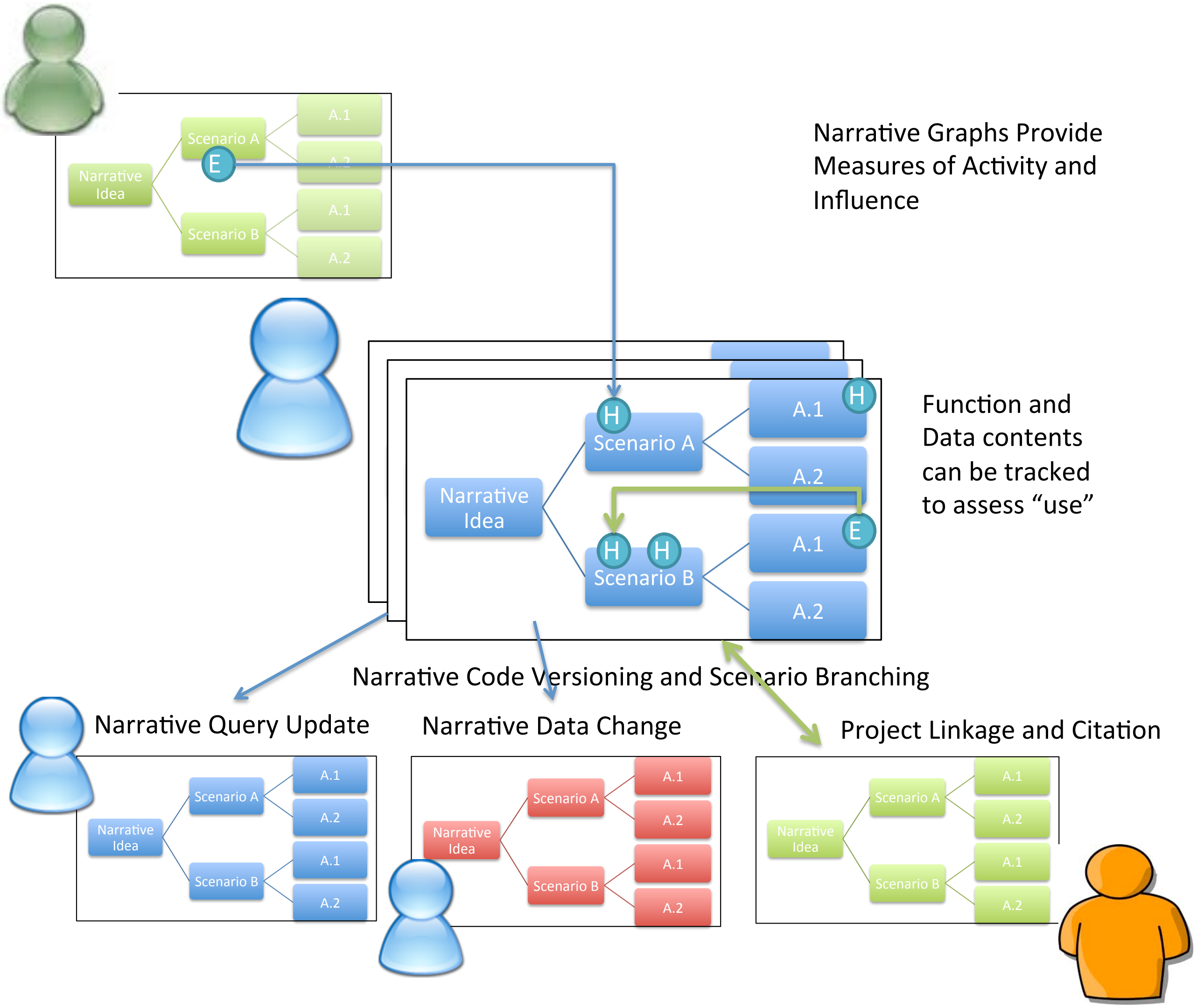
Your narratives have been:

commented on	14 times
linked to	13 times
cloned	3 times
viewed	38 times

Your usage:
23 of 200 GB of storage
23 of a possible 500 compute hours this month
learn how to increase your allocation

You have:
67 Contribution Points
based on the following:
- 1 published narrative
- 24 comments
- 13 evidence links
- 2 public datasets
- 0 user-defined functions

Concept: Interactive community knowledge



Narrative Graphs Provide Measures of Activity and Influence

Function and Data contents can be tracked to assess “use”

Narrative Code Versioning and Scenario Branching

Narrative Query Update

Narrative Data Change

Project Linkage and Citation

Into the future

- The data model must evolve to support the modeling mission
- A framework for turning bioinformatics algorithms' output into models needs further development
- **Theory for integrative, cross-scale predictive biology under development**
- Much better ontologies for nearly everything
- Building the social tools.

- Need to improve data import, quality assessment and metadata
- Efficient incorporation of new third party algorithms and support with scalable compute.
- Growing a strong external development community while maintaining quality, stability and vision.
- Launching the KBase Foundation to ease licensing and growth of KBase user participation.

↳ A more concrete view of the "Narrative" interface

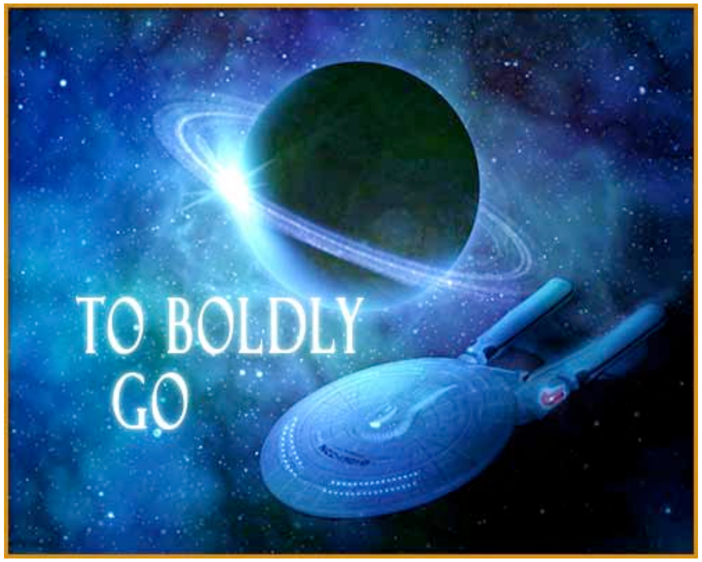
Thanks to our Team and DOE!

Arkin, Adam	LBNL	Management
Baumohl, Jason	LBNL	Microbes
Best, Aaron	Hope	Microbes
Bischof, Jared	ANL	Communities (Folker)
Bowen, Ben	LBNL	Microbes
Brenner, Steven	LBNL	Microbes
Brettin, Tom	ANL	Infrastructure
Canon, Shane	LBNL	Infrastructure
Chan, Stephen	LBNL	Infrastructure
Chandonia, John-Marc	LBNL	Microbes, Networks
Chivian, Dylan	LBNL	Communities
Cholia, Shreyas	LBNL	Infrastructure
Conrad, Neal	ANL	Infrastructure
Cottingham, Bob	ORNL	Management
Davison, Brian	ORNL	Outreach
Dehal, Paramvir	LBNL	Microbes
DeJongh, Matt	Hope	Microbes
Desai, Narayan	ANL	Infrastructure
Devoid, Scott	ANL	Infrastructure
Disz, Terry	ANL	Infrastructure
Drake, Meghan	ORNL	Outreach
Fang, Gang	Yale	Plants
Fei He	BNL	Plants
Frybarger, Paul	ANL	Microbes
Gerstein, Mark	Yale	Plants
Glass, Elizabeth Marland	ANL	Outreach
Gordon, India	ANL	Adminstration
Greiner, Annette	LBL/NERSC	Infrastructure
Gurtowski, James	CSHL	Plants
Harris, Nomi	LBNL	Management
Harrison, Travis	ANL	Communities
Henderson, Matt	LBNL	Communities
Henry, Chris	ANL	Microbes
Howe, Adina	ANL	Communities
Joachimiak, Marcin	LBNL	Microbes/Communities
Keegan, Kevin	ANL	Communities
Keller, Keith	LBNL	Microbes
Kora, Guru	ORNL	Infrastructure
Kumari, Sunita	CSHL	Plants
Land, Miriam	ORNL	Infrastructure
Leuze, Michael	ORNL	Infrastructure
Lochosvky, Lucas	Yale	Plants
Maslov, Sergei	BNL	Management
Meyer, Folker	ANL	Communities
Moulton, Steve	ORNL	Microbes
Mulligan, Silvia (Palomino)	ANL	Administration
Novichkov, Pavel	LBNL	Infrastructure/Microbes
Oh, Taeyun	UC Davis	Plants
Olsen, Gary	UIUC	Microbes
Olson, Bob	ANL	Infrastructure
Olson, Dan	ANL	Infrastructure
Overbeek, Ross	ANL	Microbes
Paczian, Tobias	ANL	Communities
Palumbo, Tony	ORNL	Outreach
Parrello, Bruce	ANL	Infrastructure
Pasternak, Shiran	CSHL	Plants
Poon, Sarah	LBL/NERSC	Infrastructure
Price, Gavin	LBL	Microbes
Ranjan, Priya	ORNL	Plants
Riehl, Bill	LBNL	Microbes
Ronald, Pamela	UC Davis	Plants
Salazar, Jennifer Fessler	ANL	Outreach
Schatz, Michael	CSHL	Plants
Seaver, Sam	ANL	Plants
Sharma, Manoj	LBNL	Infrastructure
Sharma, Rita	LBNL	Infrastructure
Sneddon, Michael	LBNL	Communities
Stevens, Rick	ANL	Management
Syed, Mustafa	ORNL	Infrastructure
Terry, Gwyenth	LBNL	Adminstration
Thomason, James	CSHL	Plants
Tintle, Nathan	Hope	Microbes
Trimble, Will	ANL	Communities
Tuskan, Jerry	ORNL	Plants
Wang, Daifeng	Yale	Plants
Ware, Doreen	CSHL	Plants
Weston, David	ORNL	Plants
Whitney, Cary	LBNL	Infrastructure
Wilke, Andreas	ANL	Communities
Wilkening, Jared	ANL	Communities
Xia, Fangfang	ANL	Microbes
Yoo, Shinjae	BNL	Plants
Yu, Dantong	BNL	Plants

Thank you!

Visit us at kbase.science.energy.gov

Contact us at outreach@kbase.us




KBase