KBase
PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

INTEGRATION
AND MODELING
for PREDICTIVE
BIOLOGY

DOE Systems Biology Knowledgebase

04/01/2015

KBase Team

Berkeley

**Adam Arkin*
**Dylan Chivian*
**Paramvir Dehal*
**Nomi Harris*
**Fernando Perez*
Jason Baumohl
Ben Bowen
Shane Canon
John-Marc Chandonia
Matt Henderson
Marcin Joachimiak
Keith Keller
Pavel Novichkov
Erik Pearson
Sarah Poon
Gavin Price
Sara Ricks
Bill Riehl
Michael Sneddon
Roman Sutormin
Gwyneth Terry

Brookhaven

**Sergei Maslov*
Fei He
Shinjae Yoo

Cold Spring Harbor

**Doreen Ware*
James Gurtowski
Vivek Kumar
Sunita Kumari
Michael Schatz
Srividya Ramakrishnan
James Thomason

Argonne

**Rick Stevens*
**Tom Brettin*
**Elizabeth Glass*
**Chris Henry*
**Folker Meyer*
Emily Dietrich
Jared Bischof
Sebastien Boisvert
Chris Bun
Neal Conrad
Janaka Edirisinghe
Wolfgang Gerlach
India Gordon
Travis Harrison
Kevin Keegan
Silvia Mulligan
Bob Olson
Dan Olson
Ross Overbeek
Tobias Paczian
Bruce Parello
Sam Seaver
Will Trimble
Andreas Wilke
Fangfang Xia

Hope College

Matt DeJongh
Aaron Best
Nathan Tintle

Univ. TN

Priya Ranjan

UIUC

Gary Olsen

Communities **Microbes**
Infrastructure **Plants**
Management **Outreach**
** Denotes team lead(s)*

Oak Ridge

**Robert Cottingham*
**Dave Weston*
Benjamin Allen
Meghan Drake
Holly Haun
Betty Mansfield
Marissa Mills

Office of Biological and Environmental Research

DOE Systems Biology Knowledgebase

- A Socially-Enabled Data Science Platform for Biology

What Biologists Want

- Lower the bar to analysis of their complex data sets
- Place their results in context of knowledge in the field
- Easily build on their own and others' work
- Control sharing of and credit for their work

What Agencies Want

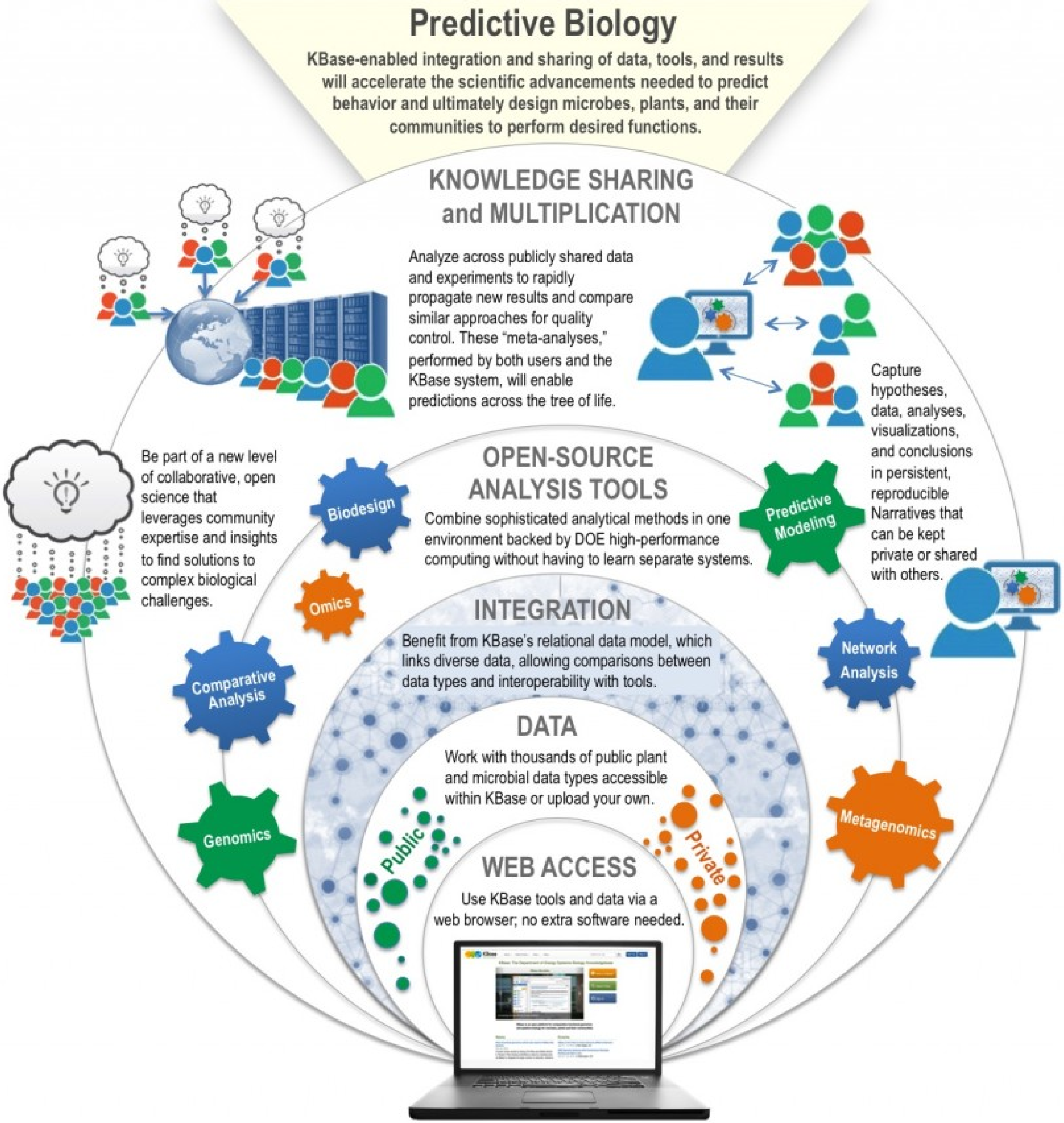
- Stewardship and communication of data and results
- Optimal use of the fruits of their investment
- Stimulation of their scientist communities
- Metrics of Impact and Utility

KBase SFA goals

- Meet these goals…
- Derive more understanding of biological function than available from any single project through continual reasoning across shared data and conclusions.

What I want:

- What determines fitness for an organism in a given target location?
- How do I engineer this organism with a desired activity while controlling fitness?



Predictive Biology

KBase-enabled integration and sharing of data, tools, and results will accelerate the scientific advancements needed to predict behavior and ultimately design microbes, plants, and their communities to perform desired functions.

KNOWLEDGE SHARING and MULTIPLICATION

Analyze across publicly shared data and experiments to rapidly propagate new results and compare similar approaches for quality control. These “meta-analyses,” performed by both users and the KBase system, will enable predictions across the tree of life.

Be part of a new level of collaborative, open science that leverages community expertise and insights to find solutions to complex biological challenges.

Capture hypotheses, data, analyses, visualizations, and conclusions in persistent, reproducible Narratives that can be kept private or shared with others.

OPEN-SOURCE ANALYSIS TOOLS

Combine sophisticated analytical methods in one environment backed by DOE high-performance computing without having to learn separate systems.

INTEGRATION

Benefit from KBase’s relational data model, which links diverse data, allowing comparisons between data types and interoperability with tools.

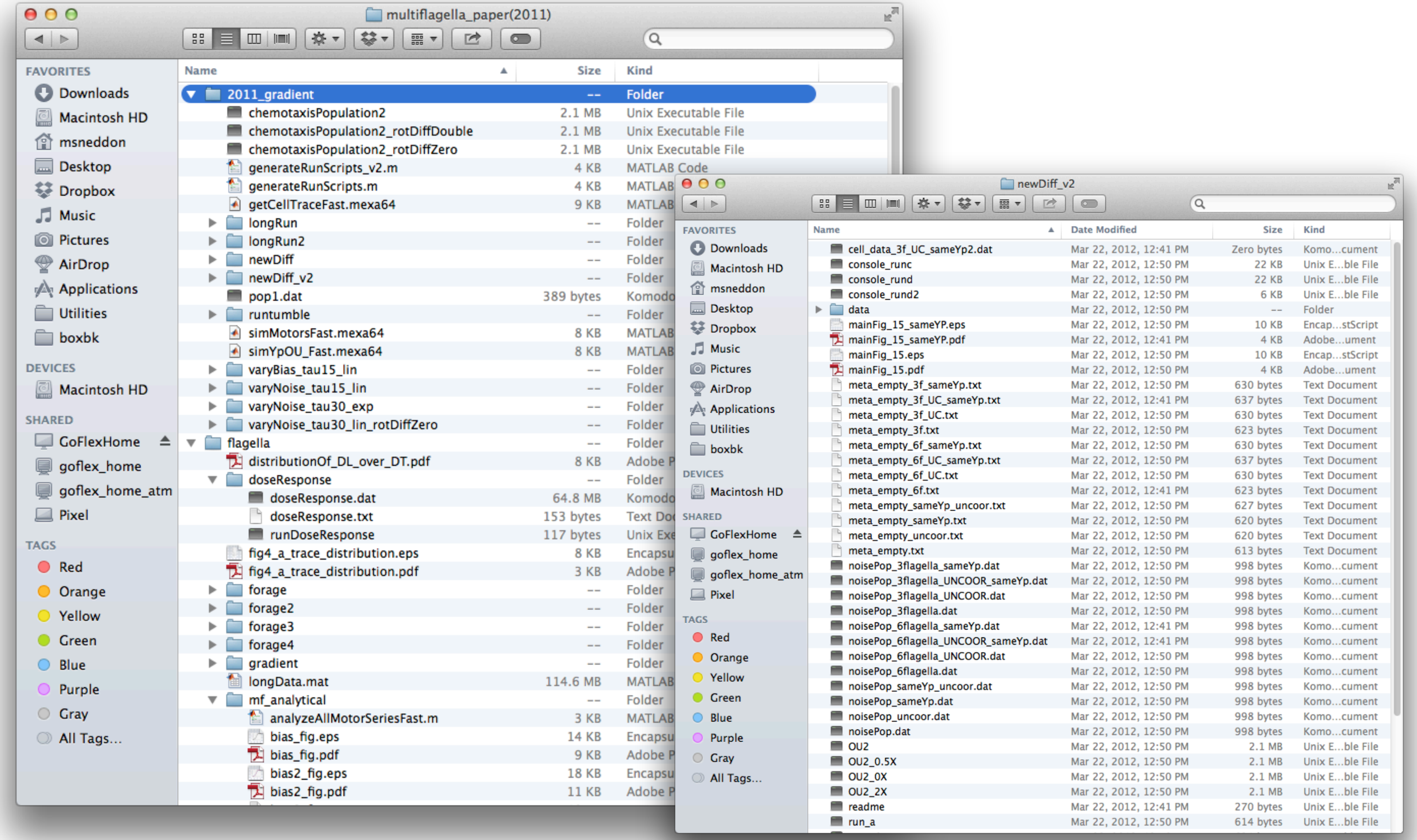
DATA

Work with thousands of public plant and microbial data types accessible within KBase or upload your own.

WEB ACCESS

Use KBase tools and data via a web browser; no extra software needed.

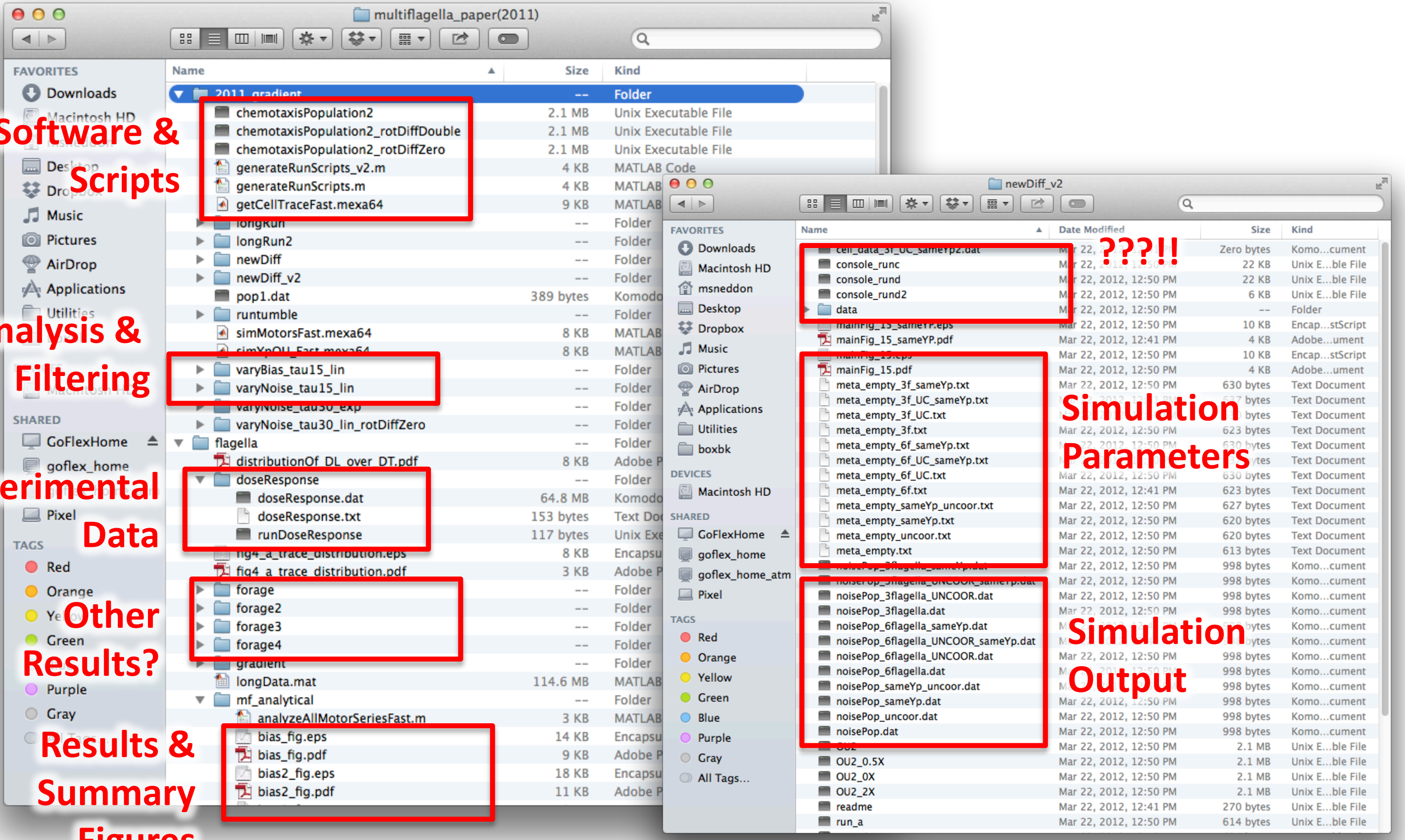
The Data Problem in Systems Biology



Name	Size	Kind
▼ 2011_gradient	--	Folder
chemotaxisPopulation2	2.1 MB	Unix Executable File
chemotaxisPopulation2_rotDiffDouble	2.1 MB	Unix Executable File
chemotaxisPopulation2_rotDiffZero	2.1 MB	Unix Executable File
generateRunScripts_v2.m	4 KB	MATLAB Code
generateRunScripts.m	4 KB	MATLAB
getCellTraceFast.mexa64	9 KB	MATLAB
▶ longRun	--	Folder
▶ longRun2	--	Folder
▶ newDiff	--	Folder
▶ newDiff_v2	--	Folder
pop1.dat	389 bytes	Komodo
▶ runtumble	--	Folder
simMotorsFast.mexa64	8 KB	MATLAB
simYpOU_Fast.mexa64	8 KB	MATLAB
▶ varyBias_tau15_lin	--	Folder
▶ varyNoise_tau15_lin	--	Folder
▶ varyNoise_tau30_exp	--	Folder
▶ varyNoise_tau30_lin_rotDiffZero	--	Folder
▼ flagella	--	Folder
distributionOf_DL_over_DT.pdf	8 KB	Adobe P
▼ doseResponse	--	Folder
doseResponse.dat	64.8 MB	Komodo
doseResponse.txt	153 bytes	Text Doc
runDoseResponse	117 bytes	Unix Exe
fig4_a_trace_distribution.eps	8 KB	Encapsu
fig4_a_trace_distribution.pdf	3 KB	Adobe P
▶ forage	--	Folder
▶ forage2	--	Folder
▶ forage3	--	Folder
▶ forage4	--	Folder
▶ gradient	--	Folder
longData.mat	114.6 MB	MATLAB
▼ mf_analytical	--	Folder
analyzeAllMotorSeriesFast.m	3 KB	MATLAB
bias_fig.eps	14 KB	Encapsu
bias_fig.pdf	9 KB	Adobe P
bias2_fig.eps	18 KB	Encapsu
bias2_fig.pdf	11 KB	Adobe P

Name	Date Modified	Size	Kind
cell_data_3f_UC_sameYp2.dat	Mar 22, 2012, 12:41 PM	Zero bytes	Komo…cument
console_runc	Mar 22, 2012, 12:50 PM	22 KB	Unix E…ble File
console_rund	Mar 22, 2012, 12:50 PM	22 KB	Unix E…ble File
console_rund2	Mar 22, 2012, 12:50 PM	6 KB	Unix E…ble File
▶ data	Mar 22, 2012, 12:50 PM	--	Folder
mainFig_15_sameYP.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap…stScript
mainFig_15_sameYP.pdf	Mar 22, 2012, 12:41 PM	4 KB	Adobe…ument
mainFig_15.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap…stScript
mainFig_15.pdf	Mar 22, 2012, 12:50 PM	4 KB	Adobe…ument
meta_empty_3f_sameYp.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_3f_UC_sameYp.txt	Mar 22, 2012, 12:41 PM	637 bytes	Text Document
meta_empty_3f_UC.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_3f.txt	Mar 22, 2012, 12:50 PM	623 bytes	Text Document
meta_empty_6f_sameYp.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_6f_UC_sameYp.txt	Mar 22, 2012, 12:50 PM	637 bytes	Text Document
meta_empty_6f_UC.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_6f.txt	Mar 22, 2012, 12:41 PM	623 bytes	Text Document
meta_empty_sameYp_uncoor.txt	Mar 22, 2012, 12:50 PM	627 bytes	Text Document
meta_empty_sameYp.txt	Mar 22, 2012, 12:50 PM	620 bytes	Text Document
meta_empty_uncoor.txt	Mar 22, 2012, 12:50 PM	620 bytes	Text Document
meta_empty.txt	Mar 22, 2012, 12:50 PM	613 bytes	Text Document
noisePop_3flagella_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_3flagella_UNCOOR_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_3flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_3flagella.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_6flagella_sameYp.dat	Mar 22, 2012, 12:41 PM	998 bytes	Komo…cument
noisePop_6flagella_UNCOOR_sameYp.dat	Mar 22, 2012, 12:41 PM	998 bytes	Komo…cument
noisePop_6flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_6flagella.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_sameYp_uncoor.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop_uncoor.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
OU2	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_0.5X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_0X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_2X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
readme	Mar 22, 2012, 12:41 PM	270 bytes	Unix E…ble File
run_a	Mar 22, 2012, 12:50 PM	614 bytes	Unix E…ble File

The Data Problem in Systems Biology



Software & Scripts

Analysis & Filtering

Experimental Data

Other Results?

Results & Summary Figures

???!!

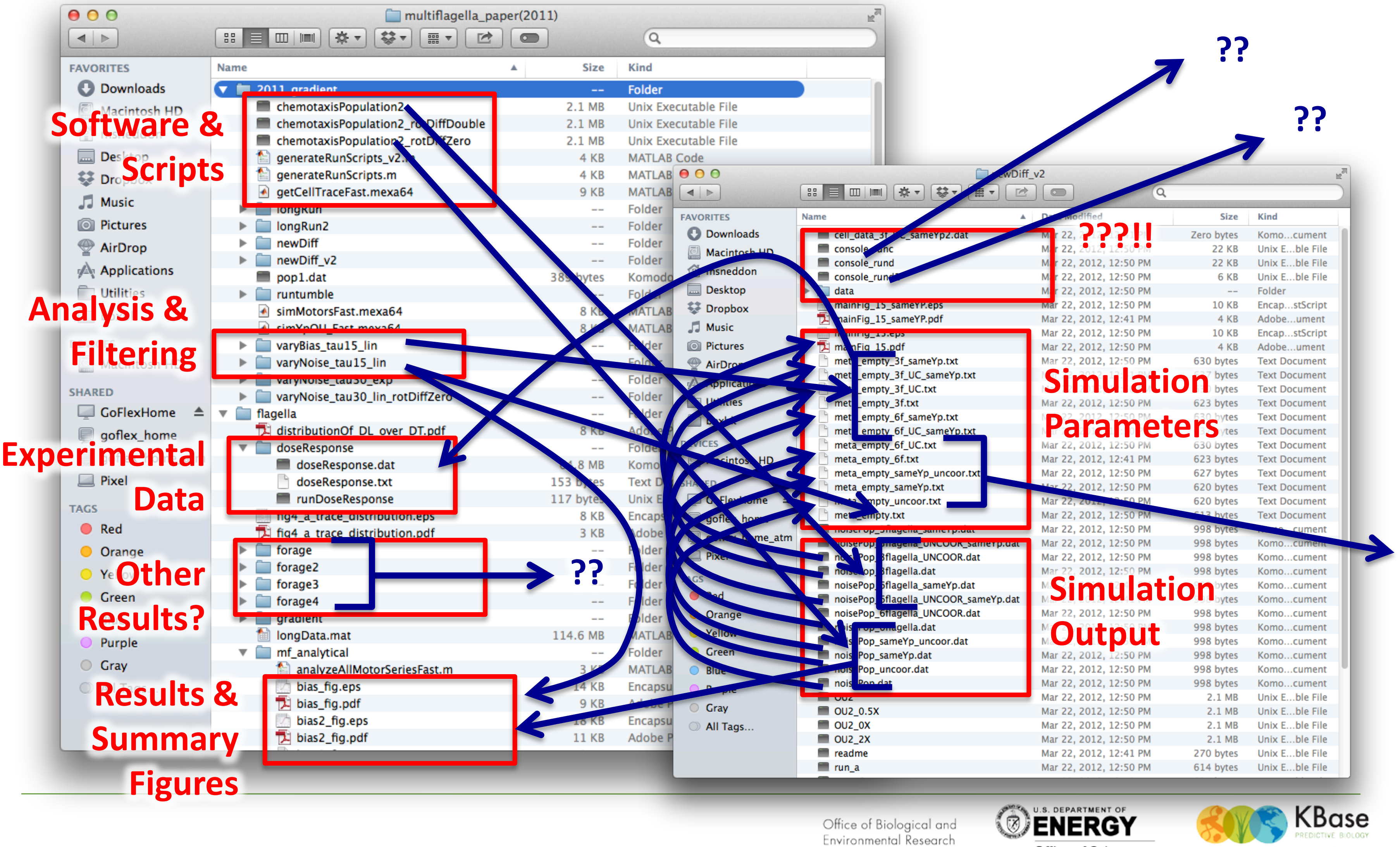
Simulation Parameters

Simulation Output

Name	Size	Kind
2011_gradient	--	Folder
chemotaxisPopulation2	2.1 MB	Unix Executable File
chemotaxisPopulation2_rotDiffDouble	2.1 MB	Unix Executable File
chemotaxisPopulation2_rotDiffZero	2.1 MB	Unix Executable File
generateRunScripts_v2.m	4 KB	MATLAB Code
generateRunScripts.m	4 KB	MATLAB
getCellTraceFast.mexa64	9 KB	MATLAB
longRun	--	Folder
longRun2	--	Folder
newDiff	--	Folder
newDiff_v2	--	Folder
pop1.dat	389 bytes	Komodo
runtumble	--	Folder
simMotorsFast.mexa64	8 KB	MATLAB
simXpOU_Fast.mexa64	8 KB	MATLAB
varyBias_tau15_lin	--	Folder
varyNoise_tau15_lin	--	Folder
varyNoise_tau30_exp	--	Folder
varyNoise_tau30_lin_rotDiffZero	--	Folder
flagella	--	Folder
distributionOf_DL_over_DT.pdf	8 KB	Adobe P
doseResponse	--	Folder
doseResponse.dat	64.8 MB	Komodo
doseResponse.txt	153 bytes	Text Doc
runDoseResponse	117 bytes	Unix Exe
fig4_a_trace_distribution.eps	8 KB	Encapsu
fig4_a_trace_distribution.pdf	3 KB	Adobe P
forage	--	Folder
forage2	--	Folder
forage3	--	Folder
forage4	--	Folder
gradient	--	Folder
longData.mat	114.6 MB	MATLAB
mf_analytical	--	Folder
analyzeAllMotorSeriesFast.m	3 KB	MATLAB
bias_fig.eps	14 KB	Encapsu
bias_fig.pdf	9 KB	Adobe P
bias2_fig.eps	18 KB	Encapsu
bias2_fig.pdf	11 KB	Adobe P

Name	Date Modified	Size	Kind
cell_data_3f_UC_sameYp2.dat	[illegible]	Zero bytes	Komo...cument
console_runc	[illegible]	22 KB	Unix E...ble File
console_rund	Mar 22, 2012, 12:50 PM	22 KB	Unix E...ble File
console_rund2	Mar 22, 2012, 12:50 PM	6 KB	Unix E...ble File
data	Mar 22, 2012, 12:50 PM	--	Folder
mainFig_15_sameYP.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap...stScript
mainFig_15_sameYP.pdf	Mar 22, 2012, 12:41 PM	4 KB	Adobe...ument
mainFig_15.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap...stScript
mainFig_15.pdf	Mar 22, 2012, 12:50 PM	4 KB	Adobe...ument
meta_empty_3f_sameYp.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_3f_UC_sameYp.txt	[illegible]	[illegible]	Text Document
meta_empty_3f_UC.txt	[illegible]	[illegible]	Text Document
meta_empty_3f.txt	Mar 22, 2012, 12:50 PM	623 bytes	Text Document
meta_empty_6f_sameYp.txt	[illegible]	630 bytes	Text Document
meta_empty_6f_UC_sameYp.txt	[illegible]	[illegible]	Text Document
meta_empty_6f_UC.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_6f.txt	Mar 22, 2012, 12:41 PM	623 bytes	Text Document
meta_empty_sameYp_uncoor.txt	Mar 22, 2012, 12:50 PM	627 bytes	Text Document
meta_empty_sameYp.txt	Mar 22, 2012, 12:50 PM	620 bytes	Text Document
meta_empty_uncoor.txt	Mar 22, 2012, 12:50 PM	620 bytes	Text Document
meta_empty.txt	Mar 22, 2012, 12:50 PM	613 bytes	Text Document
noisePop_3flagella_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_3flagella_UNCOOR_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_3flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_3flagella.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_6flagella_sameYp.dat	[illegible]	[illegible]	Komo...cument
noisePop_6flagella_UNCOOR_sameYp.dat	[illegible]	[illegible]	Komo...cument
noisePop_6flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_6flagella.dat	[illegible]	998 bytes	Komo...cument
noisePop_sameYp_uncoor.dat	[illegible]	998 bytes	Komo...cument
noisePop_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop_uncoor.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
noisePop.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo...cument
OU2	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E...ble File
OU2_0.5X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E...ble File
OU2_0X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E...ble File
OU2_2X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E...ble File
readme	Mar 22, 2012, 12:41 PM	270 bytes	Unix E...ble File
run_a	Mar 22, 2012, 12:50 PM	614 bytes	Unix E...ble File

The Data Problem in Systems Biology



Name	Size	Kind
2011_gradient	--	Folder
chemotaxisPopulation2	2.1 MB	Unix Executable File
chemotaxisPopulation2_r DiffDouble	2.1 MB	Unix Executable File
chemotaxisPopulation2_rotDiffZero	2.1 MB	Unix Executable File
generateRunScripts_v2.m	4 KB	MATLAB Code
generateRunScripts.m	4 KB	MATLAB
getCellTraceFast.mexa64	9 KB	MATLAB
longRun	--	Folder
longRun2	--	Folder
newDiff	--	Folder
newDiff_v2	--	Folder
pop1.dat	389 bytes	Komodo
runtumble	--	Fold
simMotorsFast.mexa64	8 KB	MATLAB
simXpOU_Fast.mexa64	8 KB	MATLAB
varyBias_tau15_lin	--	Folder
varyNoise_tau15_lin	--	Fold
varyNoise_tau30_exp	--	Folder
varyNoise_tau30_lin_rotDiffZero	--	Folder
flagella	--	Folder
distributionOf_DL_over_DT.pdf	8 KB	Ado
doseResponse	--	Folde
doseResponse.dat	4.8 MB	Komo
doseResponse.txt	153 bytes	Text D
runDoseResponse	117 bytes	Unix E
fig4_a_trace_distribution.eps	8 KB	Encaps
fig4_a_trace_distribution.pdf	3 KB	Adobe
forage	--	Folder
forage2	--	Folder
forage3	--	Folder
forage4	--	Folder
gradient	--	Folder
longData.mat	114.6 MB	MATLAB
mf_analytical	--	Folder
analyzeAllMotorSeriesFast.m	3 KB	MATLAB
bias_fig.eps	14 KB	Encapsu
bias_fig.pdf	9 KB	Adobe
bias2_fig.eps	18 KB	Encapsu
bias2_fig.pdf	11 KB	Adobe P

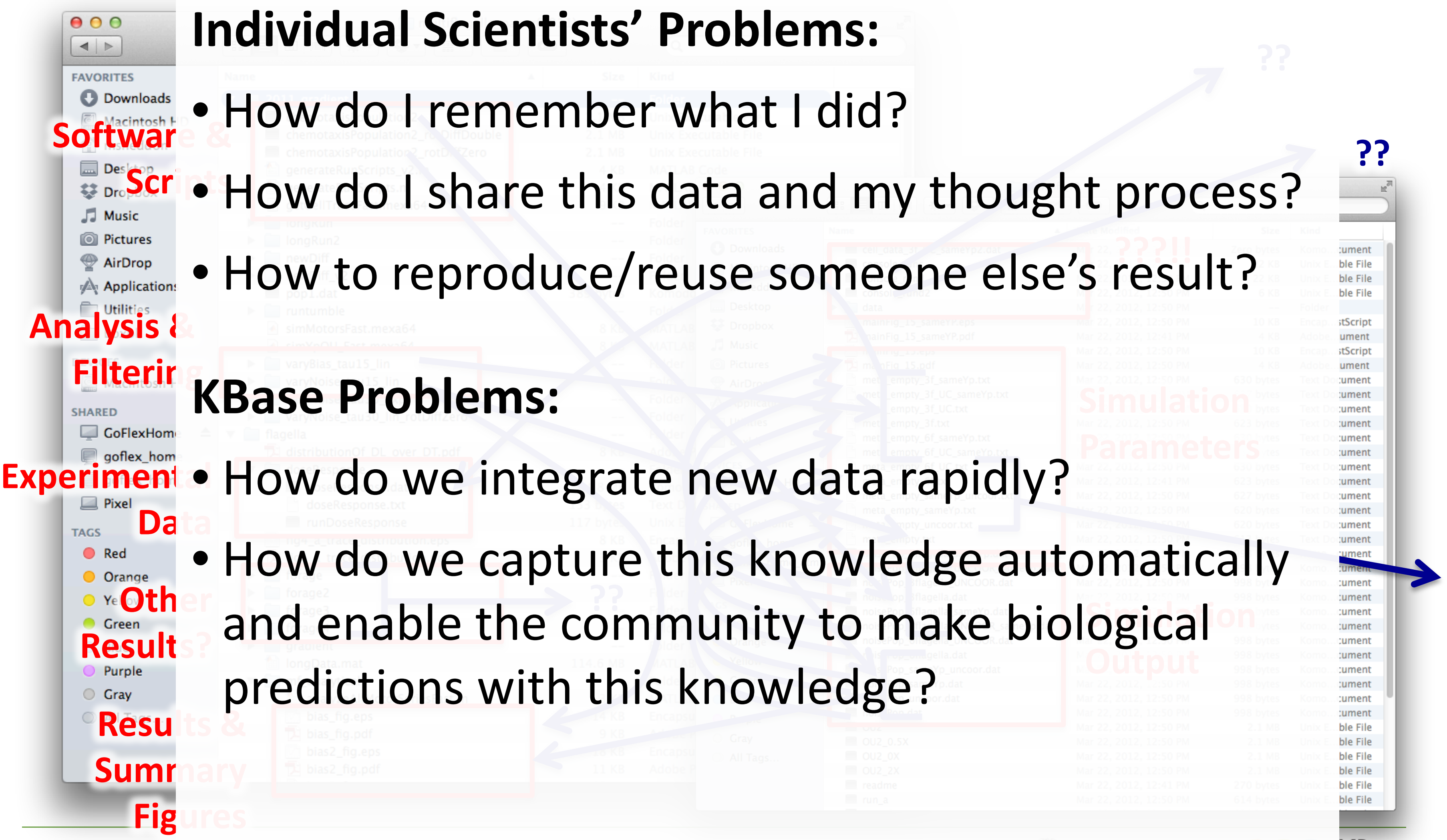
Name	Date Modified	Size	Kind
cell_data_3f C_sameYp2.dat	Mar 22,	Zero bytes	Komo…cument
console unc	Mar 22,	22 KB	Unix E…ble File
console_rund	Mar 22, 2012, 12:50 PM	22 KB	Unix E…ble File
console_run	Mar 22, 2012, 12:50 PM	6 KB	Unix E…ble File
data	Mar 22, 2012, 12:50 PM	--	Folder
mainFig_15_sameYP.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap…stScript
mainFig_15_sameYP.pdf	Mar 22, 2012, 12:41 PM	4 KB	Adobe…ument
mainFig_15.eps	Mar 22, 2012, 12:50 PM	10 KB	Encap…stScript
mainFig_15.pdf	Mar 22, 2012, 12:50 PM	4 KB	Adobe…ument
met _empty_3f_sameYp.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
met _empty_3f_UC_sameYp.txt	[illegible]	bytes	Text Document
_empty_3f_UC.txt	[illegible]	bytes	Text Document
met _empty_3f.txt	Mar 22, 2012, 12:50 PM	623 bytes	Text Document
met _empty_6f_sameYp.txt	[illegible]	630 bytes	Text Document
met _empty_6f_UC_sameYp.txt	[illegible]	ytes	Text Document
meta_empty_6f_UC.txt	Mar 22, 2012, 12:50 PM	630 bytes	Text Document
meta_empty_6f.txt	Mar 22, 2012, 12:41 PM	623 bytes	Text Document
meta_empty_sameYp_uncoor.txt	Mar 22, 2012, 12:50 PM	627 bytes	Text Document
meta_empty_sameYp.txt	Mar 22, 2012, 12:50 PM	620 bytes	Text Document
mpty_uncoor.txt	Mar 22,	620 bytes	Text Document
met empty.txt	Mar 22, 2012, 12:50 PM	bytes	Text Document
[illegible]	Mar 22, 2012, 12:50 PM	998 bytes	cument
[illegible]	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
Pop flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
op flagella.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
noisePop flagella_sameYp.dat	[illegible]	bytes	Komo…cument
noisePop flagella_UNCOOR_sameYp.dat	[illegible]	bytes	Komo…cument
noisePop_6flagella_UNCOOR.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
agella.dat	[illegible]	998 bytes	Komo…cument
Pop_sameYp_uncoor.dat	[illegible]	998 bytes	Komo…cument
nois Pop_sameYp.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
Pop_uncoor.dat	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
nois	Mar 22, 2012, 12:50 PM	998 bytes	Komo…cument
OU2	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_0.5X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_0X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
OU2_2X	Mar 22, 2012, 12:50 PM	2.1 MB	Unix E…ble File
readme	Mar 22, 2012, 12:41 PM	270 bytes	Unix E…ble File
run_a	Mar 22, 2012, 12:50 PM	614 bytes	Unix E…ble File

Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY
Office of Science

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

The Data Problem in Systems Biology

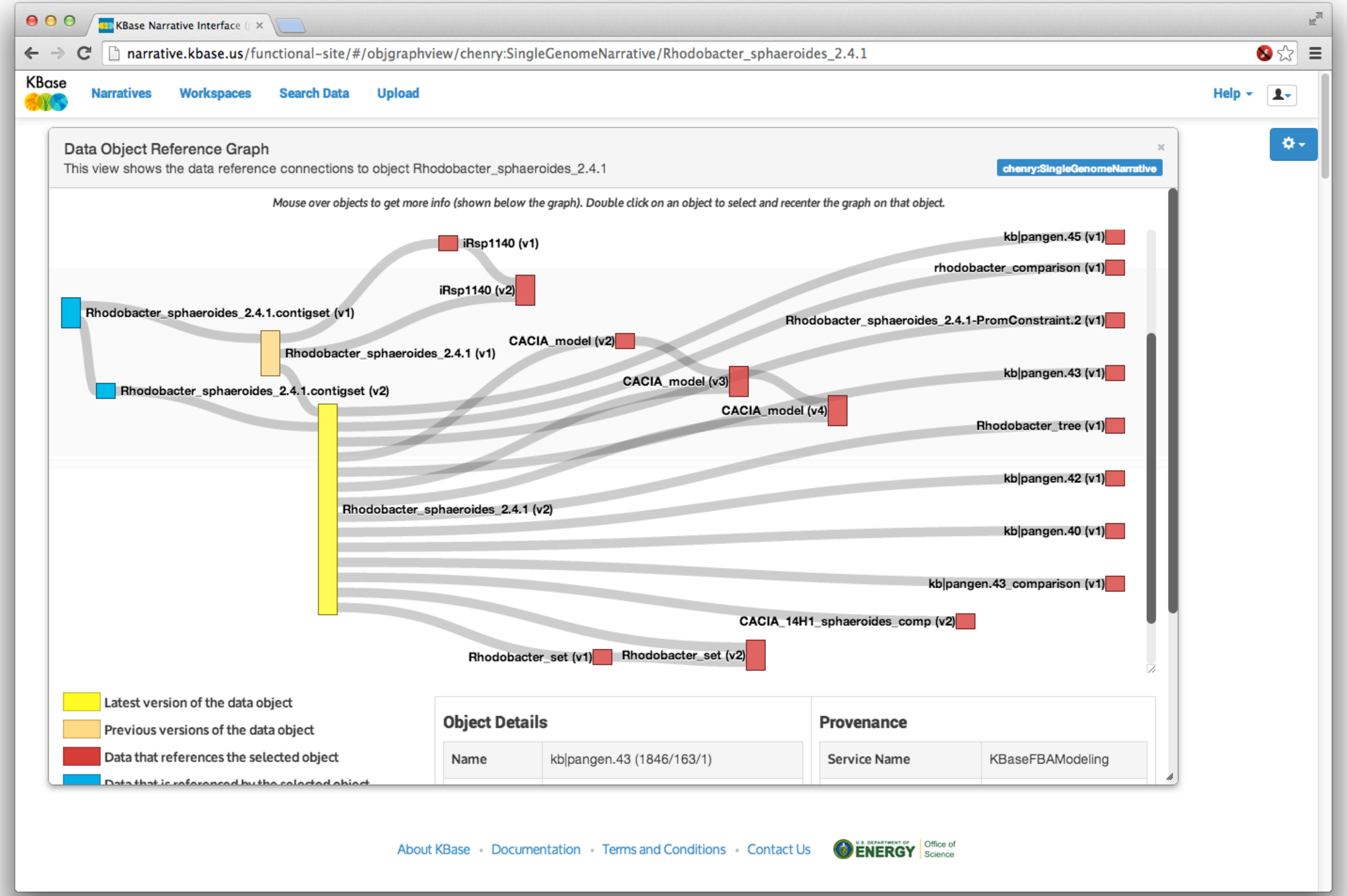


Individual Scientists' Problems:

- How do I remember what I did?
- How do I share this data and my thought process?
- How to reproduce/reuse someone else's result?

KBase Problems:

- How do we integrate new data rapidly?
- How do we capture this knowledge automatically and enable the community to make biological predictions with this knowledge?



Data Object Reference Graph

This view shows the data reference connections to object Rhodobacter_sphaeroides_2.4.1

Mouse over objects to get more info (shown below the graph). Double click on an object to select and recenter the graph on that object.

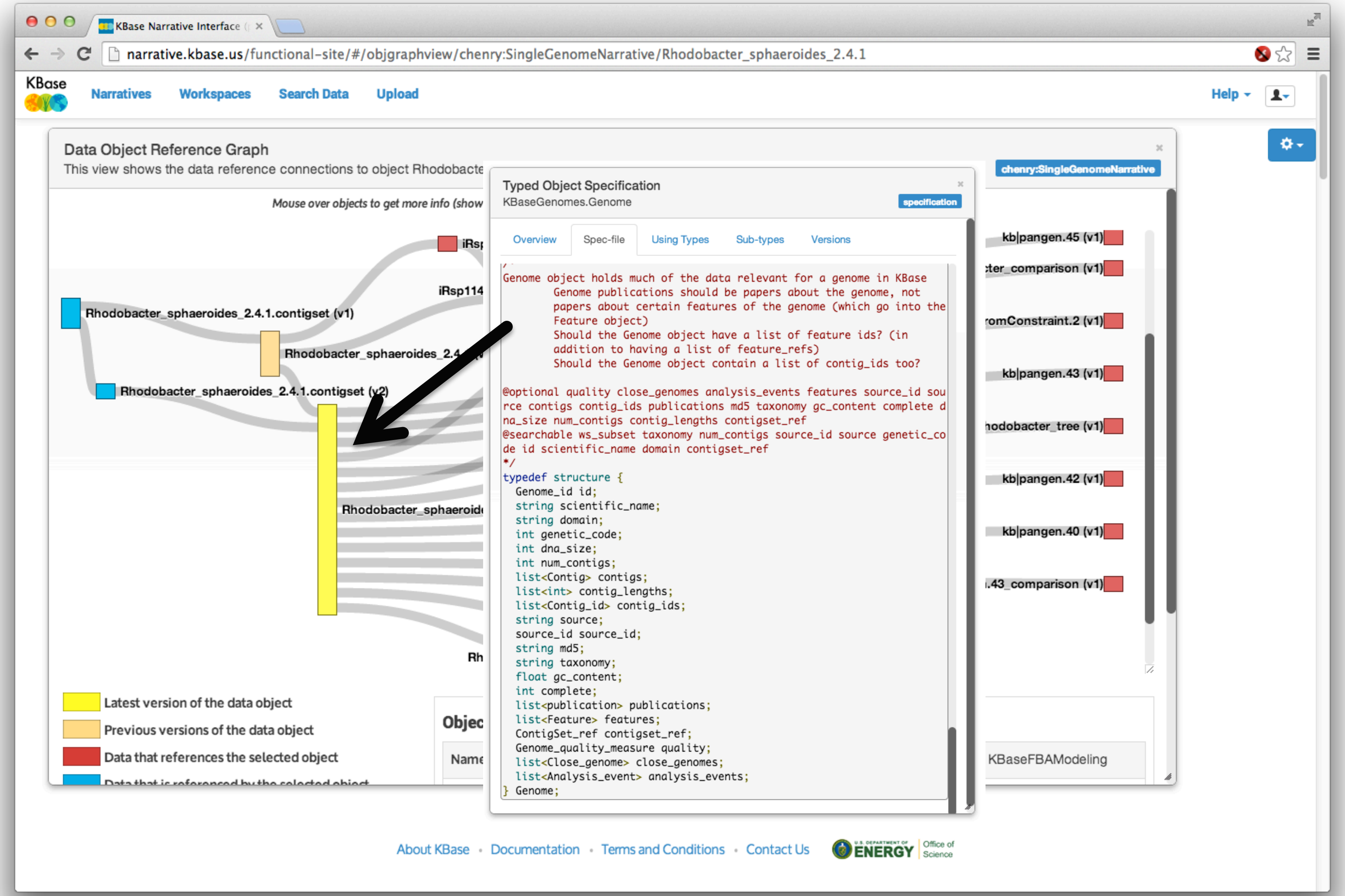
- 🟨 Latest version of the data object
- 🟧 Previous versions of the data object
- 🟥 Data that references the selected object
- 🟦 Data that is referenced by the selected object

Object Details

Name	kb|pangen.43 (1846/163/1)

Provenance

Service Name	KBaseFBAModeling



Data Object Reference Graph

This view shows the data reference connections to object Rhodobacte

Mouse over objects to get more info (show

🟨 Latest version of the data object

🟧 Previous versions of the data object

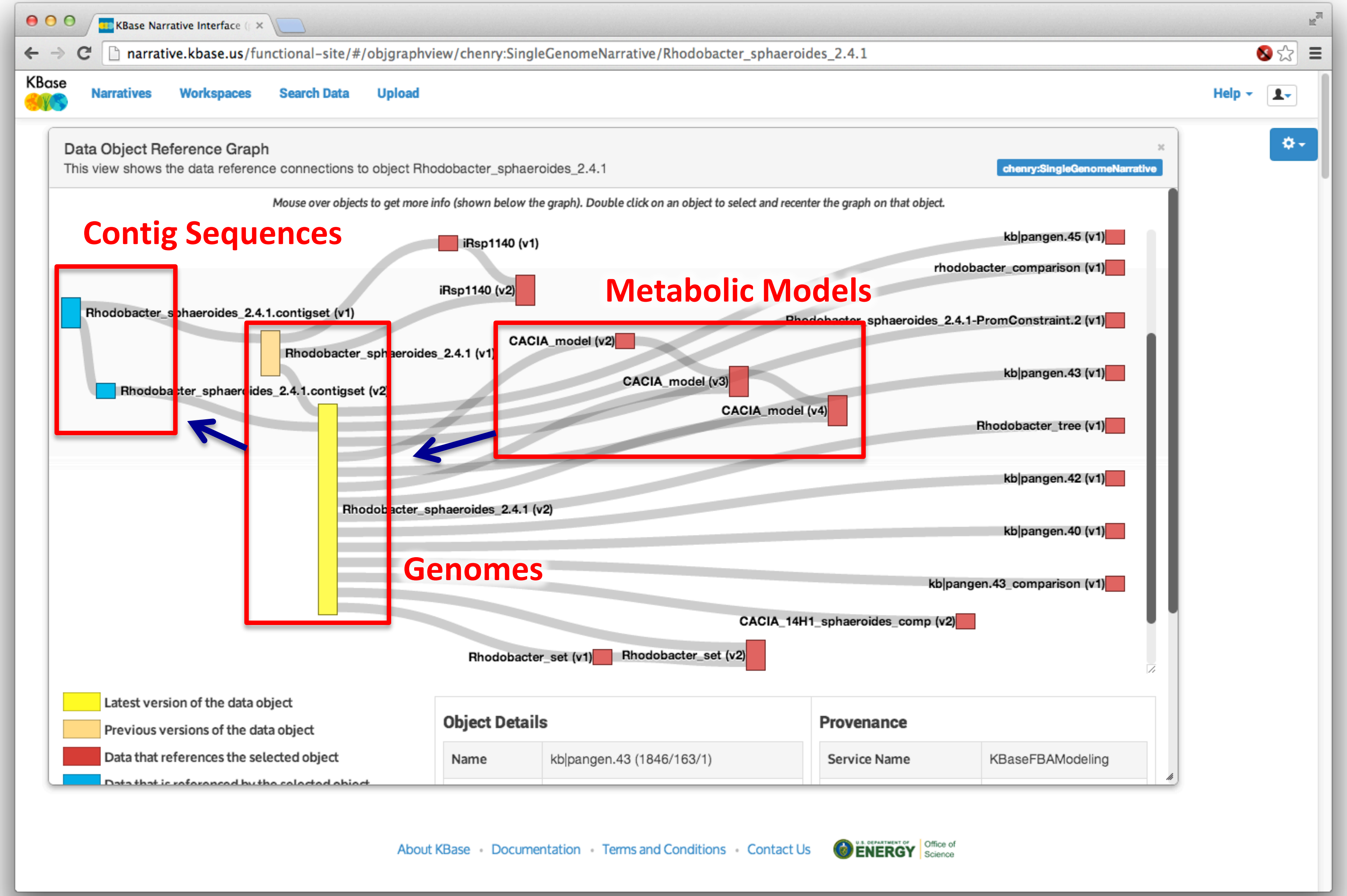
🟥 Data that references the selected object

🟦 Data that is referenced by the selected object

Typed Object Specification

KBaseGenomes.Genome

```
Genome object holds much of the data relevant for a genome in KBase
        Genome publications should be papers about the genome, not
        papers about certain features of the genome (which go into the
        Feature object)
        Should the Genome object have a list of feature ids? (in
        addition to having a list of feature_refs)
        Should the Genome object contain a list of contig_ids too?

@optional quality close_genomes analysis_events features source_id sou
rce contigs contig_ids publications md5 taxonomy gc_content complete d
na_size num_contigs contig_lengths contigset_ref
@searchable ws_subset taxonomy num_contigs source_id source genetic_co
de id scientific_name domain contigset_ref
*/
typedef structure {
  Genome_id id;
  string scientific_name;
  string domain;
  int genetic_code;
  int dna_size;
  int num_contigs;
  list<Contig> contigs;
  list<int> contig_lengths;
  list<Contig_id> contig_ids;
  string source;
  source_id source_id;
  string md5;
  string taxonomy;
  float gc_content;
  int complete;
  list<publication> publications;
  list<Feature> features;
  ContigSet_ref contigset_ref;
  Genome_quality_measure quality;
  list<Close_genome> close_genomes;
  list<Analysis_event> analysis_events;
} Genome;
```



Data Object Reference Graph

This view shows the data reference connections to object Rhodobacter_sphaeroides_2.4.1

Mouse over objects to get more info (shown below the graph). Double click on an object to select and recenter the graph on that object.

🟨 Latest version of the data object

🟨 Previous versions of the data object

🟥 Data that references the selected object

🟦 Data that is referenced by the selected object

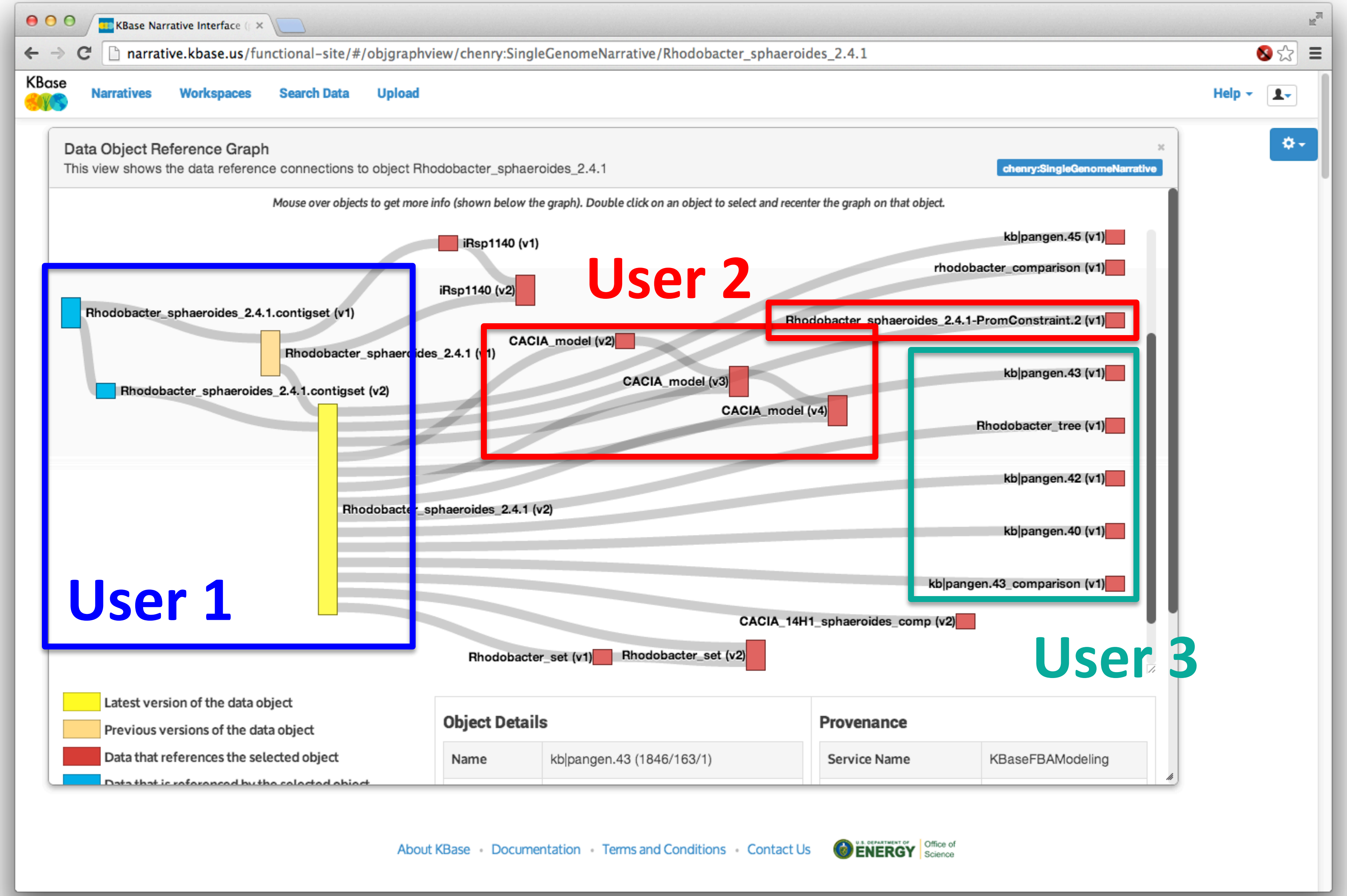
Object Details

Name	kb|pangen.43 (1846/163/1)

Provenance

Service Name	KBaseFBAModeling

About KBase · Documentation · Terms and Conditions · Contact Us



Data Object Reference Graph

This view shows the data reference connections to object Rhodobacter_sphaeroides_2.4.1

Mouse over objects to get more info (shown below the graph). Double click on an object to select and recenter the graph on that object.

🟨 Latest version of the data object

🟧 Previous versions of the data object

🟥 Data that references the selected object

🟦 Data that is referenced by the selected object

Object Details

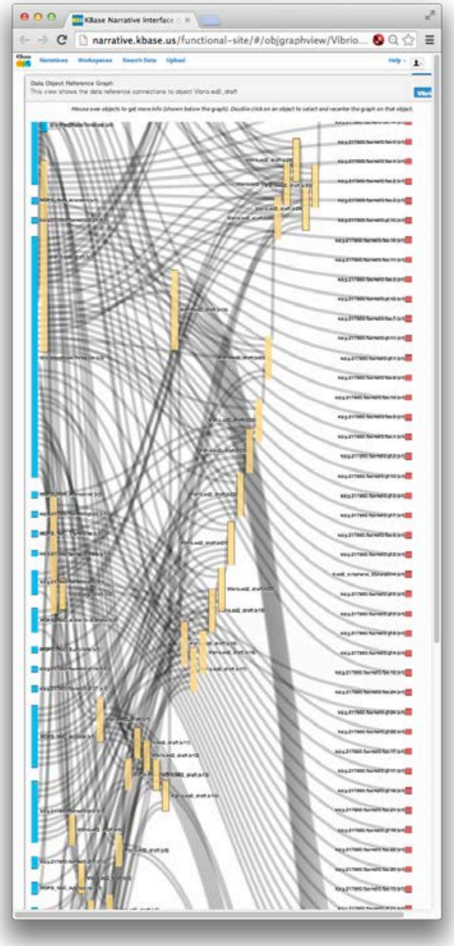
Name	kb|pangen.43 (1846/163/1)

Provenance

Service Name	KBaseFBAModeling

About KBase · Documentation · Terms and Conditions · Contact Us

The KBase Data Platform



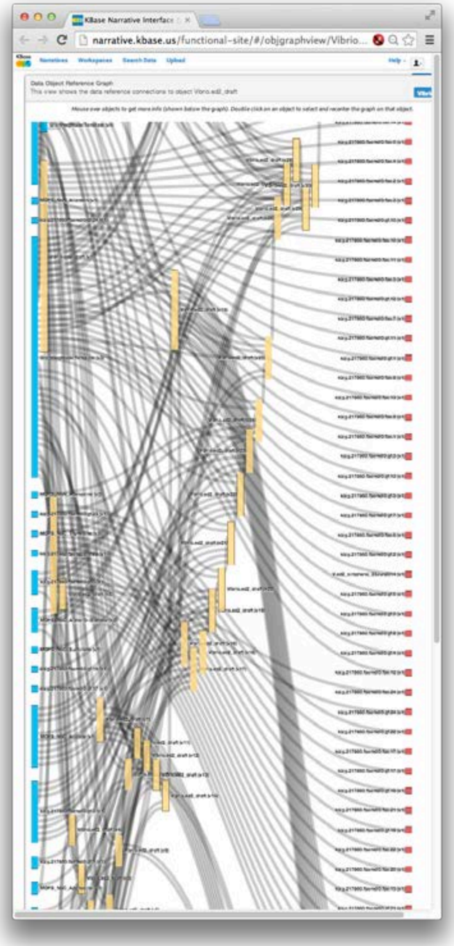
Data Object Reference Graph

This view shows the data reference connections to object Vibrio.ed2_draft

Mouse over objects to get more info (shown below the graph). Double click on an object to select and recenter the graph on that object.

- Rich, extensible, underlying data model for reference and user data
- Data objects involved in computation are typed, versioned, validated, persistent and cross-referenced automatically based on the data model.
- The "data object network" is social

The KBase Data Platform

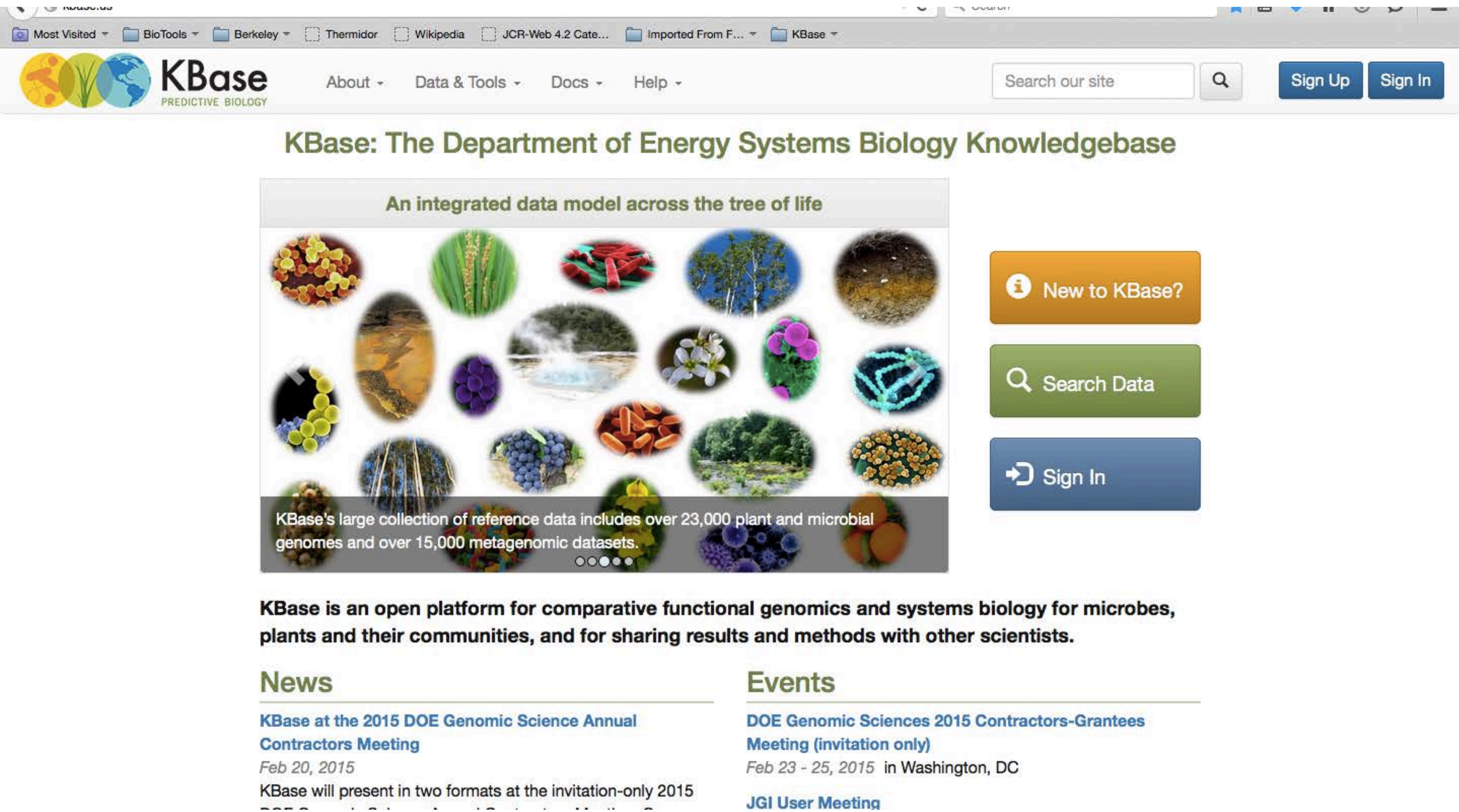


Data Object Reference Graph

This view shows the data reference connections to object Vibrio.ed2_draft

Mouse over objects to get more info (shown below the graph). Double click on an object to select and recenter the graph on that object.

- Not just an organizational tool for the user!
- Relationships between data objects imply biological assertions.
- KBase is making these assertions computable, which is the first step towards extracting new knowledge across the entire system.



KBase: The Department of Energy Systems Biology Knowledgebase

An integrated data model across the tree of life

KBase's large collection of reference data includes over 23,000 plant and microbial genomes and over 15,000 metagenomic datasets.

KBase is an open platform for comparative functional genomics and systems biology for microbes, plants and their communities, and for sharing results and methods with other scientists.

News

KBase at the 2015 DOE Genomic Science Annual Contractors Meeting

Feb 20, 2015

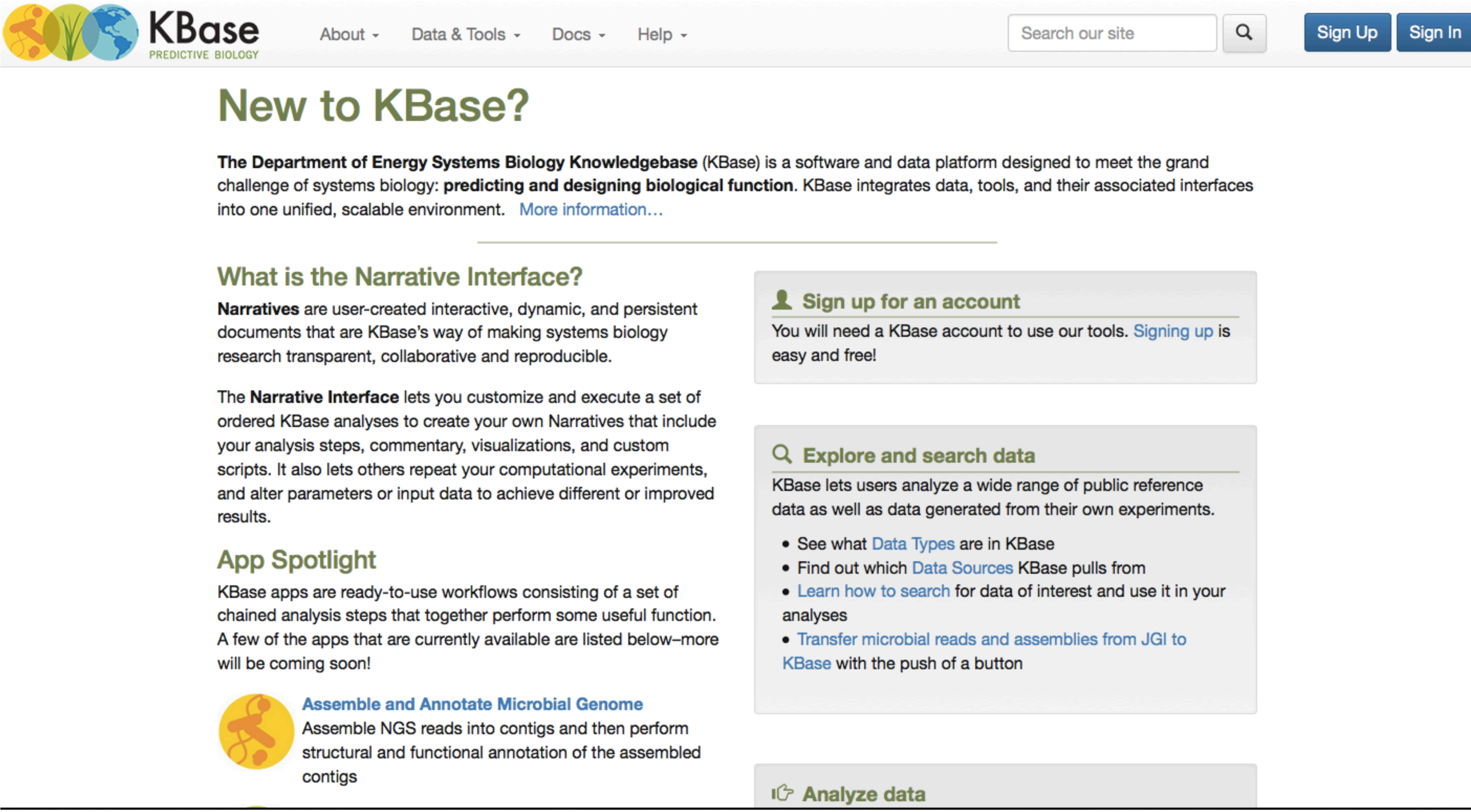
KBase will present in two formats at the invitation-only 2015

Events

DOE Genomic Sciences 2015 Contractors-Grantees Meeting (invitation only)

Feb 23 - 25, 2015 in Washington, DC

JGI User Meeting



New to KBase?

The Department of Energy Systems Biology Knowledgebase (KBase) is a software and data platform designed to meet the grand challenge of systems biology: **predicting and designing biological function**. KBase integrates data, tools, and their associated interfaces into one unified, scalable environment. More information…

What is the Narrative Interface?

Narratives are user-created interactive, dynamic, and persistent documents that are KBase's way of making systems biology research transparent, collaborative and reproducible.

The **Narrative Interface** lets you customize and execute a set of ordered KBase analyses to create your own Narratives that include your analysis steps, commentary, visualizations, and custom scripts. It also lets others repeat your computational experiments, and alter parameters or input data to achieve different or improved results.

App Spotlight

KBase apps are ready-to-use workflows consisting of a set of chained analysis steps that together perform some useful function. A few of the apps that are currently available are listed below–more will be coming soon!

Assemble and Annotate Microbial Genome

Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs

Sign up for an account

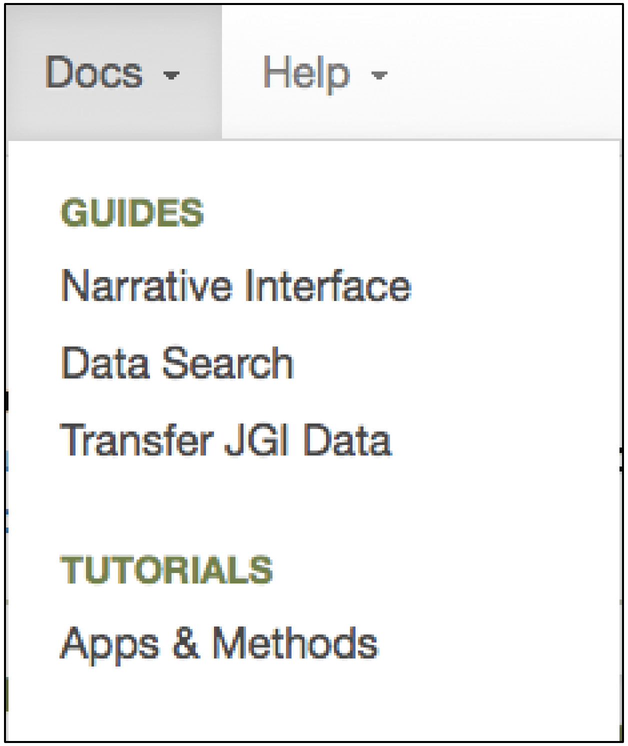
You will need a KBase account to use our tools. Signing up is easy and free!

Explore and search data

KBase lets users analyze a wide range of public reference data as well as data generated from their own experiments.

- See what Data Types are in KBase
- Find out which Data Sources KBase pulls from
- Learn how to search for data of interest and use it in your analyses
- Transfer microbial reads and assemblies from JGI to KBase with the push of a button

Analyze data


Docs
Help
GUIDES
Narrative Interface
Data Search
Transfer JGI Data
TUTORIALS
Apps & Methods

Tutorials for Apps and Methods

KBase's Narrative Interface enables researchers to design and carry out computational experiments while creating interactive, reproducible records of the data, steps, and thought processes underpinning their results. These records can be shared with collaborators and published as "active papers" called Narratives that let others repeat the computational experiment and even alter parameters or input data to achieve different or improved results.

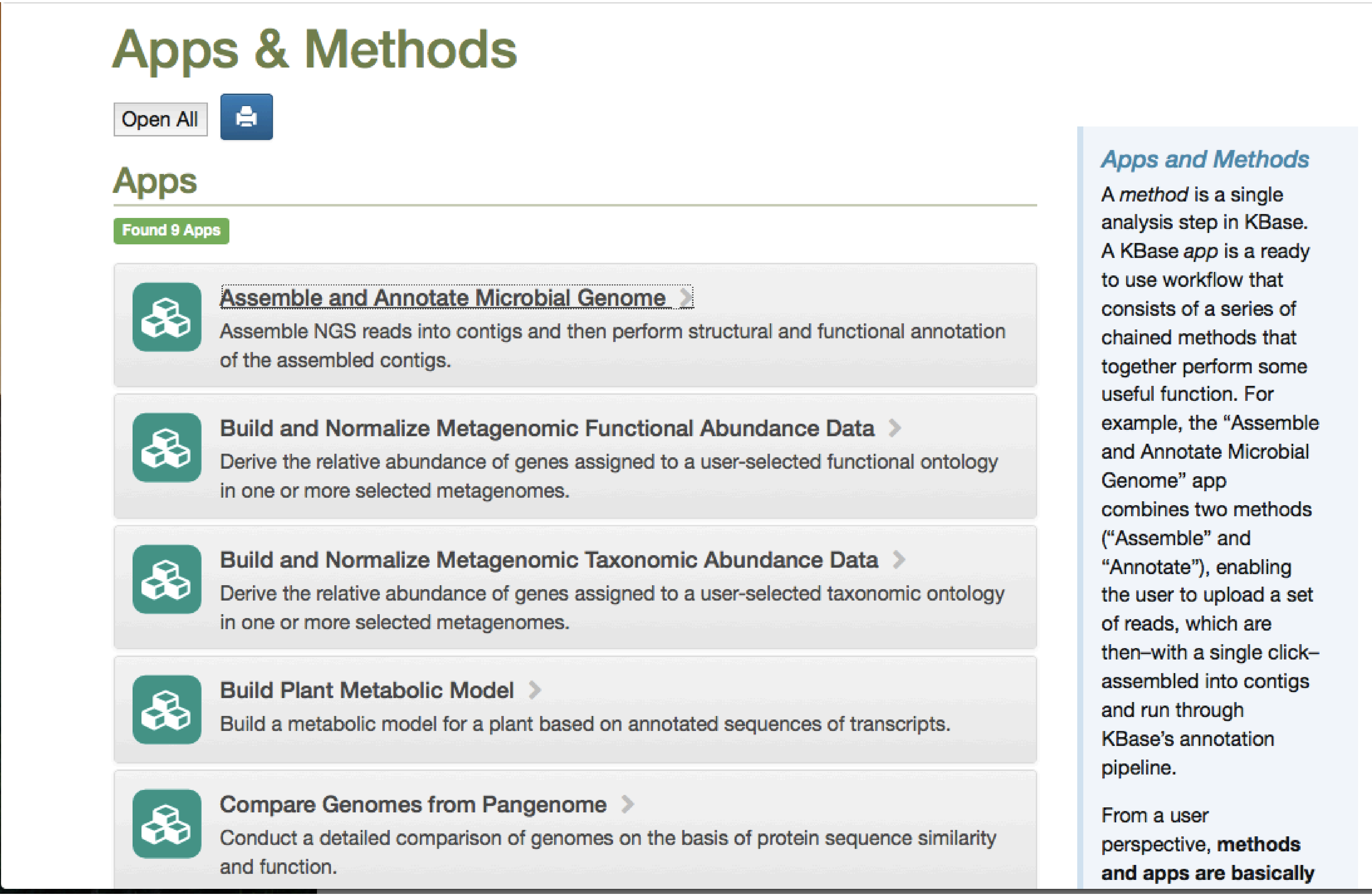
Narratives can include any number of analysis steps in the form of apps, methods, or a combination of the two. A KBase method is a single analysis step (e.g., "run flux balance analysis" or "compare two pangenomes"), whereas an app consists of a series of chained methods that together perform some useful function. For example, the *Assemble and Annotate Microbial Genome* app combines two methods that are often useful together. Apps and methods can be used interchangeably in your Narrative.

All of the apps, and many of the methods, have tutorials that provide step-by-step examples of how they can be used to analyze your data (or example data available in the system). The current set of tutorials is listed here—more will be added soon.

Also see the Narrative Interface User Guide for detailed instructions on working with Narratives.

App and Method Tutorials

- Annotate Microbial Contigs Method
- Annotate Microbial Genome Method
- Annotate Plant Transcripts with Metabolic Functions Method
- Assemble and Annotate Microbial Genome
- Build Plant Metabolic Model App
- Build and Normalize Metagenomic Functional Abundance Data App
- Build and Normalize Metagenomic Taxonomic Abundance Data App
- Compare Genomes from Pangenome
- Compare Two Metabolic Models Method
- Compare Two Proteomes Method
- Generate Boxplots for Abundance Data Method
- Generate Functional Abundance Data Method
- Generate PCoA from Abundance Data Method
- Generate Taxonomic Abundance Data Method
- Insert Genomes into Species Tree
- Normalize Abundance Data Method
- Propagate Genome-scale Model to Close Genome App
- Reconstruct Community Metabolic Model
- Reconstruct Genome-scale Metabolic Model
- Run Flux Balance Analysis Method
- Simulate Growth on Phenotype Data Method



Apps & Methods

Apps

Found 9 Apps

Assemble and Annotate Microbial Genome
Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs.

Build and Normalize Metagenomic Functional Abundance Data
Derive the relative abundance of genes assigned to a user-selected functional ontology in one or more selected metagenomes.

Build and Normalize Metagenomic Taxonomic Abundance Data
Derive the relative abundance of genes assigned to a user-selected taxonomic ontology in one or more selected metagenomes.

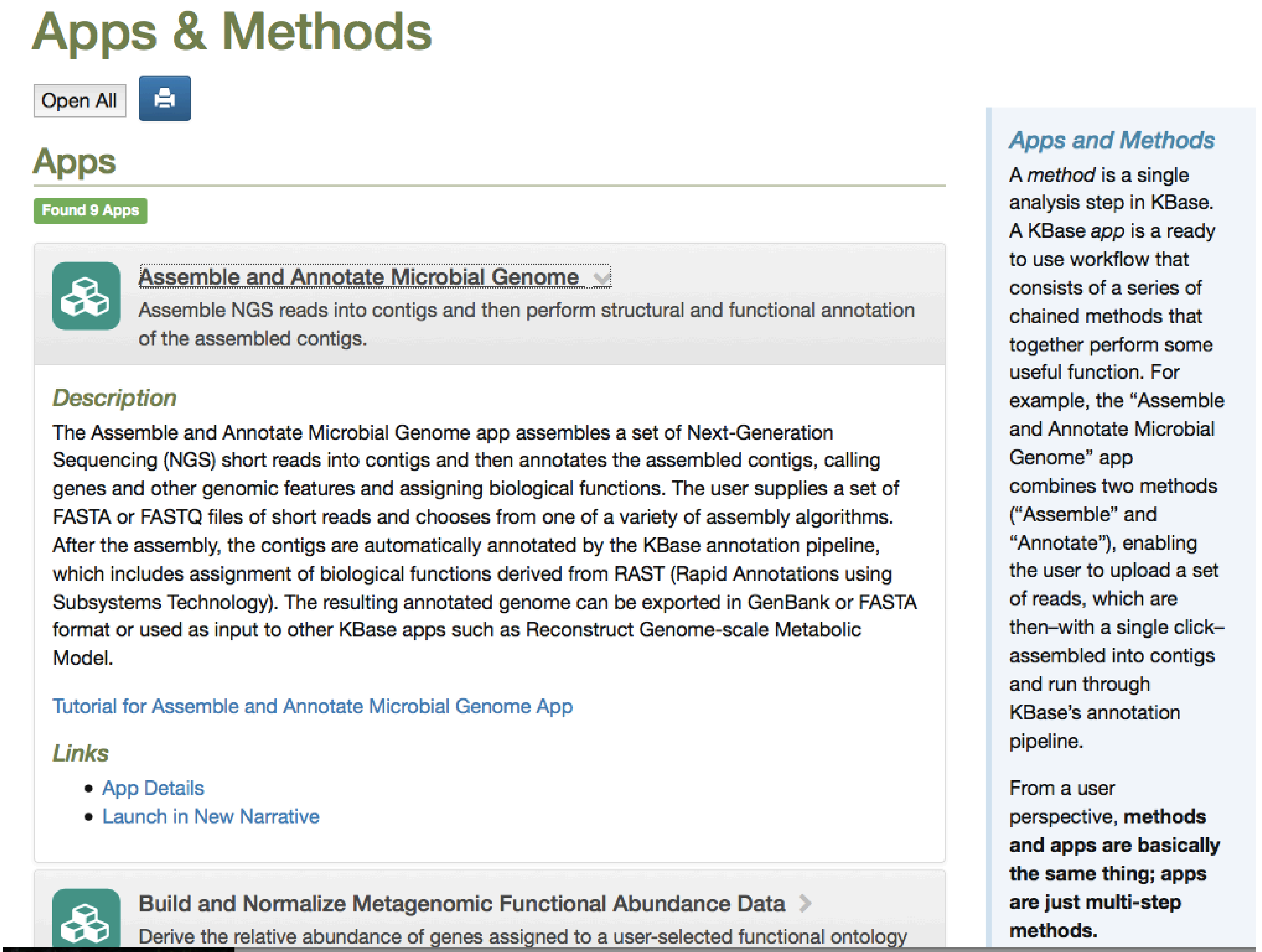
Build Plant Metabolic Model
Build a metabolic model for a plant based on annotated sequences of transcripts.

Compare Genomes from Pangenome
Conduct a detailed comparison of genomes on the basis of protein sequence similarity and function.

Apps and Methods

A *method* is a single analysis step in KBase. A KBase *app* is a ready to use workflow that consists of a series of chained methods that together perform some useful function. For example, the "Assemble and Annotate Microbial Genome" app combines two methods ("Assemble" and "Annotate"), enabling the user to upload a set of reads, which are then–with a single click–assembled into contigs and run through KBase's annotation pipeline.

From a user perspective, **methods and apps are basically**



Apps & Methods

Apps

Found 9 Apps

Assemble and Annotate Microbial Genome

Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs.

Description

The Assemble and Annotate Microbial Genome app assembles a set of Next-Generation Sequencing (NGS) short reads into contigs and then annotates the assembled contigs, calling genes and other genomic features and assigning biological functions. The user supplies a set of FASTA or FASTQ files of short reads and chooses from one of a variety of assembly algorithms. After the assembly, the contigs are automatically annotated by the KBase annotation pipeline, which includes assignment of biological functions derived from RAST (Rapid Annotations using Subsystems Technology). The resulting annotated genome can be exported in GenBank or FASTA format or used as input to other KBase apps such as Reconstruct Genome-scale Metabolic Model.

Tutorial for Assemble and Annotate Microbial Genome App

Links

- App Details
- Launch in New Narrative

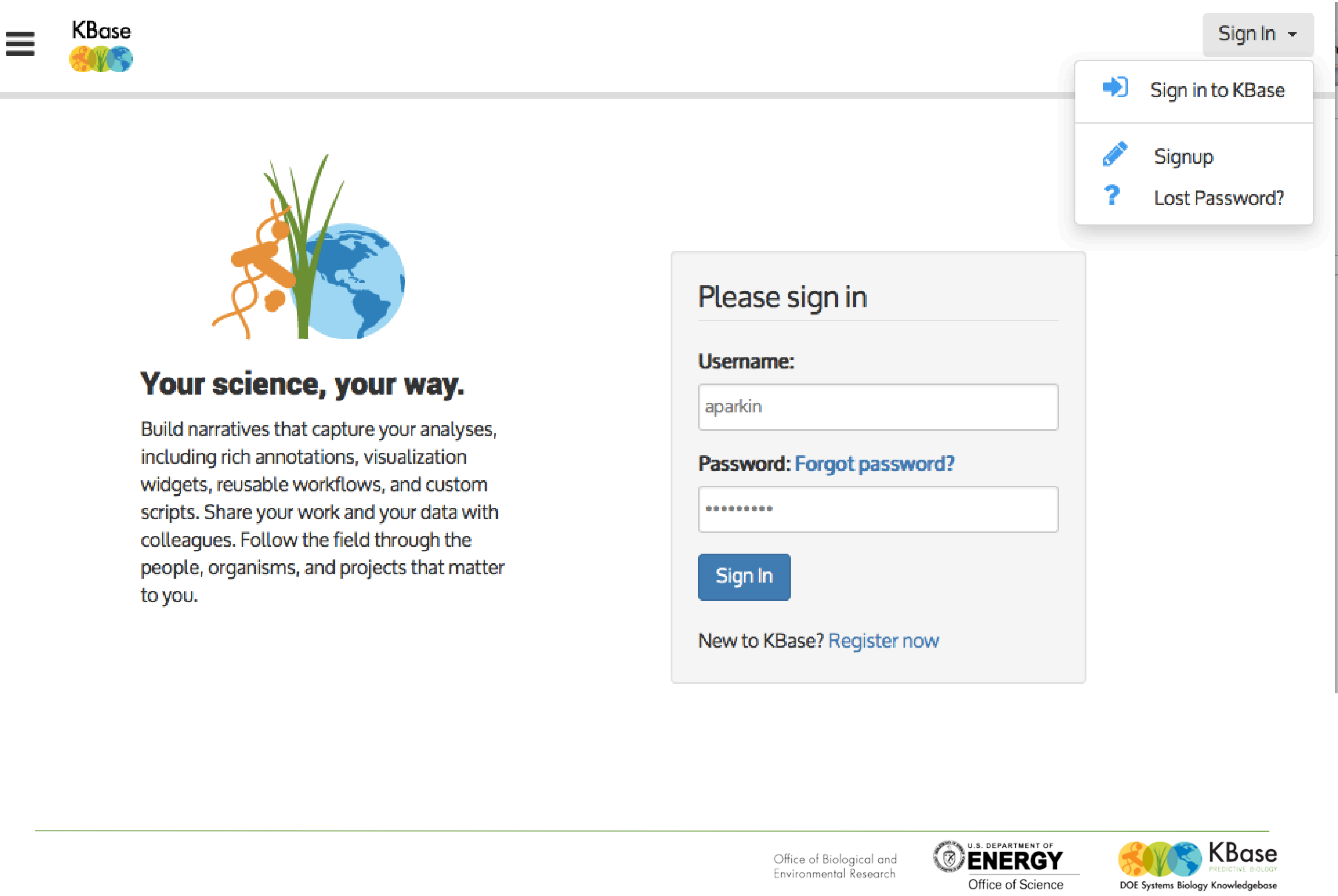
Build and Normalize Metagenomic Functional Abundance Data

Derive the relative abundance of genes assigned to a user-selected functional ontology

Apps and Methods

A *method* is a single analysis step in KBase. A KBase *app* is a ready to use workflow that consists of a series of chained methods that together perform some useful function. For example, the “Assemble and Annotate Microbial Genome” app combines two methods (“Assemble” and “Annotate”), enabling the user to upload a set of reads, which are then–with a single click–assembled into contigs and run through KBase’s annotation pipeline.

From a user perspective, **methods and apps are basically the same thing; apps are just multi-step methods.**



Your science, your way.

Build narratives that capture your analyses, including rich annotations, visualization widgets, reusable workflows, and custom scripts. Share your work and your data with colleagues. Follow the field through the people, organisms, and projects that matter to you.

Please sign in

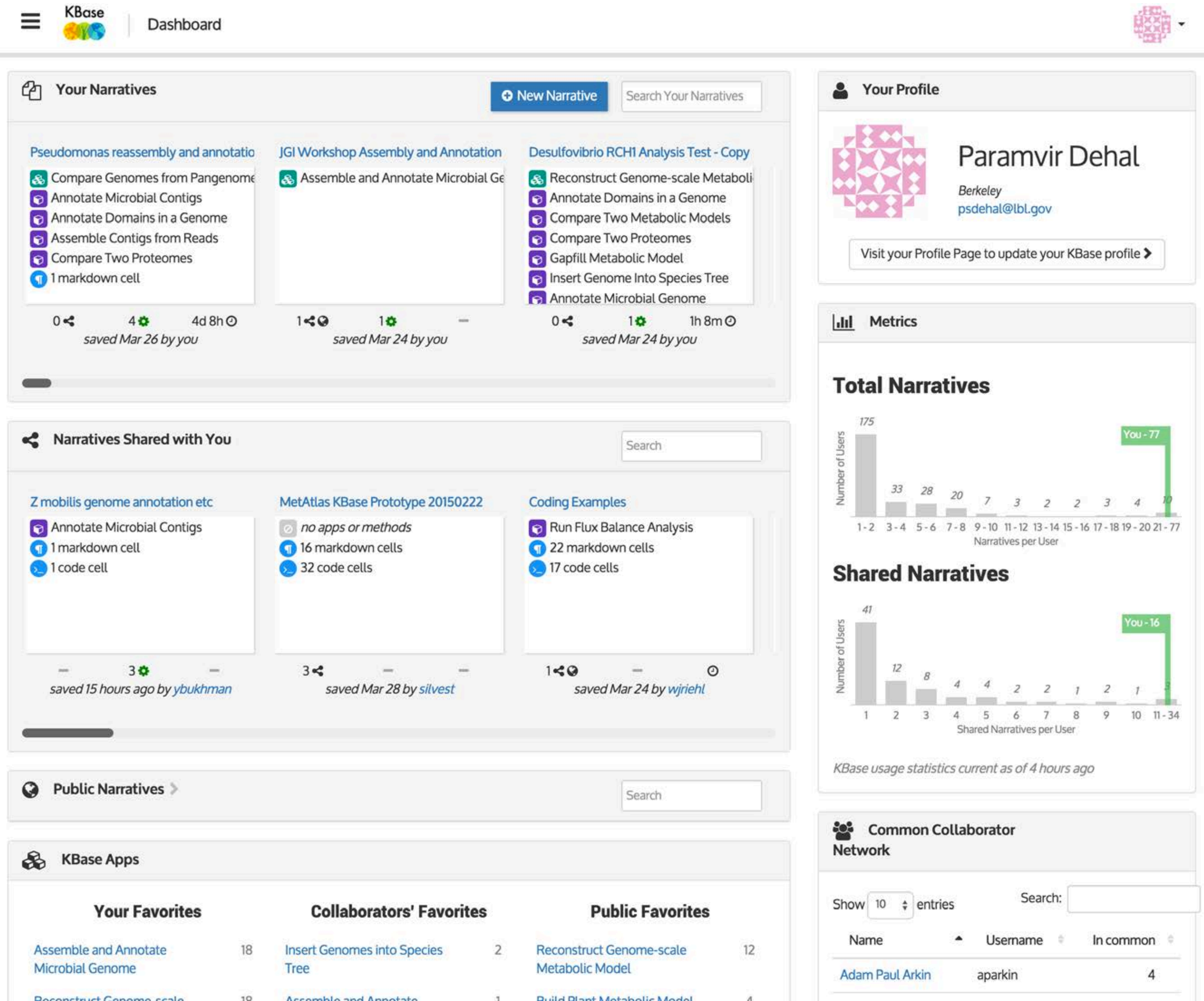
Username: aparkin

Password: Forgot password?

••••••••••

Sign In

New to KBase? Register now



KBase Dashboard

Your Narratives

Pseudomonas reassembly and annotatio	JGI Workshop Assembly and Annotation	Desulfovibrio RCH1 Analysis Test - Copy
Compare Genomes from Pangenome Annotate Microbial Contigs Annotate Domains in a Genome Assemble Contigs from Reads Compare Two Proteomes 1 markdown cell	Assemble and Annotate Microbial Ge	Reconstruct Genome-scale Metaboli Annotate Domains in a Genome Compare Two Metabolic Models Compare Two Proteomes Gapfill Metabolic Model Insert Genome Into Species Tree Annotate Microbial Genome
0 4 4d 8h	1 1 —	0 1 1h 8m
saved Mar 26 by you	*saved Mar 24 by you*	*saved Mar 24 by you*

Narratives Shared with You

Z mobilis genome annotation etc	MetAtlas KBase Prototype 20150222	Coding Examples
Annotate Microbial Contigs 1 markdown cell 1 code cell	*no apps or methods* 16 markdown cells 32 code cells	Run Flux Balance Analysis 22 markdown cells 17 code cells
— 3 —	3 — —	1 —
saved 15 hours ago by ybukhman	*saved Mar 28 by silvest*	*saved Mar 24 by wjriehl*

Public Narratives

KBase Apps

Your Favorites		Collaborators' Favorites		Public Favorites	
Assemble and Annotate Microbial Genome	18	Insert Genomes into Species Tree	2	Reconstruct Genome-scale Metabolic Model	12
Reconstruct Genome-scale	18	Assemble and Annotate	1	Build Plant Metabolic Model	4

Your Profile

Paramvir Dehal

Berkeley

psdehal@lbl.gov

Visit your Profile Page to update your KBase profile

Metrics

Total Narratives

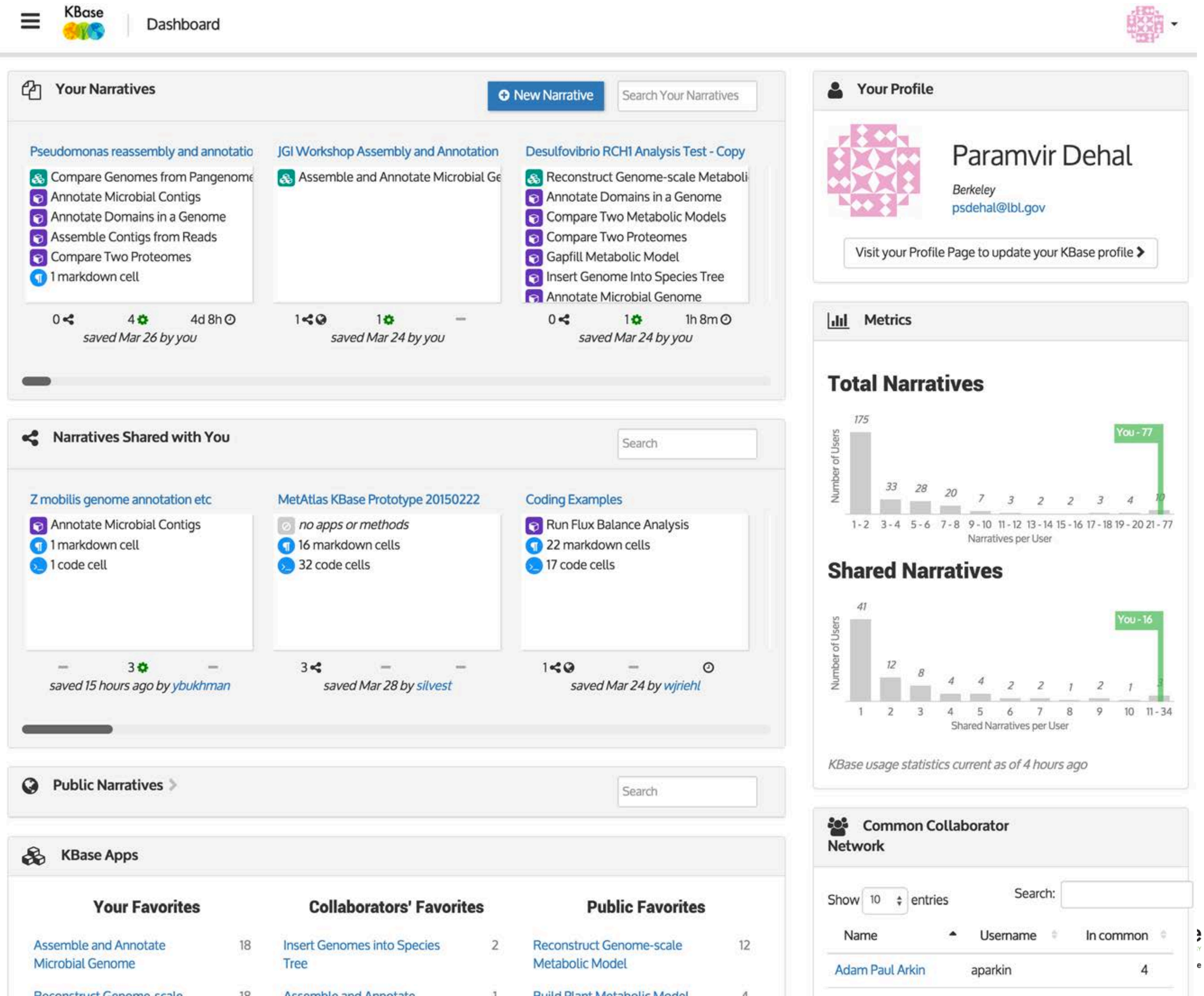
Shared Narratives

KBase usage statistics current as of 4 hours ago

Common Collaborator Network

Show 10 entries Search:

Name	Username	In common
Adam Paul Arkin	aparkin	4



Your Narratives

Pseudomonas reassembly and annotatio
- Compare Genomes from Pangenome
- Annotate Microbial Contigs
- Annotate Domains in a Genome
- Assemble Contigs from Reads
- Compare Two Proteomes
- 1 markdown cell

0 4 4d 8h
saved Mar 26 by you

JGI Workshop Assembly and Annotation
- Assemble and Annotate Microbial Ge

1 1 –
saved Mar 24 by you

Desulfovibrio RCH1 Analysis Test - Copy
- Reconstruct Genome-scale Metaboli
- Annotate Domains in a Genome
- Compare Two Metabolic Models
- Compare Two Proteomes
- Gapfill Metabolic Model
- Insert Genome Into Species Tree
- Annotate Microbial Genome

0 1 1h 8m
saved Mar 24 by you

Narratives Shared with You

Z mobilis genome annotation etc
- Annotate Microbial Contigs
- 1 markdown cell
- 1 code cell

– 3 –
saved 15 hours ago by ybukhman

MetAtlas KBase Prototype 20150222
- no apps or methods
- 16 markdown cells
- 32 code cells

3 – –
saved Mar 28 by silvest

Coding Examples
- Run Flux Balance Analysis
- 22 markdown cells
- 17 code cells

1 –
saved Mar 24 by wjriehl

Public Narratives

KBase Apps

Your Favorites		Collaborators' Favorites		Public Favorites	
Assemble and Annotate Microbial Genome	18	Insert Genomes into Species Tree	2	Reconstruct Genome-scale Metabolic Model	12
Reconstruct Genome-scale	18	Assemble and Annotate	1	Build Plant Metabolic Model	4

Your Profile

Paramvir Dehal

Berkeley

psdehal@lbl.gov

Metrics

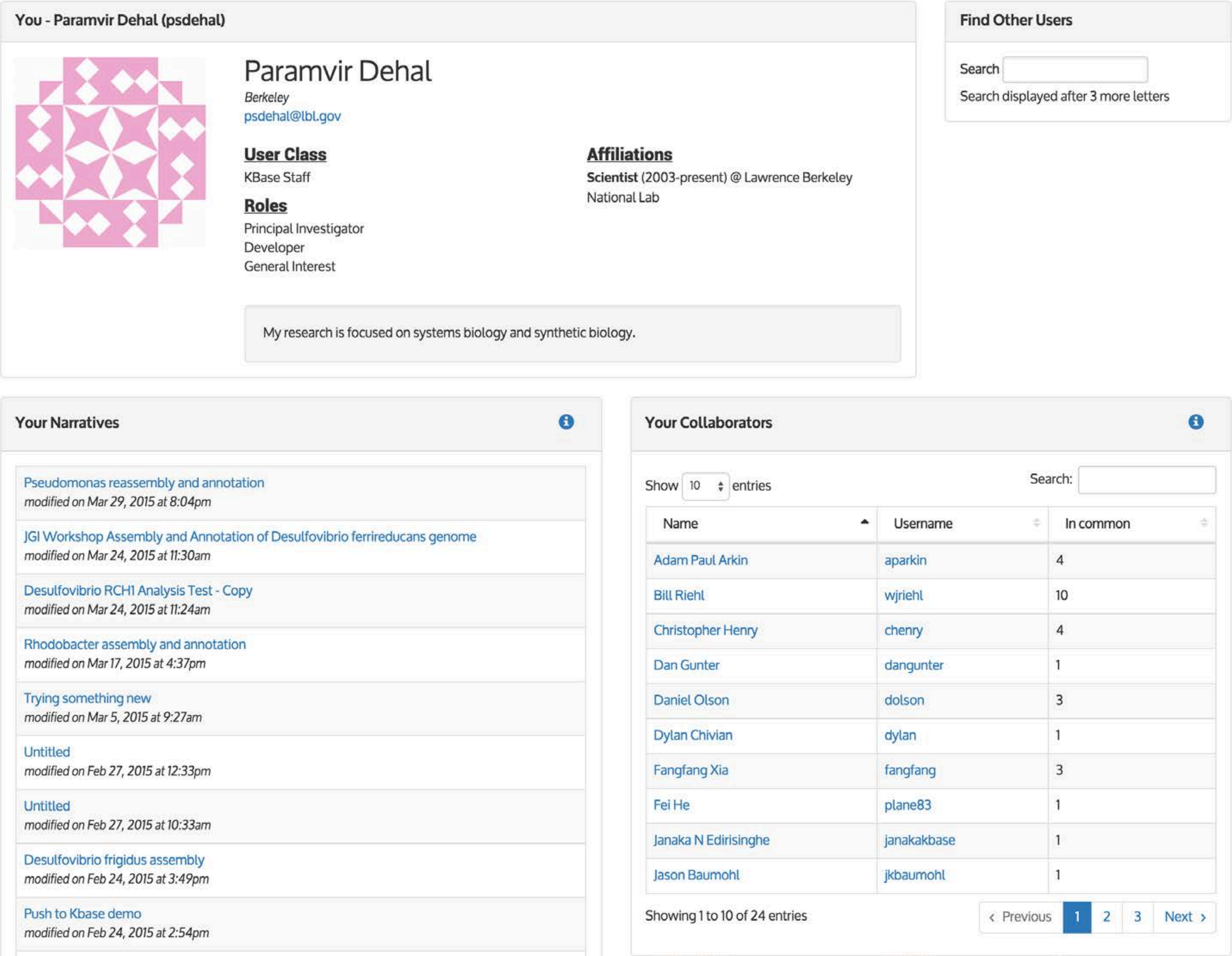
Total Narratives

Shared Narratives

KBase usage statistics current as of 4 hours ago

Common Collaborator Network

Name	Username	In common
Adam Paul Arkin	aparkin	4



You - Paramvir Dehal (psdehal)

Find Other Users

Search

Search displayed after 3 more letters

Paramvir Dehal

Berkeley

psdehal@lbl.gov

User Class

KBase Staff

Roles

Principal Investigator
Developer
General Interest

Affiliations

Scientist (2003-present) @ Lawrence Berkeley National Lab

My research is focused on systems biology and synthetic biology.

Your Narratives

Pseudomonas reassembly and annotation
modified on Mar 29, 2015 at 8:04pm

JGI Workshop Assembly and Annotation of Desulfovibrio ferrireducans genome
modified on Mar 24, 2015 at 11:30am

Desulfovibrio RCH1 Analysis Test - Copy
modified on Mar 24, 2015 at 11:24am

Rhodobacter assembly and annotation
modified on Mar 17, 2015 at 4:37pm

Trying something new
modified on Mar 5, 2015 at 9:27am

Untitled
modified on Feb 27, 2015 at 12:33pm

Untitled
modified on Feb 27, 2015 at 10:33am

Desulfovibrio frigidus assembly
modified on Feb 24, 2015 at 3:49pm

Push to Kbase demo
modified on Feb 24, 2015 at 2:54pm

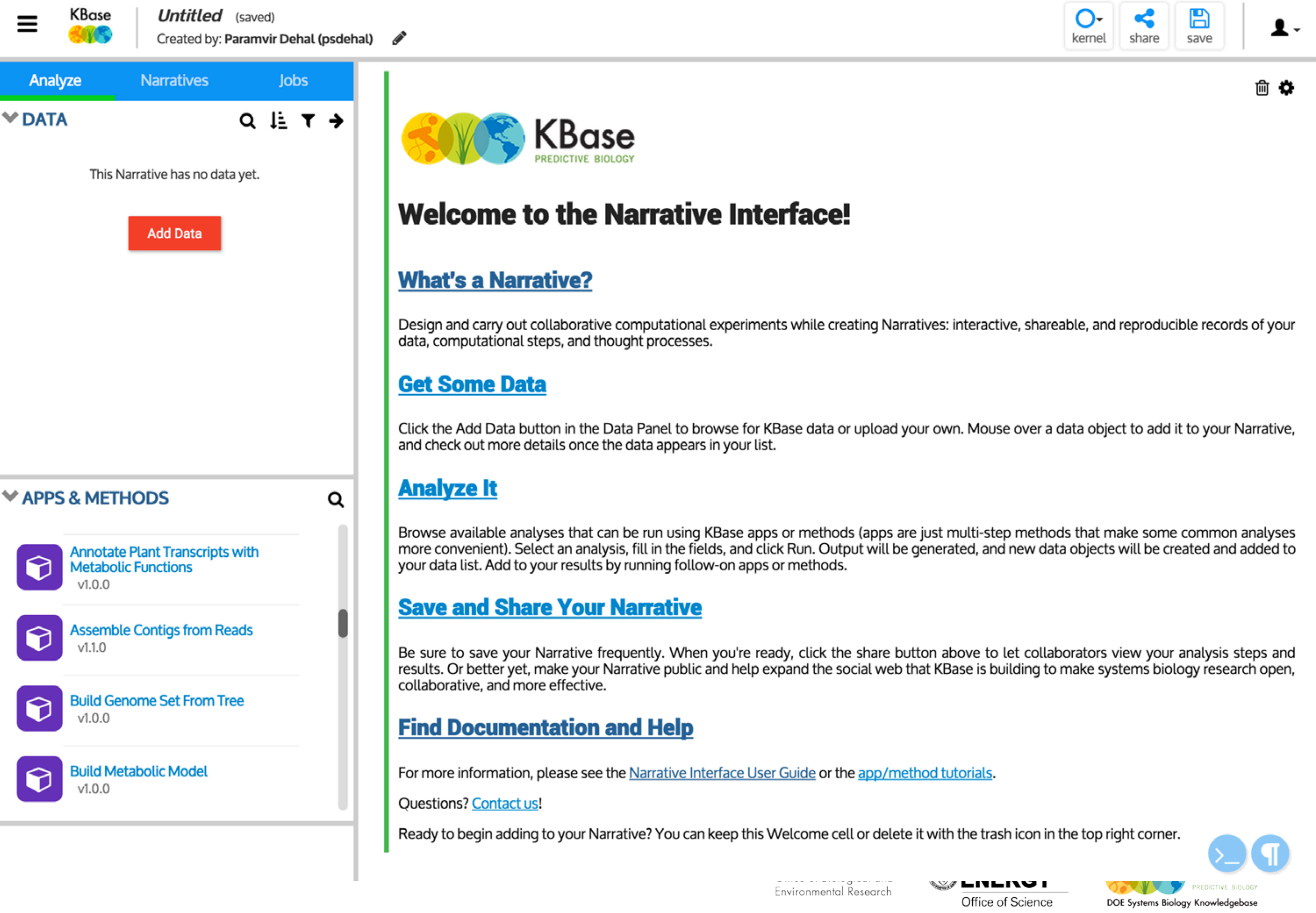
Your Collaborators

Show 10 entries

Search:

Name	Username	In common
Adam Paul Arkin	aparkin	4
Bill Riehl	wjriehl	10
Christopher Henry	chenry	4
Dan Gunter	dangunter	1
Daniel Olson	dolson	3
Dylan Chivian	dylan	1
Fangfang Xia	fangfang	3
Fei He	plane83	1
Janaka N Edirisinghe	janakakbase	1
Jason Baumohl	jkbaumohl	1

Showing 1 to 10 of 24 entries



Untitled (saved)

Created by: **Paramvir Dehal (psdehal)**

This Narrative has no data yet.

Welcome to the Narrative Interface!

What's a Narrative?

Design and carry out collaborative computational experiments while creating Narratives: interactive, shareable, and reproducible records of your data, computational steps, and thought processes.

Get Some Data

Click the Add Data button in the Data Panel to browse for KBase data or upload your own. Mouse over a data object to add it to your Narrative, and check out more details once the data appears in your list.

Analyze It

Browse available analyses that can be run using KBase apps or methods (apps are just multi-step methods that make some common analyses more convenient). Select an analysis, fill in the fields, and click Run. Output will be generated, and new data objects will be created and added to your data list. Add to your results by running follow-on apps or methods.

Save and Share Your Narrative

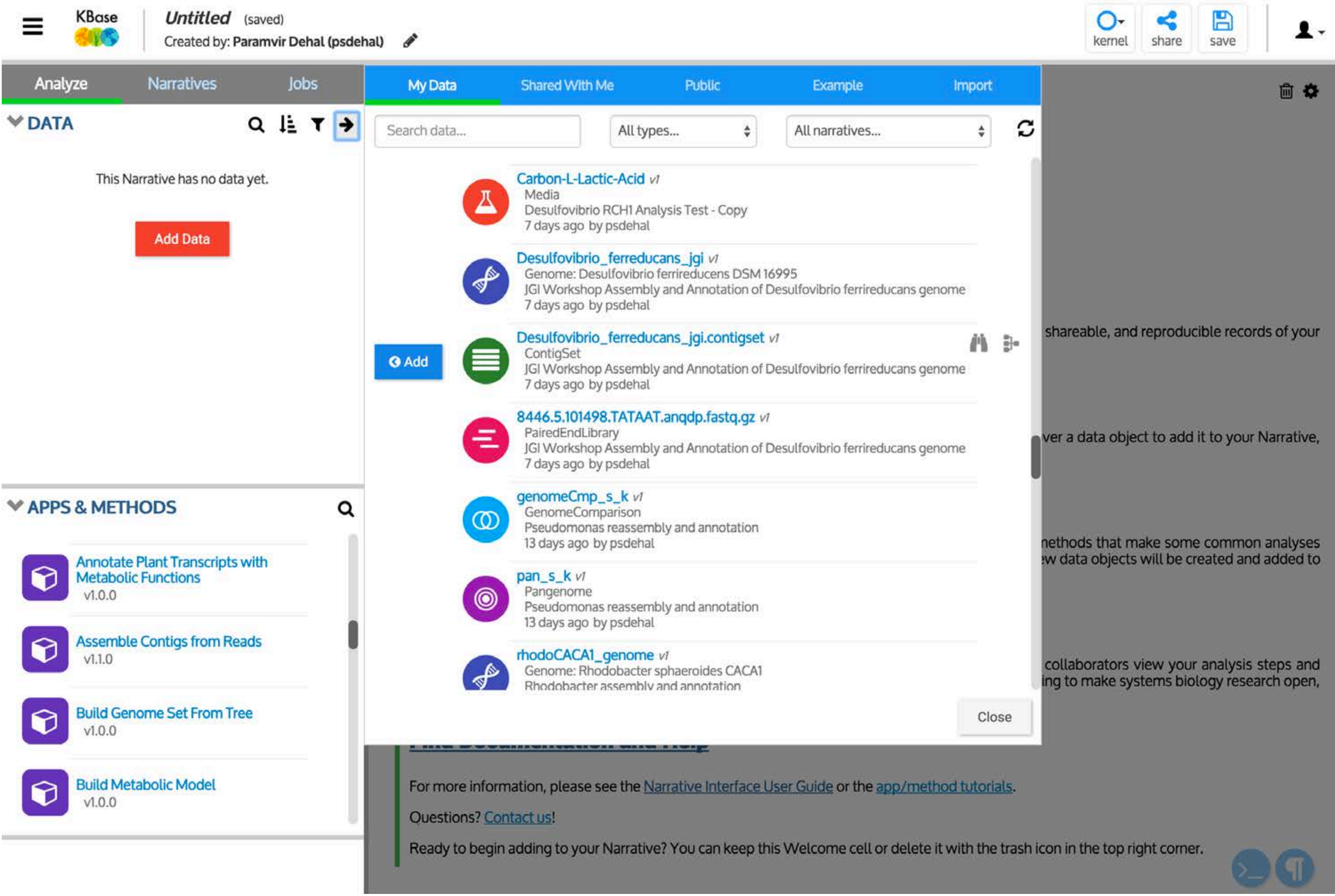
Be sure to save your Narrative frequently. When you're ready, click the share button above to let collaborators view your analysis steps and results. Or better yet, make your Narrative public and help expand the social web that KBase is building to make systems biology research open, collaborative, and more effective.

Find Documentation and Help

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



Carbon-L-Lactic-Acid v1
Media
Desulfovibrio RCH1 Analysis Test - Copy
7 days ago by psdehal

Desulfovibrio_ferreducans_jgi v1
Genome: Desulfovibrio ferrireducens DSM 16995
JGI Workshop Assembly and Annotation of Desulfovibrio ferrireducans genome
7 days ago by psdehal

Desulfovibrio_ferreducans_jgi.contigset v1
ContigSet
JGI Workshop Assembly and Annotation of Desulfovibrio ferrireducans genome
7 days ago by psdehal

8446.5.101498.TATAAT.anqdp.fastq.gz v1
PairedEndLibrary
JGI Workshop Assembly and Annotation of Desulfovibrio ferrireducans genome
7 days ago by psdehal

genomeCmp_s_k v1
GenomeComparison
Pseudomonas reassembly and annotation
13 days ago by psdehal

pan_s_k v1
Pangenome
Pseudomonas reassembly and annotation
13 days ago by psdehal

rhodoCACA1_genome v1
Genome: Rhodobacter sphaeroides CACA1
Rhodobacter assembly and annotation

shareable, and reproducible records of your

ver a data object to add it to your Narrative,

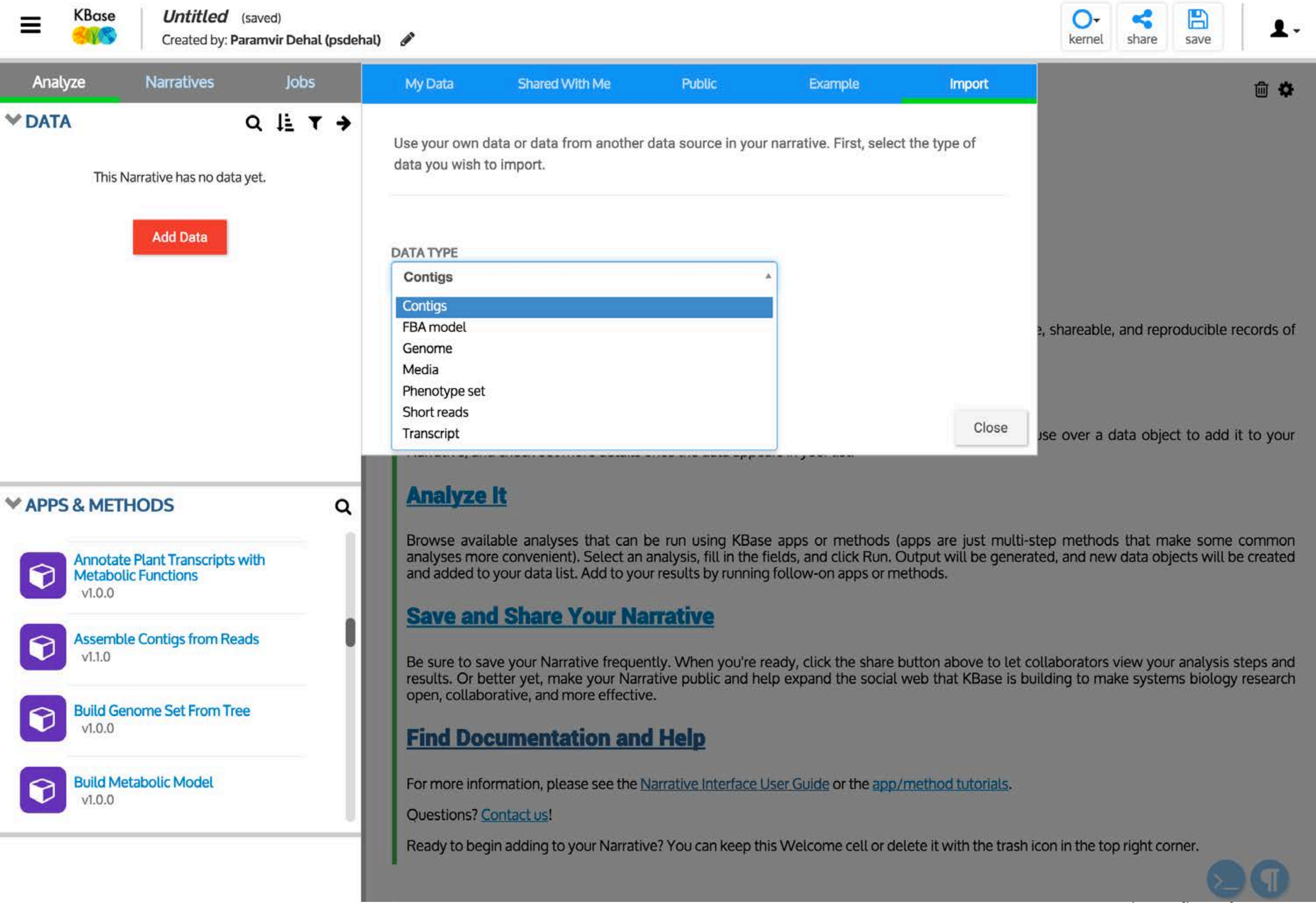
nethods that make some common analyses
w data objects will be created and added to

collaborators view your analysis steps and
ing to make systems biology research open,

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



This Narrative has no data yet.

Use your own data or data from another data source in your narrative. First, select the type of data you wish to import.

e, shareable, and reproducible records of

use over a data object to add it to your

Analyze It

Browse available analyses that can be run using KBase apps or methods (apps are just multi-step methods that make some common analyses more convenient). Select an analysis, fill in the fields, and click Run. Output will be generated, and new data objects will be created and added to your data list. Add to your results by running follow-on apps or methods.

Save and Share Your Narrative

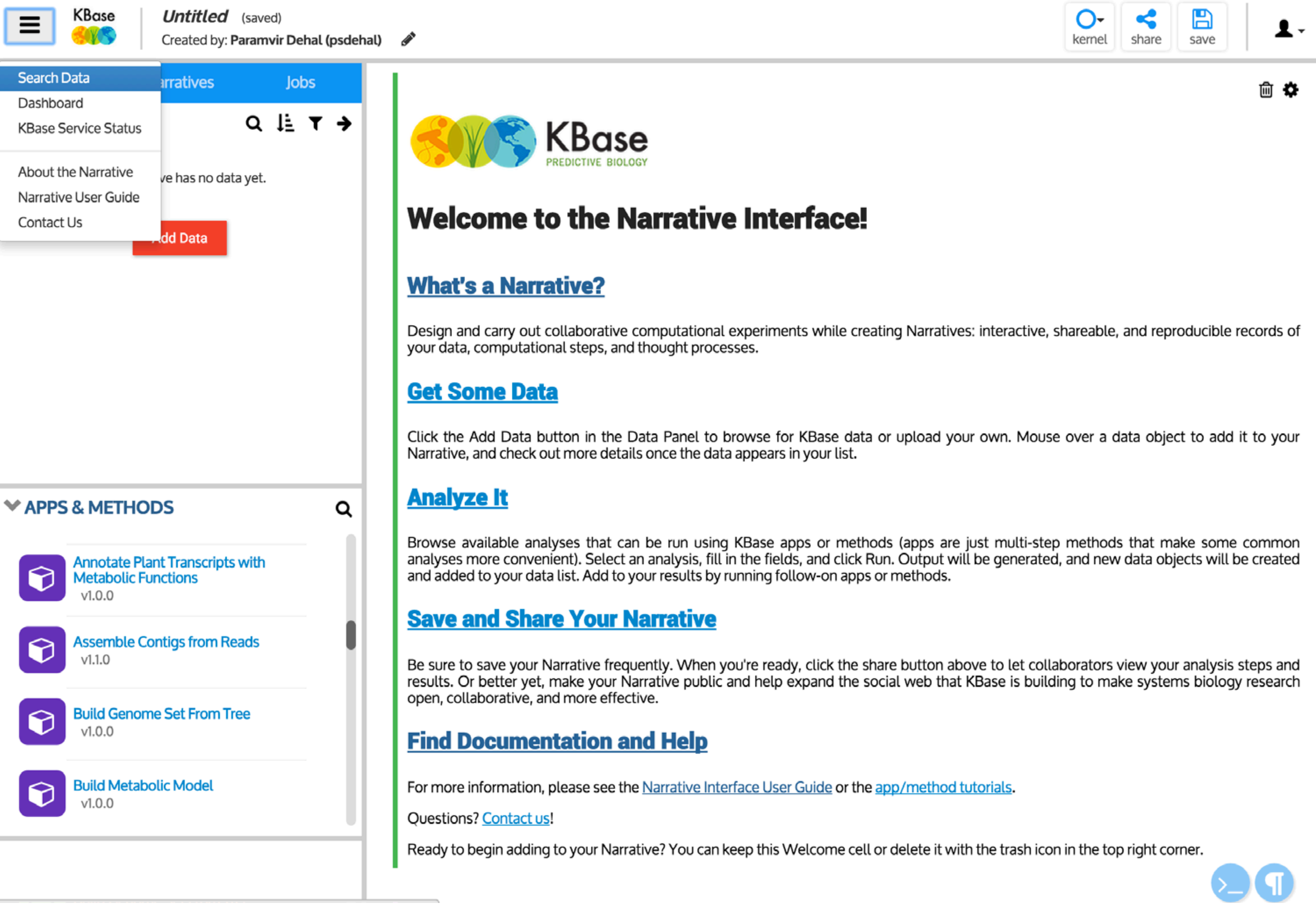
Be sure to save your Narrative frequently. When you're ready, click the share button above to let collaborators view your analysis steps and results. Or better yet, make your Narrative public and help expand the social web that KBase is building to make systems biology research open, collaborative, and more effective.

Find Documentation and Help

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



Welcome to the Narrative Interface!

What's a Narrative?

Design and carry out collaborative computational experiments while creating Narratives: interactive, shareable, and reproducible records of your data, computational steps, and thought processes.

Get Some Data

Click the Add Data button in the Data Panel to browse for KBase data or upload your own. Mouse over a data object to add it to your Narrative, and check out more details once the data appears in your list.

Analyze It

Browse available analyses that can be run using KBase apps or methods (apps are just multi-step methods that make some common analyses more convenient). Select an analysis, fill in the fields, and click Run. Output will be generated, and new data objects will be created and added to your data list. Add to your results by running follow-on apps or methods.

Save and Share Your Narrative

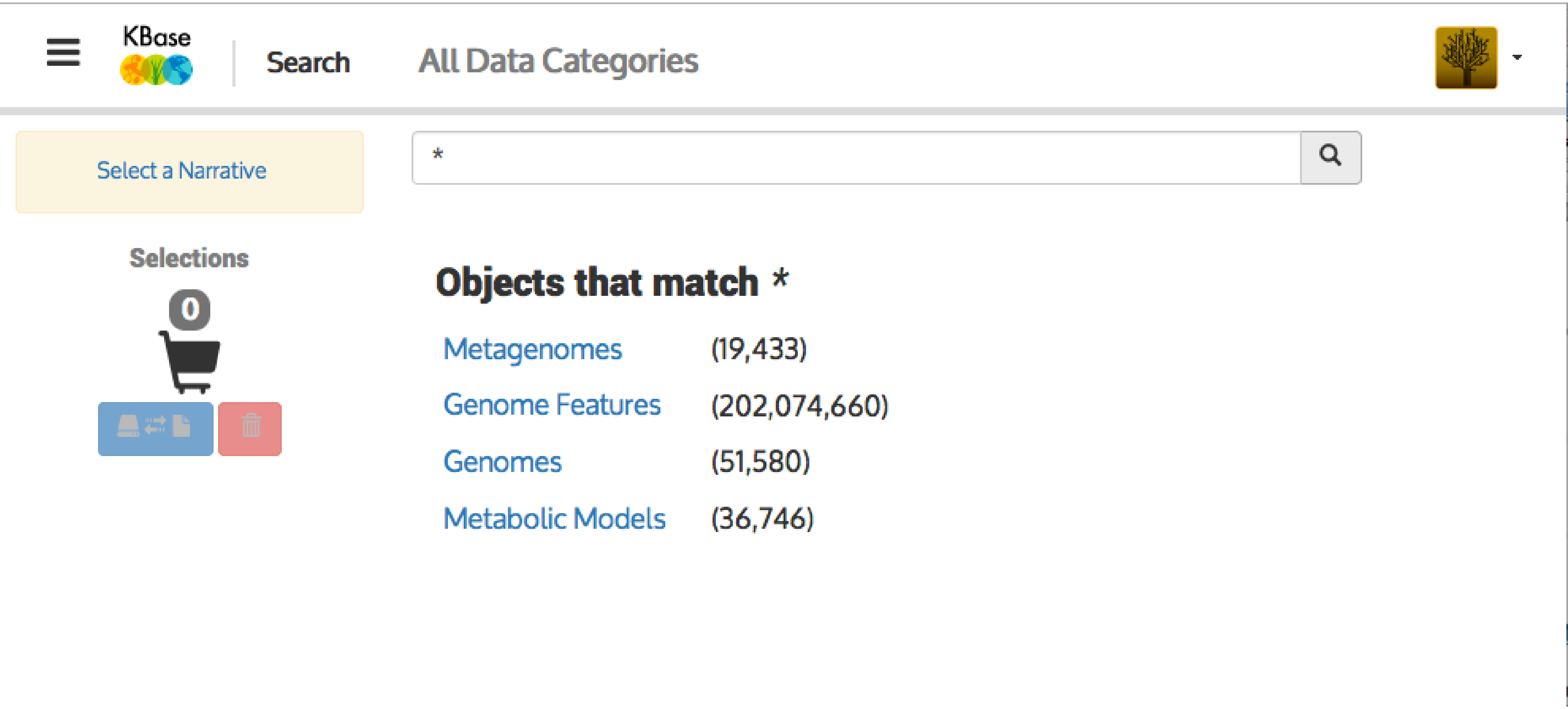
Be sure to save your Narrative frequently. When you're ready, click the share button above to let collaborators view your analysis steps and results. Or better yet, make your Narrative public and help expand the social web that KBase is building to make systems biology research open, collaborative, and more effective.

Find Documentation and Help

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

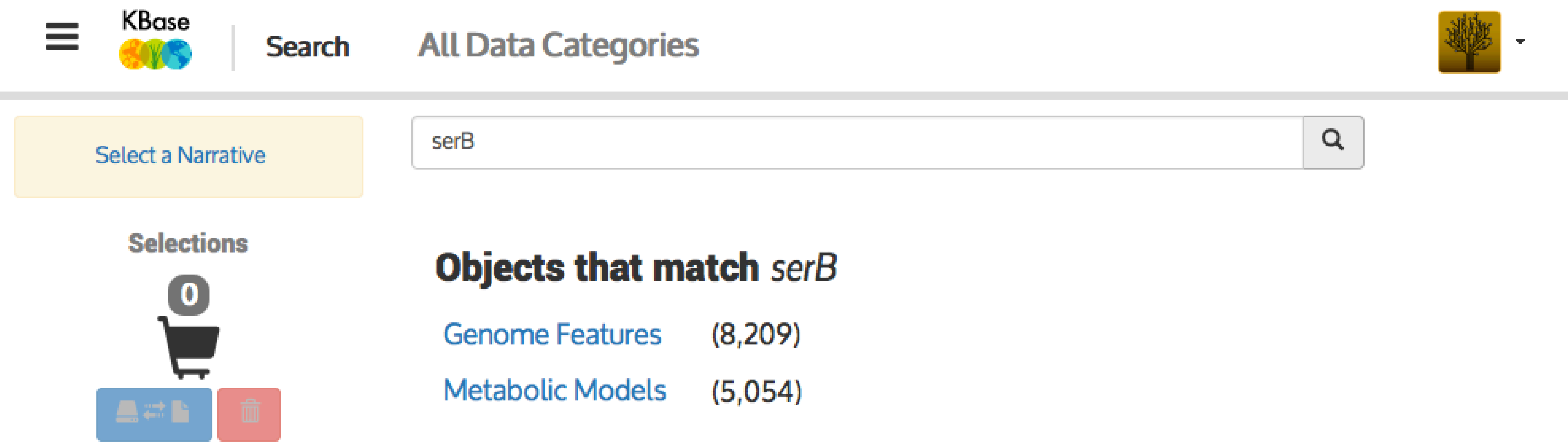
Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



Objects that match *

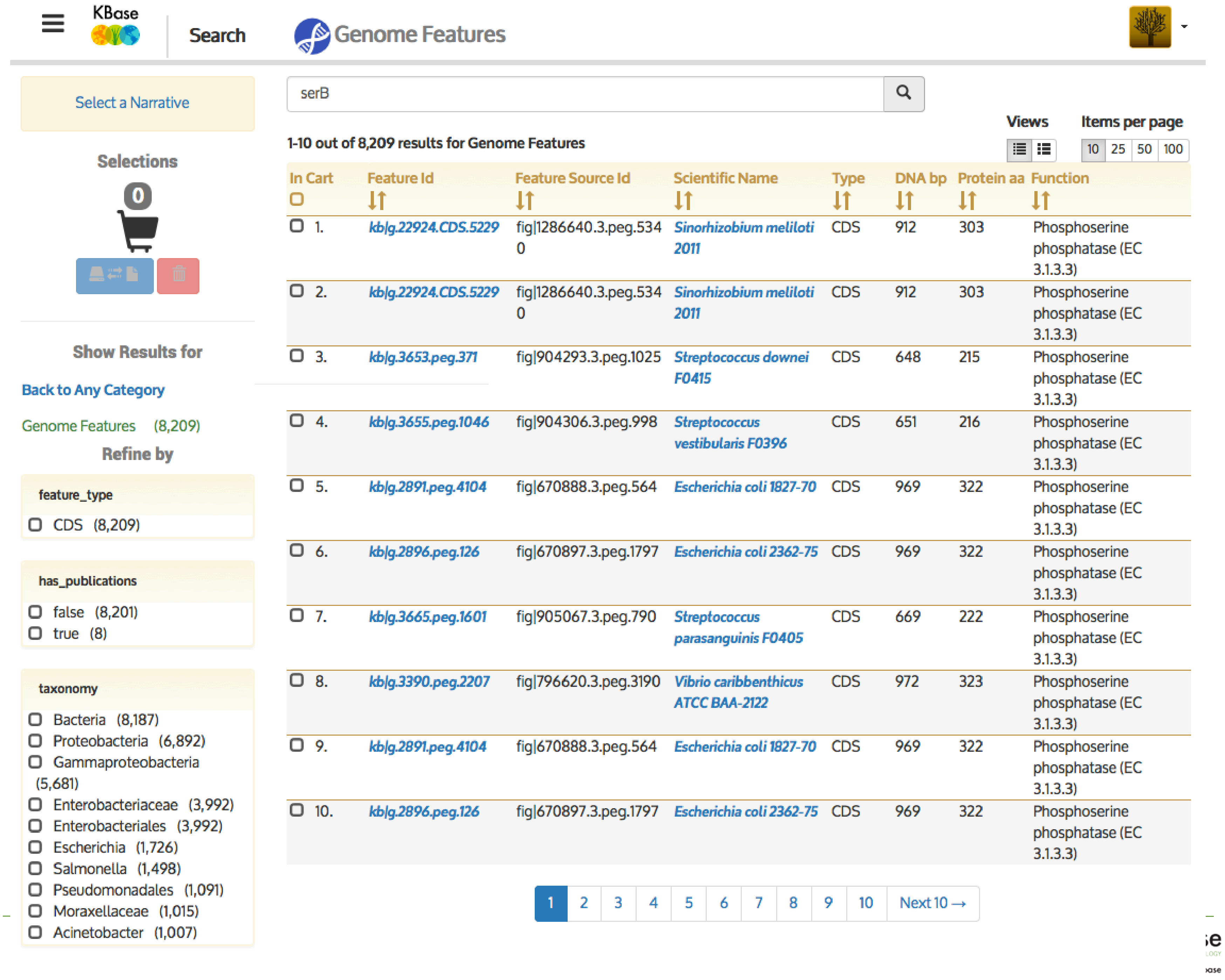
Metagenomes	(19,433)
Genome Features	(202,074,660)
Genomes	(51,580)
Metabolic Models	(36,746)



serB

Objects that match *serB*

Genome Features	(8,209)
Metabolic Models	(5,054)



1-10 out of 8,209 results for Genome Features

In Cart ☐	Feature Id	Feature Source Id	Scientific Name	Type	DNA bp	Protein aa	Function
☐ 1.	kb|g.22924.CDS.5229	fig|1286640.3.peg.5340	*Sinorhizobium meliloti 2011*	CDS	912	303	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 2.	kb|g.22924.CDS.5229	fig|1286640.3.peg.5340	*Sinorhizobium meliloti 2011*	CDS	912	303	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 3.	kb|g.3653.peg.371	fig|904293.3.peg.1025	*Streptococcus downei F0415*	CDS	648	215	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 4.	kb|g.3655.peg.1046	fig|904306.3.peg.998	*Streptococcus vestibularis F0396*	CDS	651	216	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 5.	kb|g.2891.peg.4104	fig|670888.3.peg.564	*Escherichia coli 1827-70*	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 6.	kb|g.2896.peg.126	fig|670897.3.peg.1797	*Escherichia coli 2362-75*	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 7.	kb|g.3665.peg.1601	fig|905067.3.peg.790	*Streptococcus parasanguinis F0405*	CDS	669	222	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 8.	kb|g.3390.peg.2207	fig|796620.3.peg.3190	*Vibrio caribbenthicus ATCC BAA-2122*	CDS	972	323	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 9.	kb|g.2891.peg.4104	fig|670888.3.peg.564	*Escherichia coli 1827-70*	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐ 10.	kb|g.2896.peg.126	fig|670897.3.peg.1797	*Escherichia coli 2362-75*	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)

Show Results for

Back to Any Category

Genome Features (8,209)

Refine by

feature_type

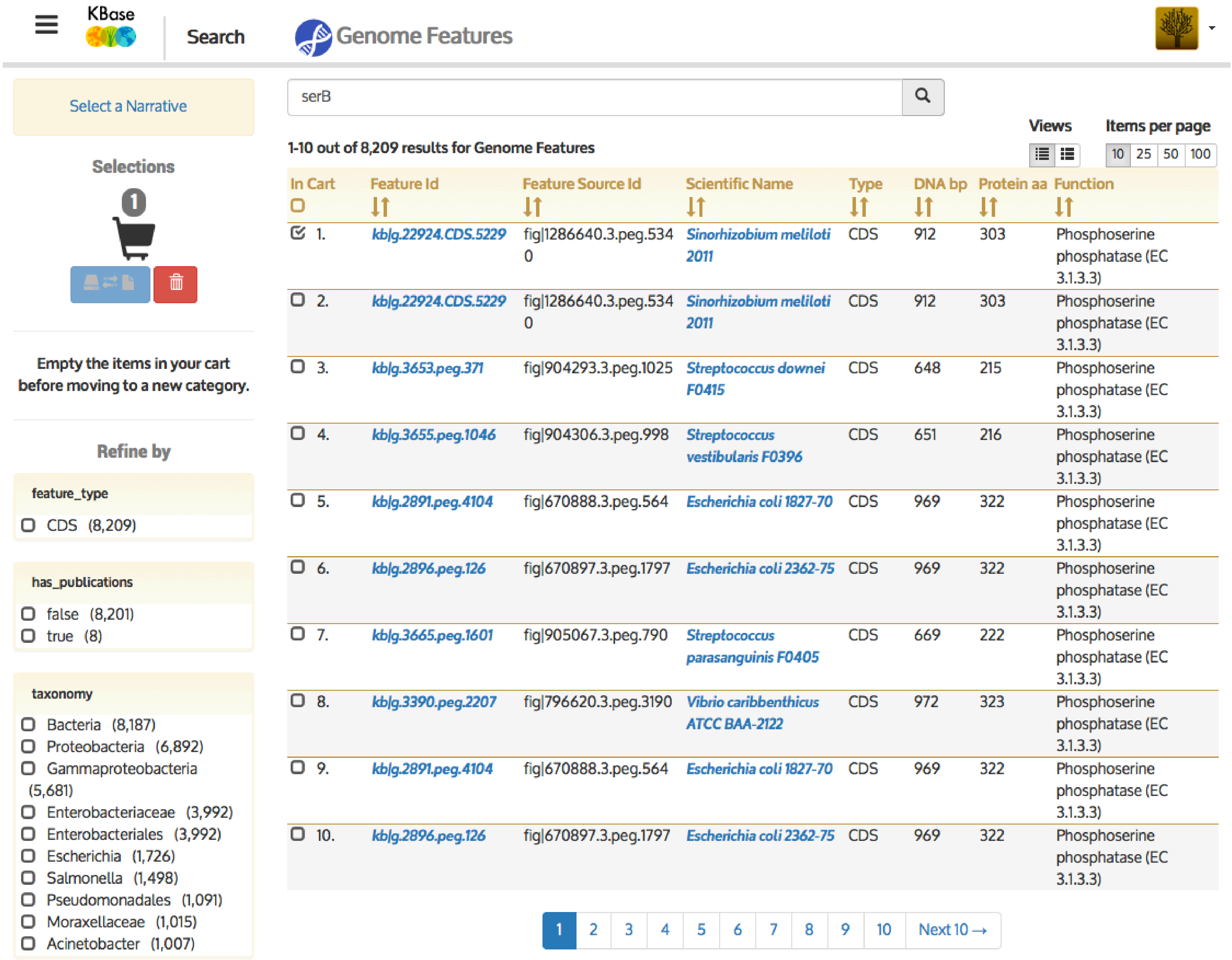
☐ CDS (8,209)

has_publications

☐ false (8,201)
☐ true (8)

taxonomy

☐ Bacteria (8,187)
☐ Proteobacteria (6,892)
☐ Gammaproteobacteria (5,681)
☐ Enterobacteriaceae (3,992)
☐ Enterobacteriales (3,992)
☐ Escherichia (1,726)
☐ Salmonella (1,498)
☐ Pseudomonadales (1,091)
☐ Moraxellaceae (1,015)
☐ Acinetobacter (1,007)



serB

1-10 out of 8,209 results for Genome Features

In Cart ☐		Feature Id	Feature Source Id	Scientific Name	Type	DNA bp	Protein aa	Function
☑	1.	kb|g.22924.CDS.5229	fig|1286640.3.peg.5340	Sinorhizobium meliloti 2011	CDS	912	303	Phosphoserine phosphatase (EC 3.1.3.3)
☐	2.	kb|g.22924.CDS.5229	fig|1286640.3.peg.5340	Sinorhizobium meliloti 2011	CDS	912	303	Phosphoserine phosphatase (EC 3.1.3.3)
☐	3.	kb|g.3653.peg.371	fig|904293.3.peg.1025	Streptococcus downei F0415	CDS	648	215	Phosphoserine phosphatase (EC 3.1.3.3)
☐	4.	kb|g.3655.peg.1046	fig|904306.3.peg.998	Streptococcus vestibularis F0396	CDS	651	216	Phosphoserine phosphatase (EC 3.1.3.3)
☐	5.	kb|g.2891.peg.4104	fig|670888.3.peg.564	Escherichia coli 1827-70	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐	6.	kb|g.2896.peg.126	fig|670897.3.peg.1797	Escherichia coli 2362-75	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐	7.	kb|g.3665.peg.1601	fig|905067.3.peg.790	Streptococcus parasanguinis F0405	CDS	669	222	Phosphoserine phosphatase (EC 3.1.3.3)
☐	8.	kb|g.3390.peg.2207	fig|796620.3.peg.3190	Vibrio caribbenthicus ATCC BAA-2122	CDS	972	323	Phosphoserine phosphatase (EC 3.1.3.3)
☐	9.	kb|g.2891.peg.4104	fig|670888.3.peg.564	Escherichia coli 1827-70	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)
☐	10.	kb|g.2896.peg.126	fig|670897.3.peg.1797	Escherichia coli 2362-75	CDS	969	322	Phosphoserine phosphatase (EC 3.1.3.3)

Empty the items in your cart before moving to a new category.

Refine by

feature_type

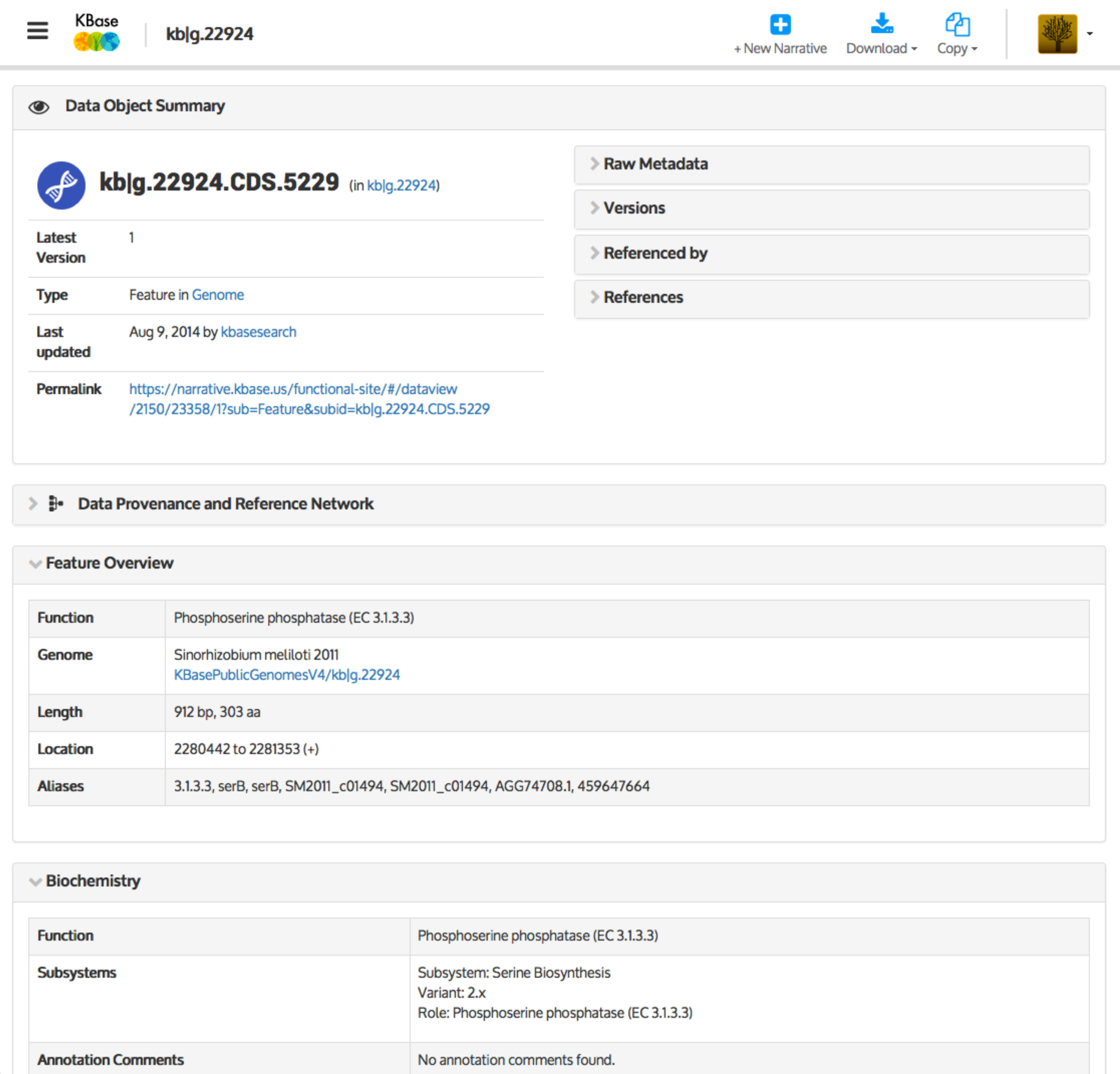
☐ CDS (8,209)

has_publications

☐ false (8,201)
☐ true (8)

taxonomy

☐ Bacteria (8,187)
☐ Proteobacteria (6,892)
☐ Gammaproteobacteria (5,681)
☐ Enterobacteriaceae (3,992)
☐ Enterobacteriales (3,992)
☐ Escherichia (1,726)
☐ Salmonella (1,498)
☐ Pseudomonadales (1,091)
☐ Moraxellaceae (1,015)
☐ Acinetobacter (1,007)



Data Object Summary

kb|g.22924.CDS.5229 (in kb|g.22924)

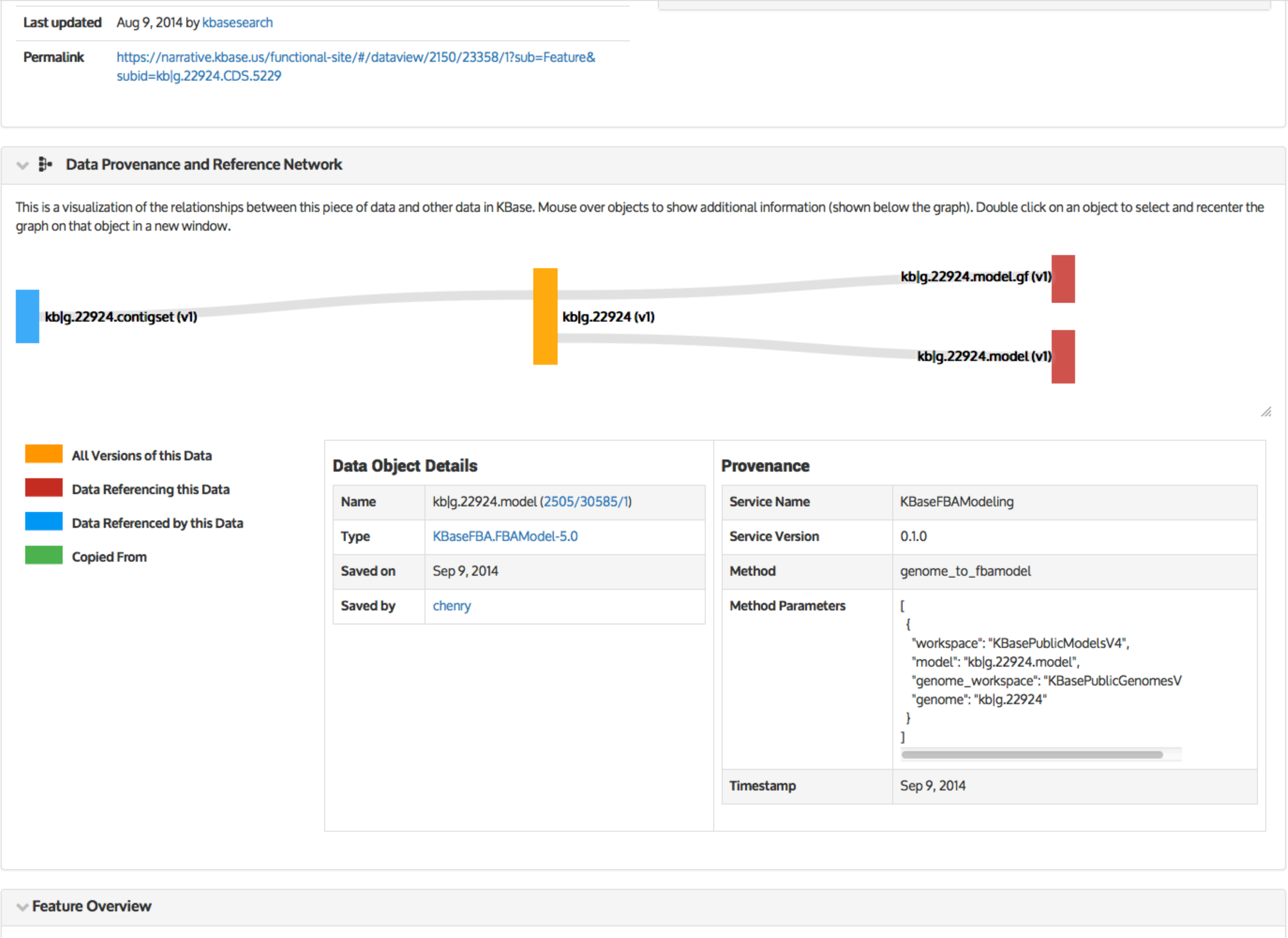
Latest Version	1
Type	Feature in Genome
Last updated	Aug 9, 2014 by kbasesearch
Permalink	https://narrative.kbase.us/functional-site/#/dataview/2150/23358/1?sub=Feature&subid=kb|g.22924.CDS.5229

Feature Overview

Function	Phosphoserine phosphatase (EC 3.1.3.3)
Genome	Sinorhizobium meliloti 2011 KBasePublicGenomesV4/kb|g.22924
Length	912 bp, 303 aa
Location	2280442 to 2281353 (+)
Aliases	3.1.3.3, serB, serB, SM2011_c01494, SM2011_c01494, AGG74708.1, 459647664

Biochemistry

Function	Phosphoserine phosphatase (EC 3.1.3.3)
Subsystems	Subsystem: Serine Biosynthesis Variant: 2.x Role: Phosphoserine phosphatase (EC 3.1.3.3)
Annotation Comments	No annotation comments found.



Last updated	Aug 9, 2014 by kbasesearch
Permalink	https://narrative.kbase.us/functional-site/#/dataview/2150/23358/1?sub=Feature&subid=kb|g.22924.CDS.5229

Data Provenance and Reference Network

This is a visualization of the relationships between this piece of data and other data in KBase. Mouse over objects to show additional information (shown below the graph). Double click on an object to select and recenter the graph on that object in a new window.

- 🟧 All Versions of this Data
- 🟥 Data Referencing this Data
- 🟦 Data Referenced by this Data
- 🟩 Copied From

Data Object Details

Name	kb|g.22924.model (2505/30585/1)
Type	KBaseFBA.FBAModel-5.0
Saved on	Sep 9, 2014
Saved by	chenry

Provenance

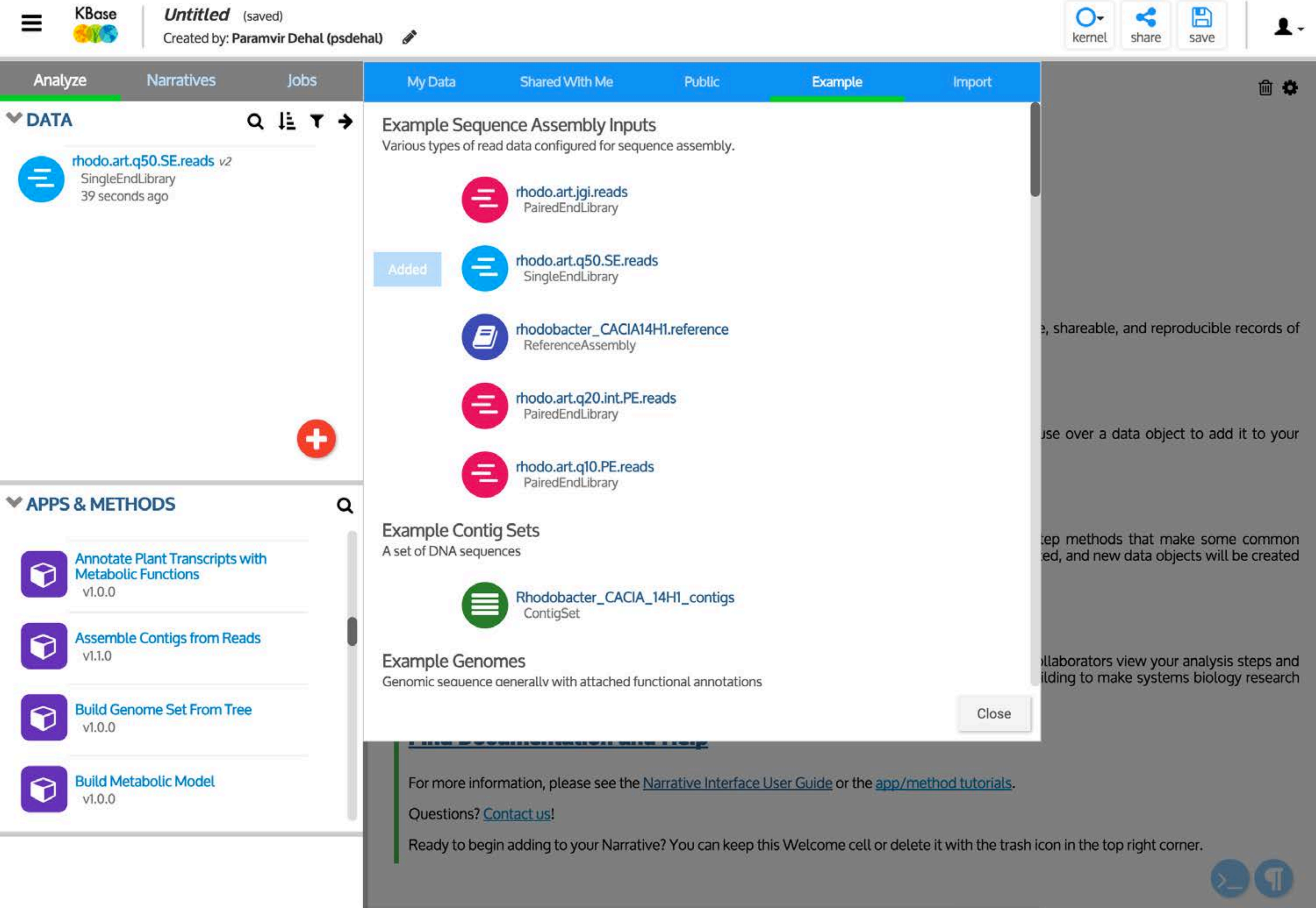
Service Name	KBaseFBAModeling
Service Version	0.1.0
Method	genome_to_fbamodel

Method Parameters

```
[
 {
  "workspace": "KBasePublicModelsV4",
  "model": "kb|g.22924.model",
  "genome_workspace": "KBasePublicGenomesV
  "genome": "kb|g.22924"
 }
]
```

Timestamp	Sep 9, 2014

Feature Overview



Example Sequence Assembly Inputs

Various types of read data configured for sequence assembly.

rhodo.art.jgi.reads
PairedEndLibrary

rhodo.art.q50.SE.reads
SingleEndLibrary

rhodobacter_CACIA14H1.reference
ReferenceAssembly

rhodo.art.q20.int.PE.reads
PairedEndLibrary

rhodo.art.q10.PE.reads
PairedEndLibrary

Example Contig Sets

A set of DNA sequences

Rhodobacter_CACIA_14H1_contigs
ContigSet

Example Genomes

Genomic sequence generally with attached functional annotations

e, shareable, and reproducible records of

use over a data object to add it to your

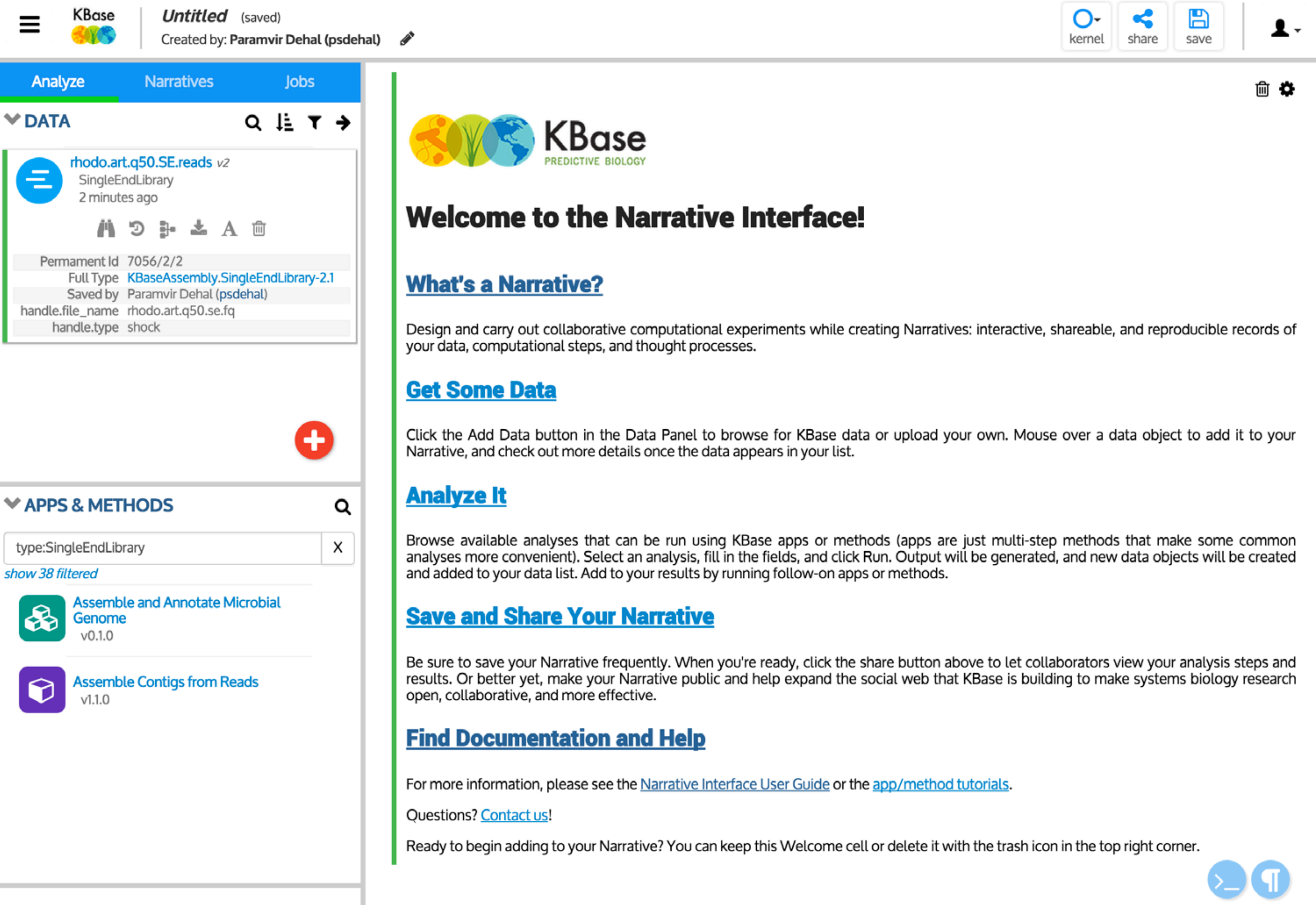
tep methods that make some common
ted, and new data objects will be created

ollaborators view your analysis steps and
ilding to make systems biology research

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
2 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary X

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

KBase PREDICTIVE BIOLOGY

Welcome to the Narrative Interface!

What's a Narrative?

Design and carry out collaborative computational experiments while creating Narratives: interactive, shareable, and reproducible records of your data, computational steps, and thought processes.

Get Some Data

Click the Add Data button in the Data Panel to browse for KBase data or upload your own. Mouse over a data object to add it to your Narrative, and check out more details once the data appears in your list.

Analyze It

Browse available analyses that can be run using KBase apps or methods (apps are just multi-step methods that make some common analyses more convenient). Select an analysis, fill in the fields, and click Run. Output will be generated, and new data objects will be created and added to your data list. Add to your results by running follow-on apps or methods.

Save and Share Your Narrative

Be sure to save your Narrative frequently. When you're ready, click the share button above to let collaborators view your analysis steps and results. Or better yet, make your Narrative public and help expand the social web that KBase is building to make systems biology research open, collaborative, and more effective.

Find Documentation and Help

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

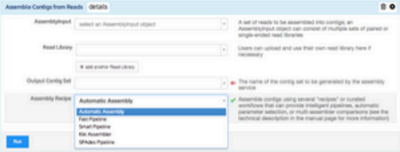
Questions? Contact us!

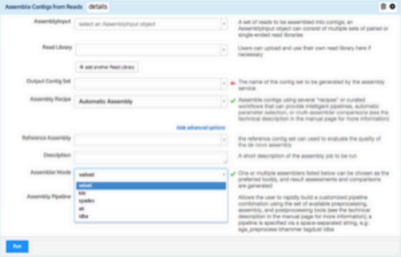
Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.

Method - Assemble Contigs from Reads

Assemble a set of short DNA reads into a set of contigs.

Help or Questions? Contact: help@kbase.us





This method can be used to perform an automatic genome assembly using the latest computational tools. Single or multiple assemblers can be invoked to compare results. Resulting assemblies are automatically processed via a collection of analysis tools developed by both KBase and the research community. The assembly method attempts to select the best (the smallest number of contigs, the longest average contig length) assembly to suggest to the user.

Several assembly workflows or "recipes" are available. These workflows have been tuned and tested to fit certain dataset types or desired analysis criteria such as throughput or rigor. The compute engine's flexible nature also enables the rapid design and emulation of other popular protocols.

Additionally, custom workflows can be designed and executed in "pipeline" mode without having to compose complicated scripts. Workflows can be composed with combinations of quality filtering or trimming, error correction, adapter removal, assembly, scaffolding, or post-processing.

Assembly Recipe Descriptions:

Automatic Assembly:

- Provides a nice balance between "fast pipeline" and "smart pipeline"
- Runs BayesHammer on reads
- Assembles with Velvet, IDBA and SPAdes
- Sorts assemblies by ALE score

Fast Pipeline:

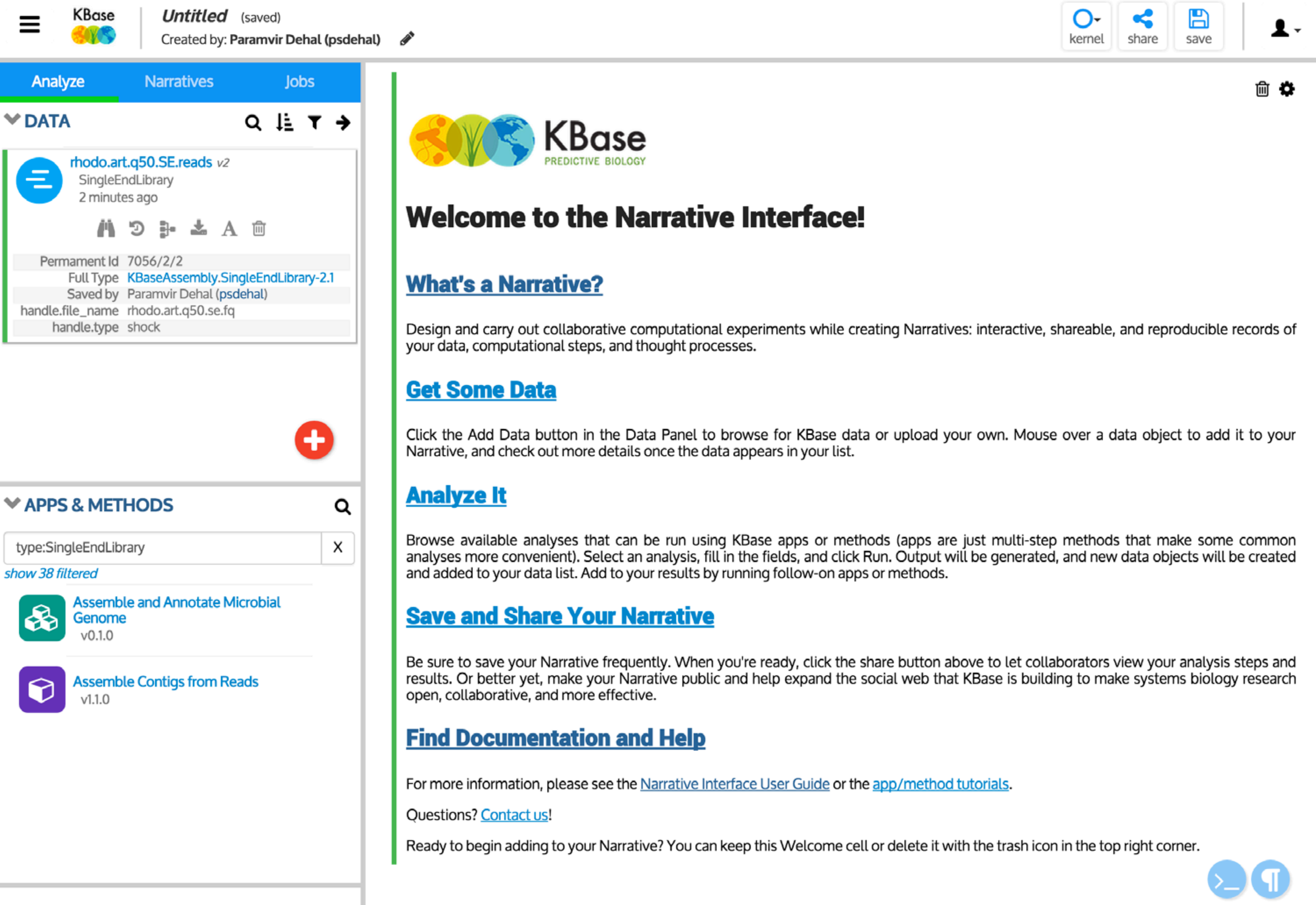
- Assembles with A6, Velvet and SPAdes (with BayesHammer for error correction)
- Results are sorted by ARAST quality score

Smart Pipeline:

- Runs BayesHammer on reads, KmerGenie to choose hash-length for Velvet
- Assembles with Velvet, IDBA and SPAdes
- Sorts assemblies by ALE score
- Merges the two best assemblies with GAM-NGS

Kiki assembler:

- Runs Kiki assembler



Permament Id 7056/2/2
Full Type KBaseAssembly.SingleEndLibrary-2.1
Saved by Paramvir Dehal (psdehal)
handle.file_name rhodo.art.q50.se.fq
handle.type shock

type:SingleEndLibrary

Welcome to the Narrative Interface!

What's a Narrative?

Design and carry out collaborative computational experiments while creating Narratives: interactive, shareable, and reproducible records of your data, computational steps, and thought processes.

Get Some Data

Click the Add Data button in the Data Panel to browse for KBase data or upload your own. Mouse over a data object to add it to your Narrative, and check out more details once the data appears in your list.

Analyze It

Browse available analyses that can be run using KBase apps or methods (apps are just multi-step methods that make some common analyses more convenient). Select an analysis, fill in the fields, and click Run. Output will be generated, and new data objects will be created and added to your data list. Add to your results by running follow-on apps or methods.

Save and Share Your Narrative

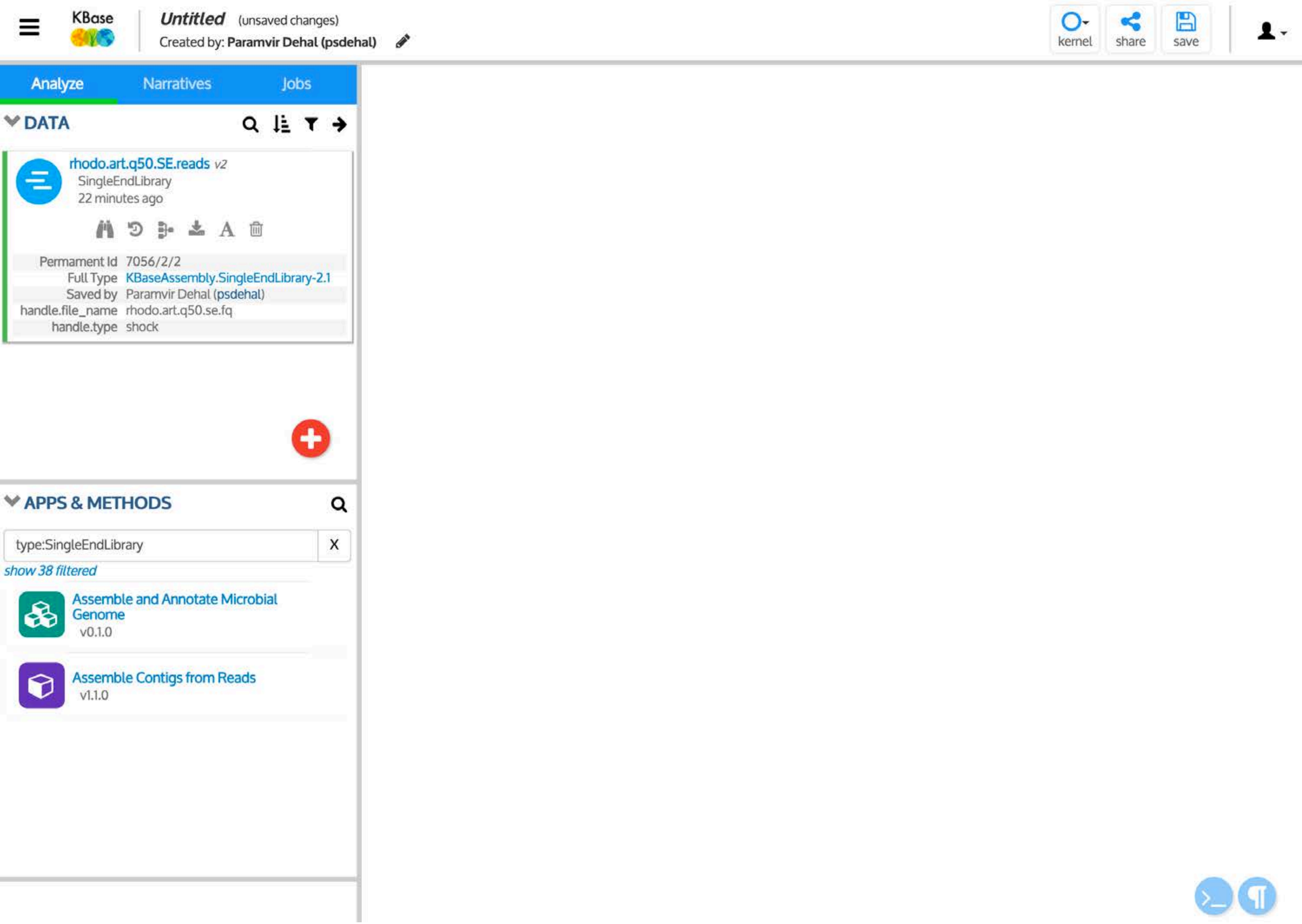
Be sure to save your Narrative frequently. When you're ready, click the share button above to let collaborators view your analysis steps and results. Or better yet, make your Narrative public and help expand the social web that KBase is building to make systems biology research open, collaborative, and more effective.

Find Documentation and Help

For more information, please see the Narrative Interface User Guide or the app/method tutorials.

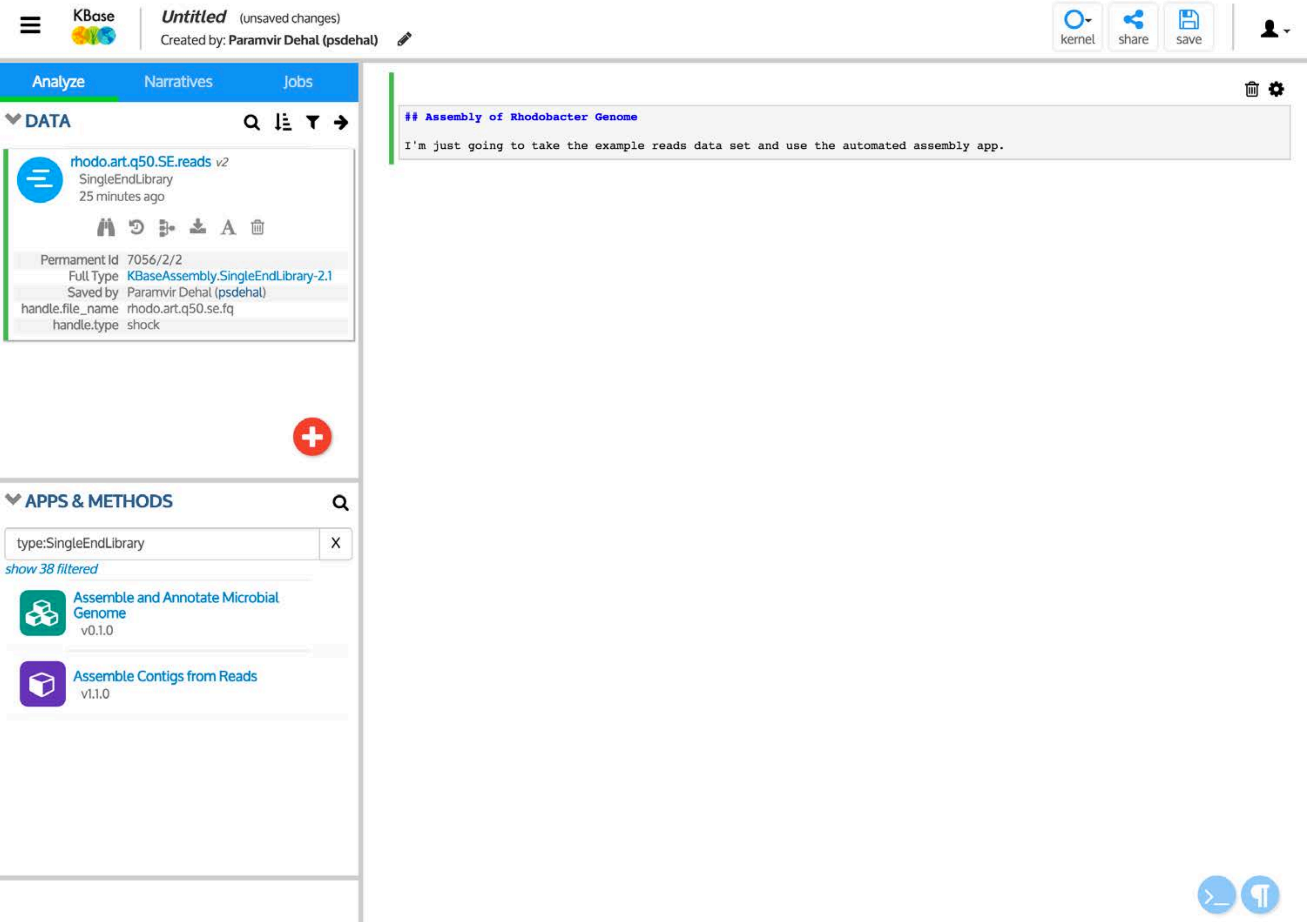
Questions? Contact us!

Ready to begin adding to your Narrative? You can keep this Welcome cell or delete it with the trash icon in the top right corner.



Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary



Untitled (unsaved changes)
Created by: **Paramvir Dehal (psdehal)**

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
25 minutes ago

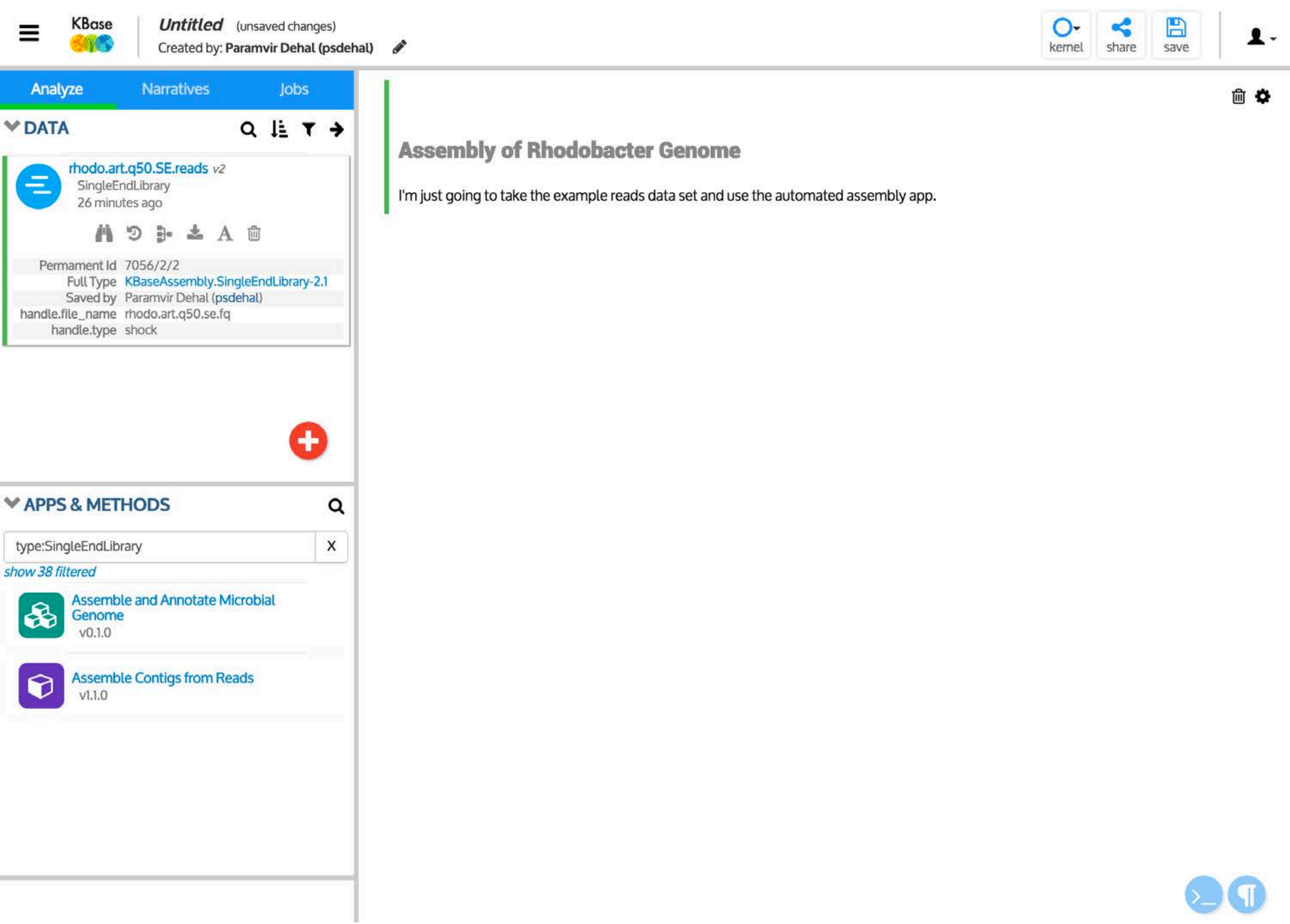
Permanent Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary X
show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

```
## Assembly of Rhodobacter Genome

I'm just going to take the example reads data set and use the automated assembly app.
```



Untitled (unsaved changes)
Created by: **Paramvir Dehal (psdehal)**

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
26 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

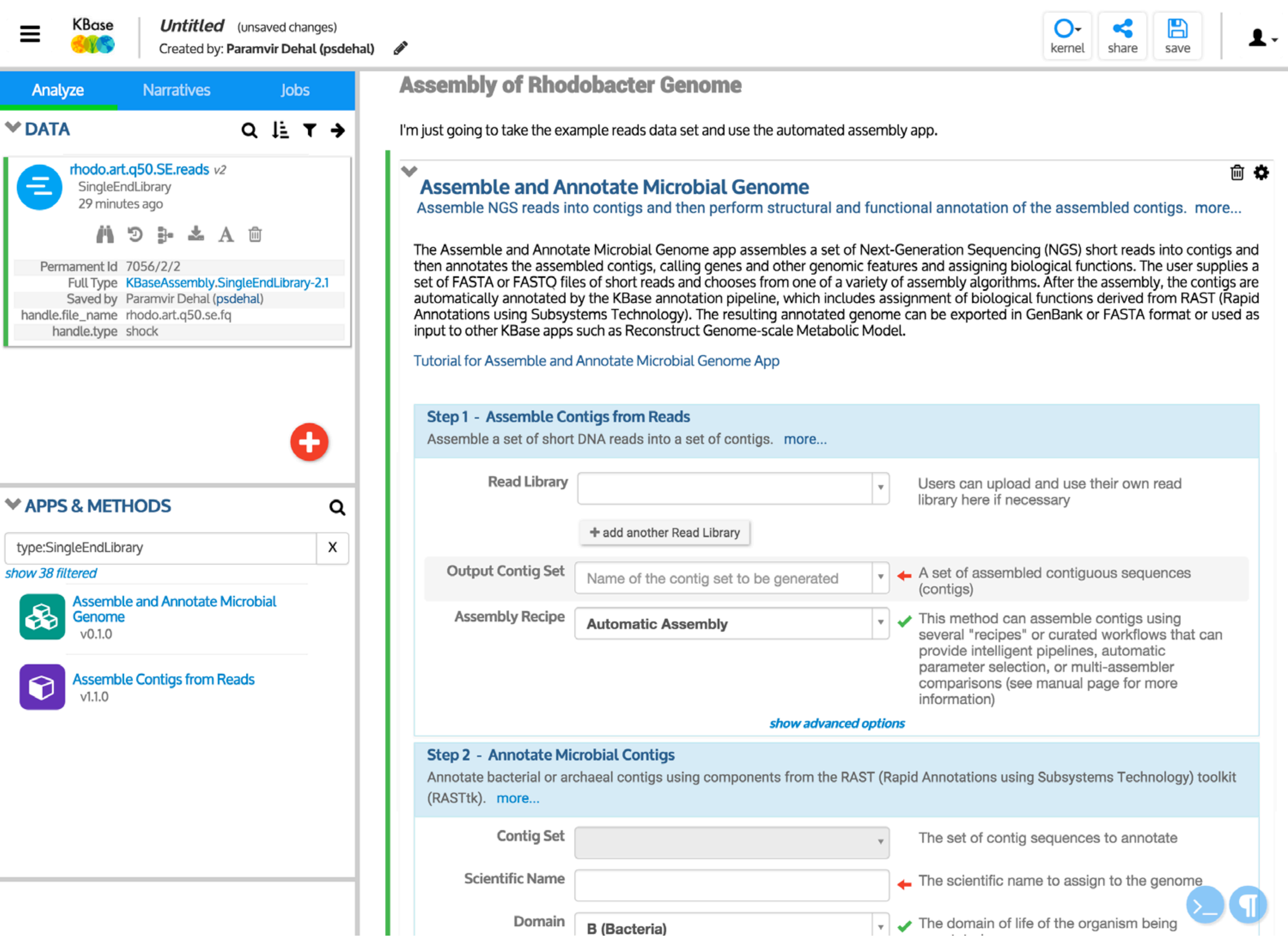
type:SingleEndLibrary

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

Assembly of Rhodobacter Genome

I'm just going to take the example reads data set and use the automated assembly app.



DATA

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
29 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

APPS & METHODS

type:SingleEndLibrary X

show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

Assembly of Rhodobacter Genome

I'm just going to take the example reads data set and use the automated assembly app.

Assemble and Annotate Microbial Genome

Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs. more...

The Assemble and Annotate Microbial Genome app assembles a set of Next-Generation Sequencing (NGS) short reads into contigs and then annotates the assembled contigs, calling genes and other genomic features and assigning biological functions. The user supplies a set of FASTA or FASTQ files of short reads and chooses from one of a variety of assembly algorithms. After the assembly, the contigs are automatically annotated by the KBase annotation pipeline, which includes assignment of biological functions derived from RAST (Rapid Annotations using Subsystems Technology). The resulting annotated genome can be exported in GenBank or FASTA format or used as input to other KBase apps such as Reconstruct Genome-scale Metabolic Model.

Tutorial for Assemble and Annotate Microbial Genome App

Step 1 - Assemble Contigs from Reads

Assemble a set of short DNA reads into a set of contigs. more...

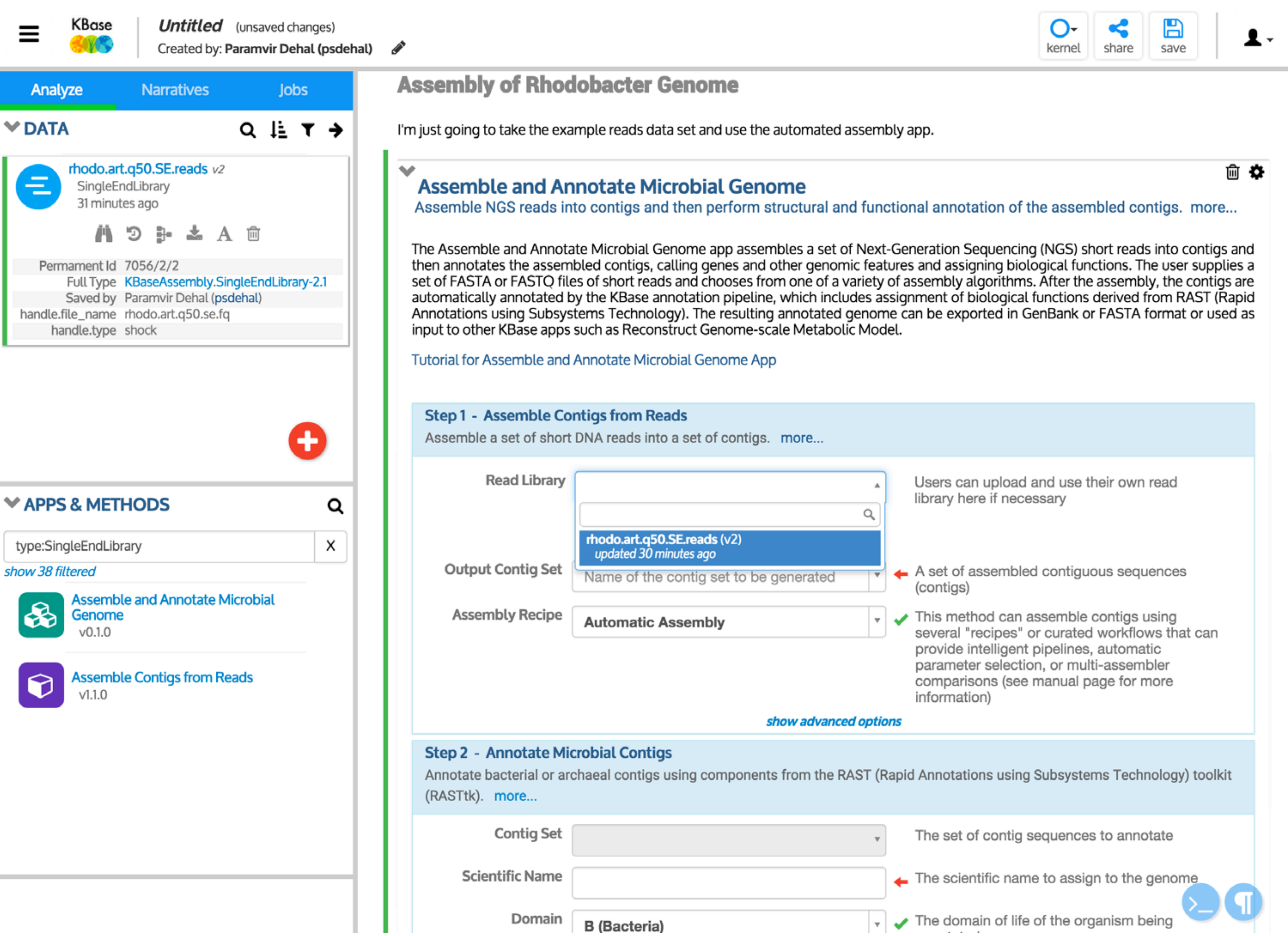
Read Library		Users can upload and use their own read library here if necessary
	+ add another Read Library	
Output Contig Set	Name of the contig set to be generated	A set of assembled contiguous sequences (contigs)
Assembly Recipe	Automatic Assembly	✓ This method can assemble contigs using several "recipes" or curated workflows that can provide intelligent pipelines, automatic parameter selection, or multi-assembler comparisons (see manual page for more information)

show advanced options

Step 2 - Annotate Microbial Contigs

Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set		The set of contig sequences to annotate
Scientific Name		The scientific name to assign to the genome
Domain	B (Bacteria)	✓ The domain of life of the organism being



Untitled (unsaved changes)
Created by: **Paramvir Dehal (psdehal)**

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
31 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary X
show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

Assembly of Rhodobacter Genome

I'm just going to take the example reads data set and use the automated assembly app.

Assemble and Annotate Microbial Genome

Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs. more...

The Assemble and Annotate Microbial Genome app assembles a set of Next-Generation Sequencing (NGS) short reads into contigs and then annotates the assembled contigs, calling genes and other genomic features and assigning biological functions. The user supplies a set of FASTA or FASTQ files of short reads and chooses from one of a variety of assembly algorithms. After the assembly, the contigs are automatically annotated by the KBase annotation pipeline, which includes assignment of biological functions derived from RAST (Rapid Annotations using Subsystems Technology). The resulting annotated genome can be exported in GenBank or FASTA format or used as input to other KBase apps such as Reconstruct Genome-scale Metabolic Model.

Tutorial for Assemble and Annotate Microbial Genome App

Step 1 - Assemble Contigs from Reads

Assemble a set of short DNA reads into a set of contigs. more...

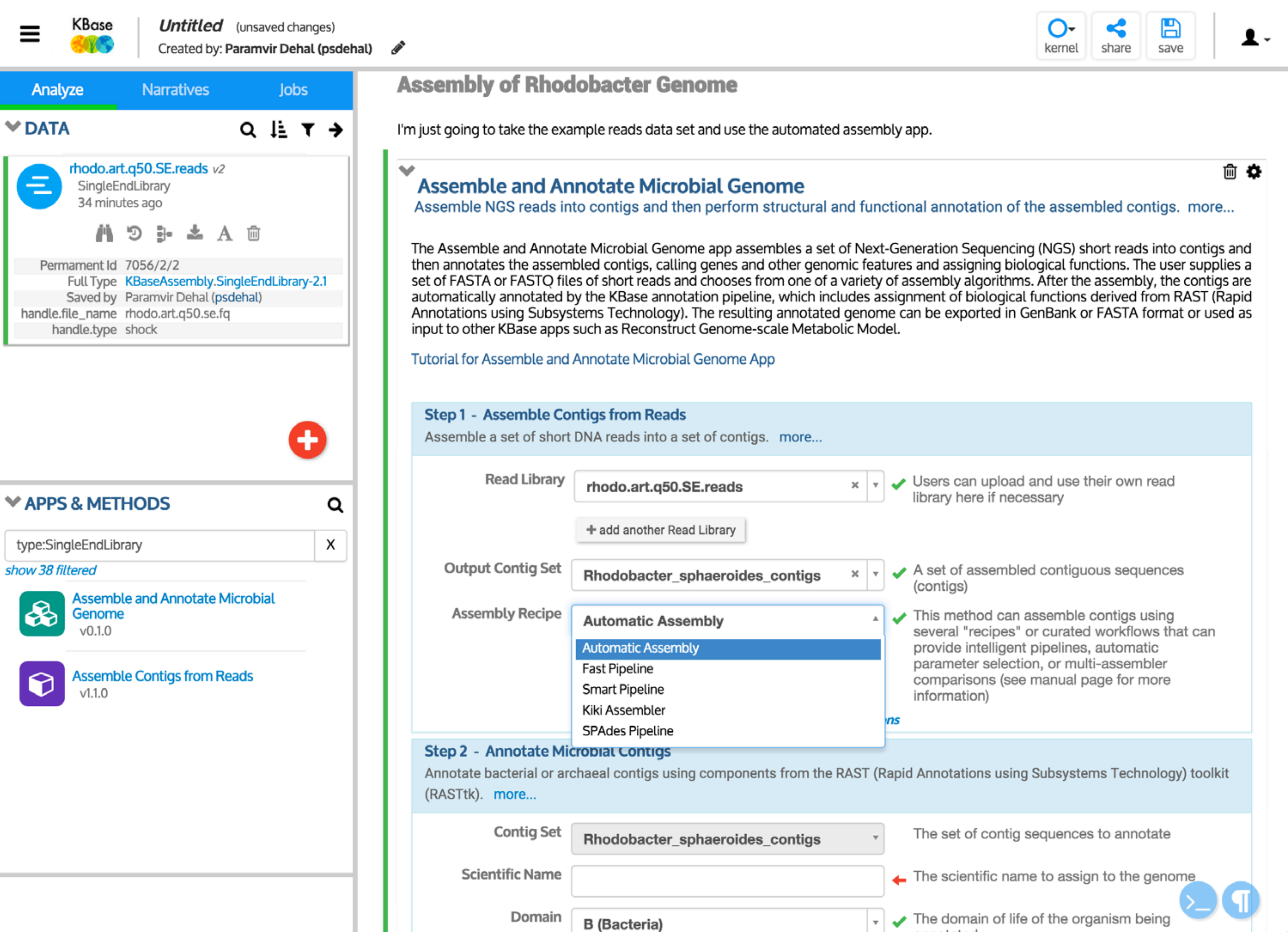
Read Library	**rhodo.art.q50.SE.reads (v2)** *updated 30 minutes ago*	Users can upload and use their own read library here if necessary
Output Contig Set	Name of the contig set to be generated	A set of assembled contiguous sequences (contigs)
Assembly Recipe	**Automatic Assembly**	This method can assemble contigs using several "recipes" or curated workflows that can provide intelligent pipelines, automatic parameter selection, or multi-assembler comparisons (see manual page for more information)

show advanced options

Step 2 - Annotate Microbial Contigs

Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set		The set of contig sequences to annotate
Scientific Name		The scientific name to assign to the genome
Domain	**B (Bacteria)**	The domain of life of the organism being



Untitled (unsaved changes)
Created by: **Paramvir Dehal (psdehal)**

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
34 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary X

show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

Assembly of Rhodobacter Genome

I'm just going to take the example reads data set and use the automated assembly app.

Assemble and Annotate Microbial Genome

Assemble NGS reads into contigs and then perform structural and functional annotation of the assembled contigs. more...

The Assemble and Annotate Microbial Genome app assembles a set of Next-Generation Sequencing (NGS) short reads into contigs and then annotates the assembled contigs, calling genes and other genomic features and assigning biological functions. The user supplies a set of FASTA or FASTQ files of short reads and chooses from one of a variety of assembly algorithms. After the assembly, the contigs are automatically annotated by the KBase annotation pipeline, which includes assignment of biological functions derived from RAST (Rapid Annotations using Subsystems Technology). The resulting annotated genome can be exported in GenBank or FASTA format or used as input to other KBase apps such as Reconstruct Genome-scale Metabolic Model.

Tutorial for Assemble and Annotate Microbial Genome App

Step 1 - Assemble Contigs from Reads

Assemble a set of short DNA reads into a set of contigs. more...

Read Library rhodo.art.q50.SE.reads ✓ Users can upload and use their own read library here if necessary

Output Contig Set Rhodobacter_sphaeroides_contigs ✓ A set of assembled contiguous sequences (contigs)

Assembly Recipe Automatic Assembly ✓ This method can assemble contigs using several "recipes" or curated workflows that can provide intelligent pipelines, automatic parameter selection, or multi-assembler comparisons (see manual page for more information)

- Automatic Assembly
- Fast Pipeline
- Smart Pipeline
- Kiki Assembler
- SPAdes Pipeline

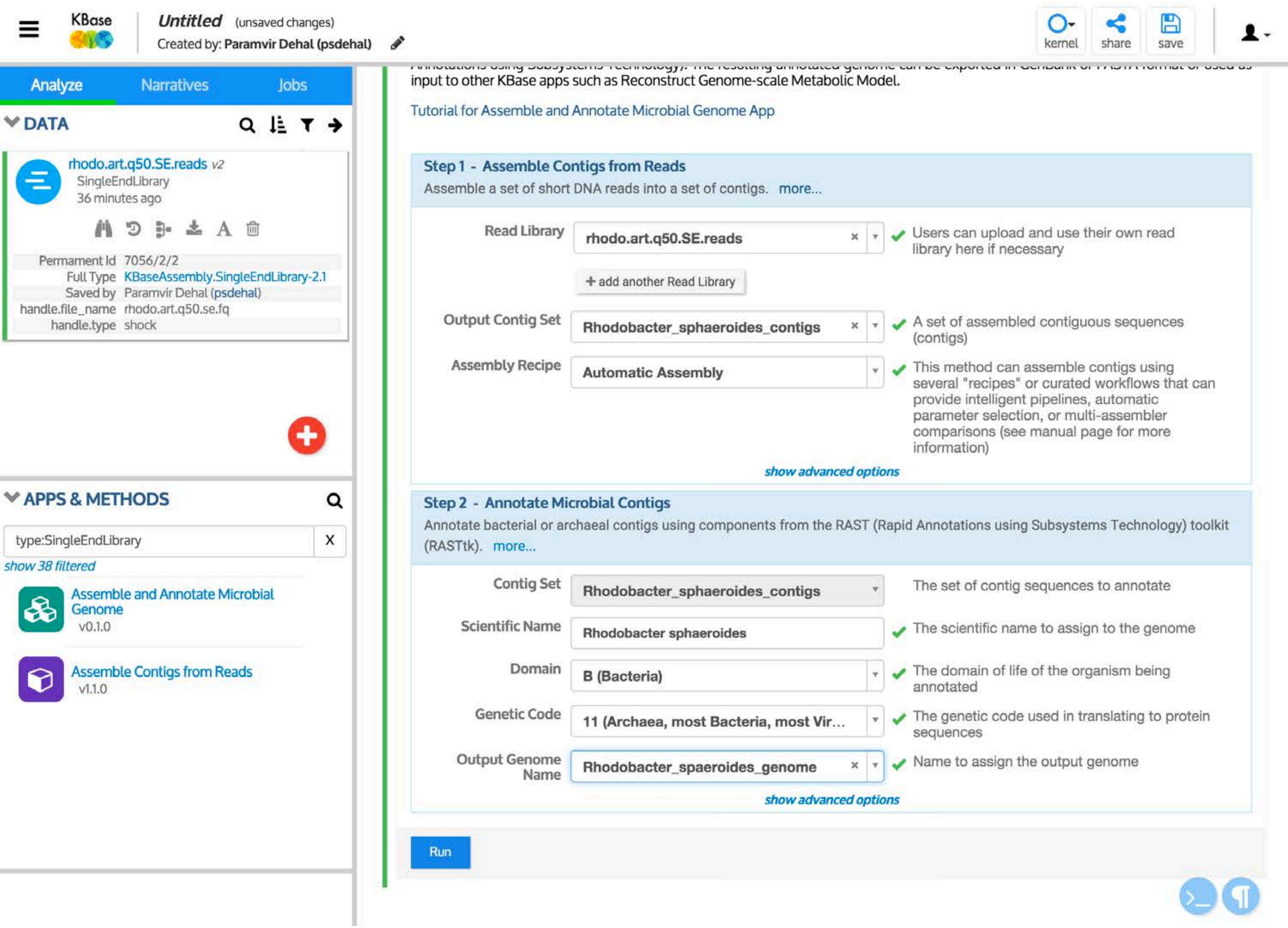
Step 2 - Annotate Microbial Contigs

Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set Rhodobacter_sphaeroides_contigs The set of contig sequences to annotate

Scientific Name The scientific name to assign to the genome

Domain B (Bacteria) ✓ The domain of life of the organism being



Untitled (unsaved changes)
Created by: **Paramvir Dehal (psdehal)**

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
36 minutes ago

Permament Id	7056/2/2
Full Type	KBaseAssembly.SingleEndLibrary-2.1
Saved by	Paramvir Dehal (psdehal)
handle.file_name	rhodo.art.q50.se.fq
handle.type	shock

type:SingleEndLibrary X

show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

input to other KBase apps such as Reconstruct Genome-scale Metabolic Model.

Tutorial for Assemble and Annotate Microbial Genome App

Step 1 - Assemble Contigs from Reads
Assemble a set of short DNA reads into a set of contigs. more...

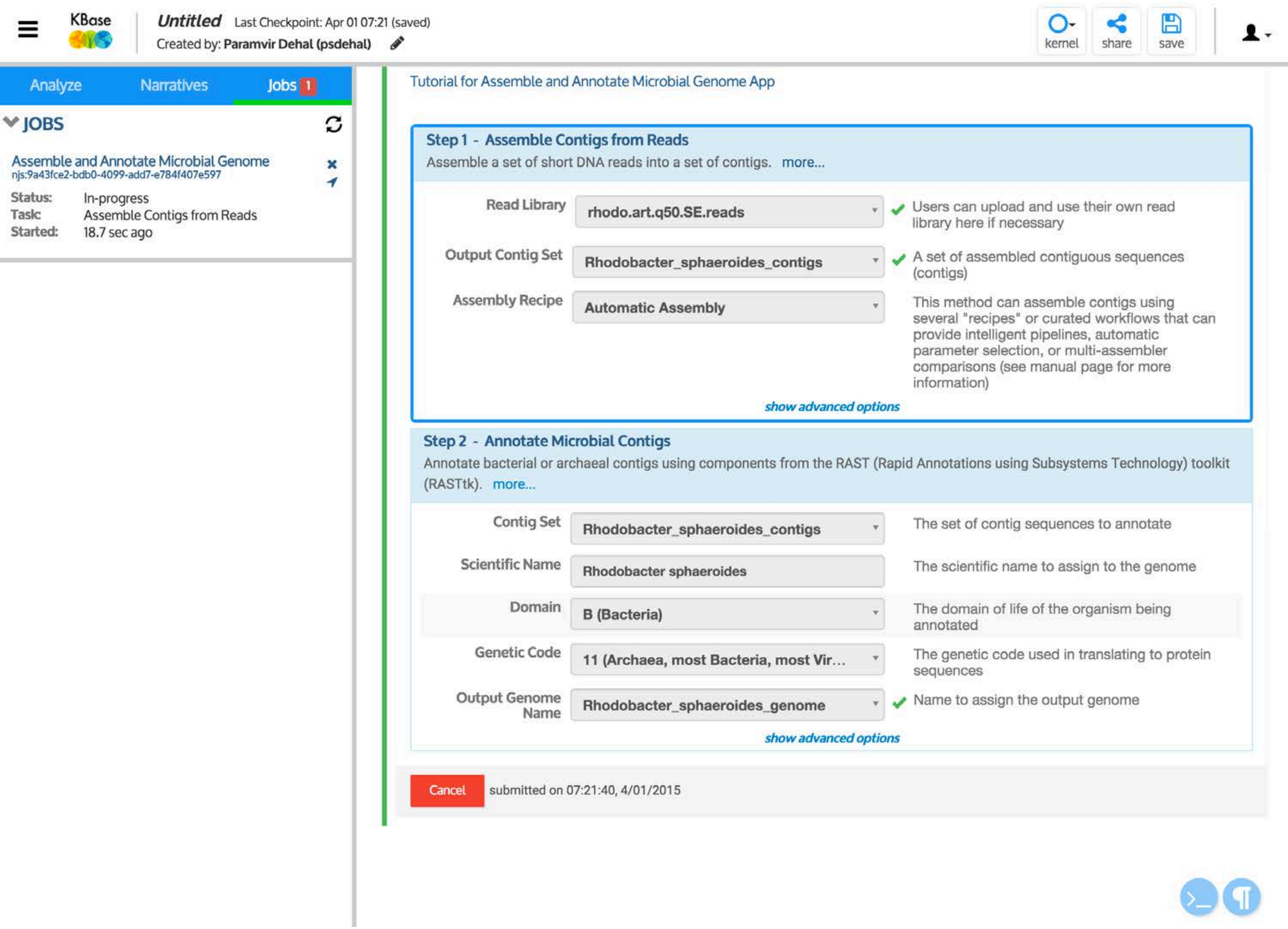
Read Library	rhodo.art.q50.SE.reads	✓ Users can upload and use their own read library here if necessary
	+ add another Read Library	
Output Contig Set	Rhodobacter_sphaeroides_contigs	✓ A set of assembled contiguous sequences (contigs)
Assembly Recipe	Automatic Assembly	✓ This method can assemble contigs using several "recipes" or curated workflows that can provide intelligent pipelines, automatic parameter selection, or multi-assembler comparisons (see manual page for more information)

show advanced options

Step 2 - Annotate Microbial Contigs
Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set	Rhodobacter_sphaeroides_contigs	The set of contig sequences to annotate
Scientific Name	Rhodobacter sphaeroides	✓ The scientific name to assign to the genome
Domain	B (Bacteria)	✓ The domain of life of the organism being annotated
Genetic Code	11 (Archaea, most Bacteria, most Vir...	✓ The genetic code used in translating to protein sequences
Output Genome Name	Rhodobacter_spaeroides_genome	✓ Name to assign the output genome

show advanced options



Untitled Last Checkpoint: Apr 01 07:21 (saved)
Created by: **Paramvir Dehal (psdehal)**

JOBS

Assemble and Annotate Microbial Genome
njs:9a43fce2-bdb0-4099-add7-e784f407e597

Status:	In-progress
Task:	Assemble Contigs from Reads
Started:	18.7 sec ago

Tutorial for Assemble and Annotate Microbial Genome App

Step 1 - Assemble Contigs from Reads

Assemble a set of short DNA reads into a set of contigs. more...

Read Library	rhodo.art.q50.SE.reads	✓ Users can upload and use their own read library here if necessary
Output Contig Set	Rhodobacter_sphaeroides_contigs	✓ A set of assembled contiguous sequences (contigs)
Assembly Recipe	Automatic Assembly	This method can assemble contigs using several "recipes" or curated workflows that can provide intelligent pipelines, automatic parameter selection, or multi-assembler comparisons (see manual page for more information)

show advanced options

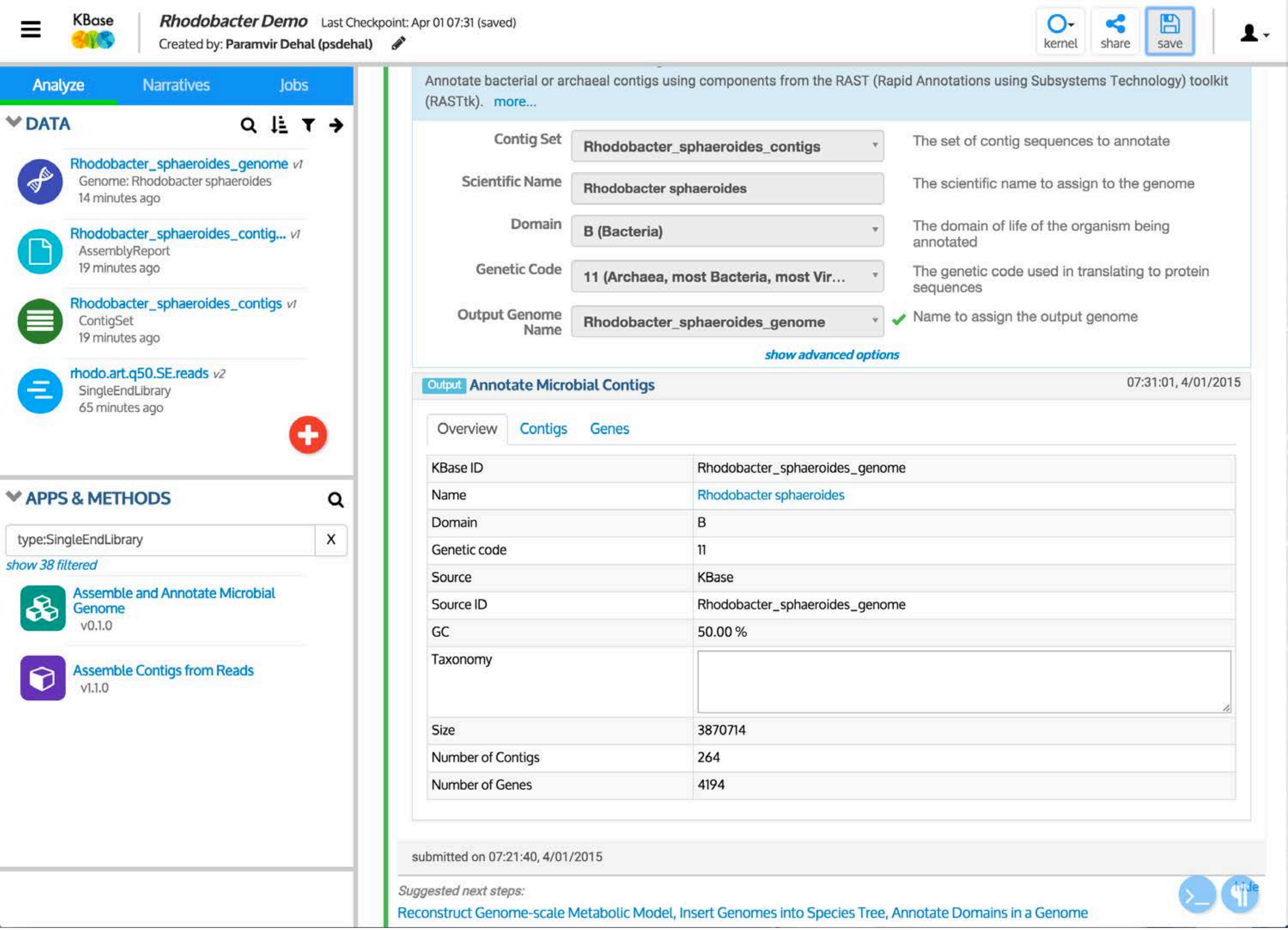
Step 2 - Annotate Microbial Contigs

Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set	Rhodobacter_sphaeroides_contigs	The set of contig sequences to annotate
Scientific Name	Rhodobacter sphaeroides	The scientific name to assign to the genome
Domain	B (Bacteria)	The domain of life of the organism being annotated
Genetic Code	11 (Archaea, most Bacteria, most Vir...	The genetic code used in translating to protein sequences
Output Genome Name	Rhodobacter_sphaeroides_genome	✓ Name to assign the output genome

show advanced options

Cancel submitted on 07:21:40, 4/01/2015



Rhodobacter Demo Last Checkpoint: Apr 01 07:31 (saved)

Created by: **Paramvir Dehal (psdehal)**

DATA

Rhodobacter_sphaeroides_genome *v1*
Genome: Rhodobacter sphaeroides
14 minutes ago

Rhodobacter_sphaeroides_contig... *v1*
AssemblyReport
19 minutes ago

Rhodobacter_sphaeroides_contigs *v1*
ContigSet
19 minutes ago

rhodo.art.q50.SE.reads *v2*
SingleEndLibrary
65 minutes ago

APPS & METHODS

type:SingleEndLibrary X

show 38 filtered

Assemble and Annotate Microbial Genome
v0.1.0

Assemble Contigs from Reads
v1.1.0

Annotate bacterial or archaeal contigs using components from the RAST (Rapid Annotations using Subsystems Technology) toolkit (RASTtk). more...

Contig Set	Rhodobacter_sphaeroides_contigs	The set of contig sequences to annotate
Scientific Name	Rhodobacter sphaeroides	The scientific name to assign to the genome
Domain	B (Bacteria)	The domain of life of the organism being annotated
Genetic Code	11 (Archaea, most Bacteria, most Vir...	The genetic code used in translating to protein sequences
Output Genome Name	Rhodobacter_sphaeroides_genome ✓	Name to assign the output genome

show advanced options

Output **Annotate Microbial Contigs** 07:31:01, 4/01/2015

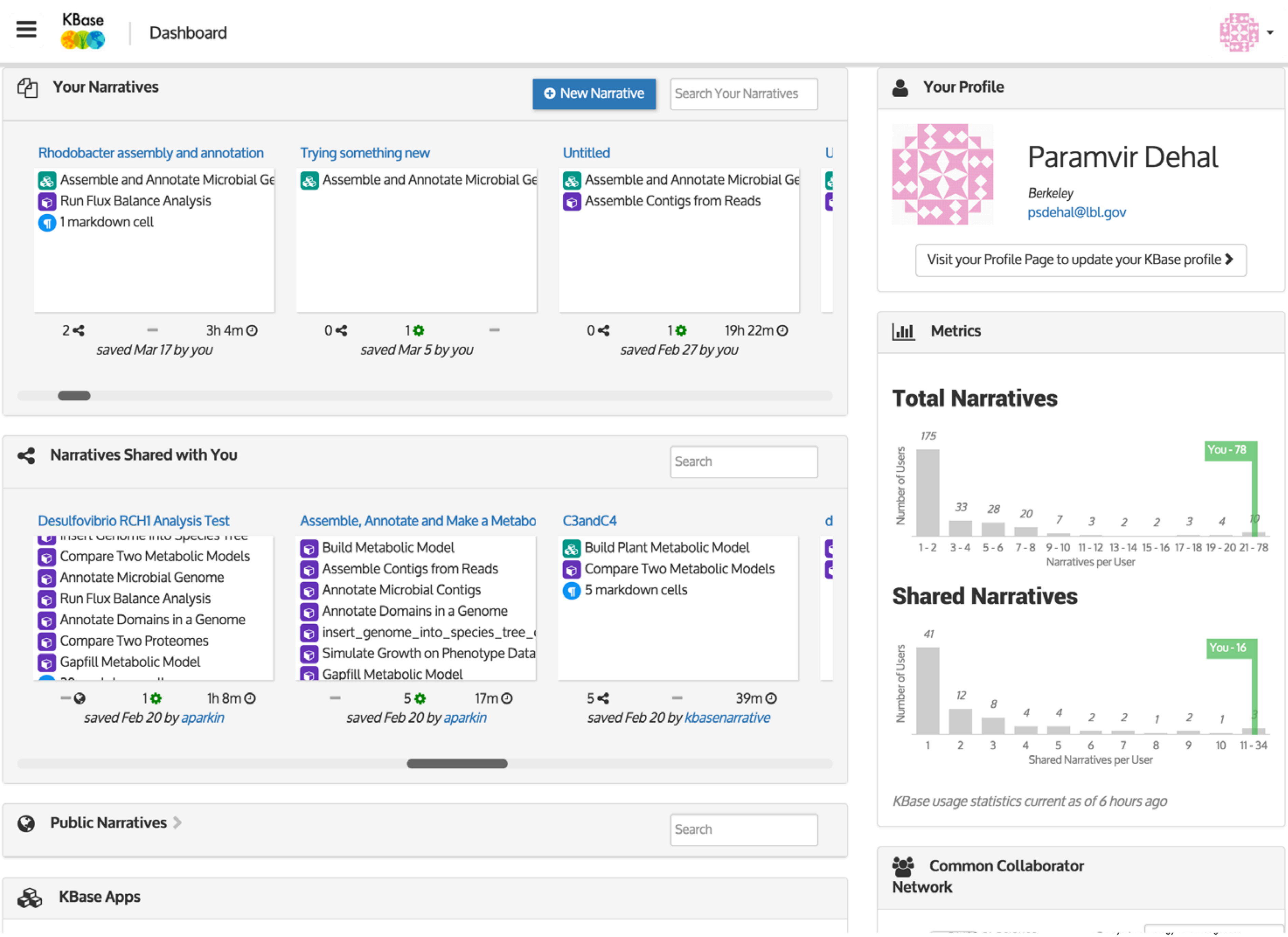
Overview Contigs Genes

KBase ID	Rhodobacter_sphaeroides_genome
Name	Rhodobacter sphaeroides
Domain	B
Genetic code	11
Source	KBase
Source ID	Rhodobacter_sphaeroides_genome
GC	50.00 %
Taxonomy	
Size	3870714
Number of Contigs	264
Number of Genes	4194

submitted on 07:21:40, 4/01/2015

Suggested next steps:

Reconstruct Genome-scale Metabolic Model, Insert Genomes into Species Tree, Annotate Domains in a Genome



Your Narratives

Rhodobacter assembly and annotation
- Assemble and Annotate Microbial Ge
- Run Flux Balance Analysis
- 1 markdown cell

2 — 3h 4m
saved Mar 17 by you

Trying something new
- Assemble and Annotate Microbial Ge

0 1 —
saved Mar 5 by you

Untitled
- Assemble and Annotate Microbial Ge
- Assemble Contigs from Reads

0 1 19h 22m
saved Feb 27 by you

Narratives Shared with You

Desulfovibrio RCH1 Analysis Test
- Insert Genome into Species Tree
- Compare Two Metabolic Models
- Annotate Microbial Genome
- Run Flux Balance Analysis
- Annotate Domains in a Genome
- Compare Two Proteomes
- Gapfill Metabolic Model

— 1 1h 8m
saved Feb 20 by aparkin

Assemble, Annotate and Make a Metabo
- Build Metabolic Model
- Assemble Contigs from Reads
- Annotate Microbial Contigs
- Annotate Domains in a Genome
- insert_genome_into_species_tree_
- Simulate Growth on Phenotype Data
- Gapfill Metabolic Model

— 5 17m
saved Feb 20 by aparkin

C3andC4
- Build Plant Metabolic Model
- Compare Two Metabolic Models
- 5 markdown cells

5 — 39m
saved Feb 20 by kbasenarrative

Public Narratives

KBase Apps

Your Profile

Paramvir Dehal

Berkeley

psdehal@lbl.gov

Visit your Profile Page to update your KBase profile

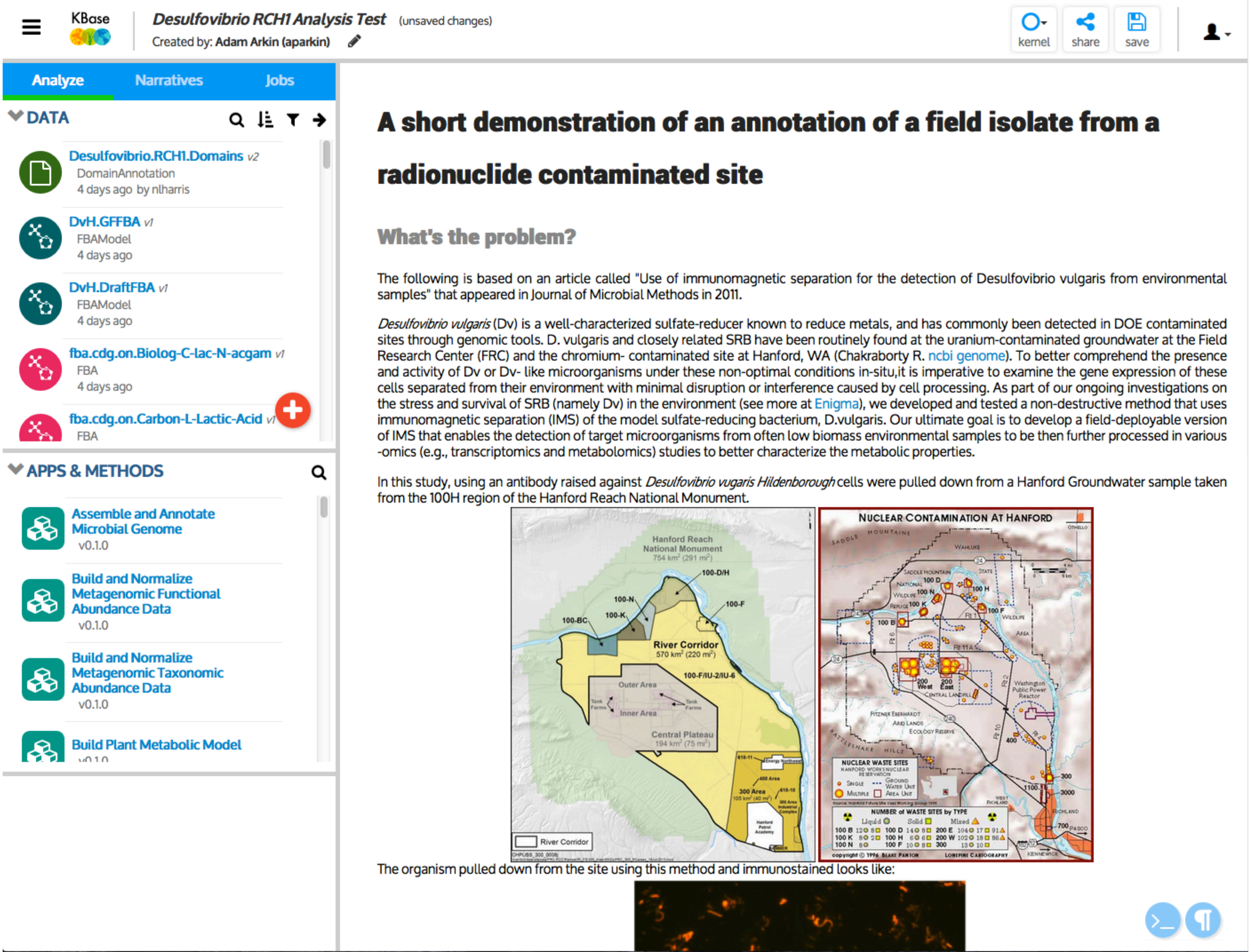
Metrics

Total Narratives

Shared Narratives

KBase usage statistics current as of 6 hours ago

Common Collaborator Network



A short demonstration of an annotation of a field isolate from a radionuclide contaminated site

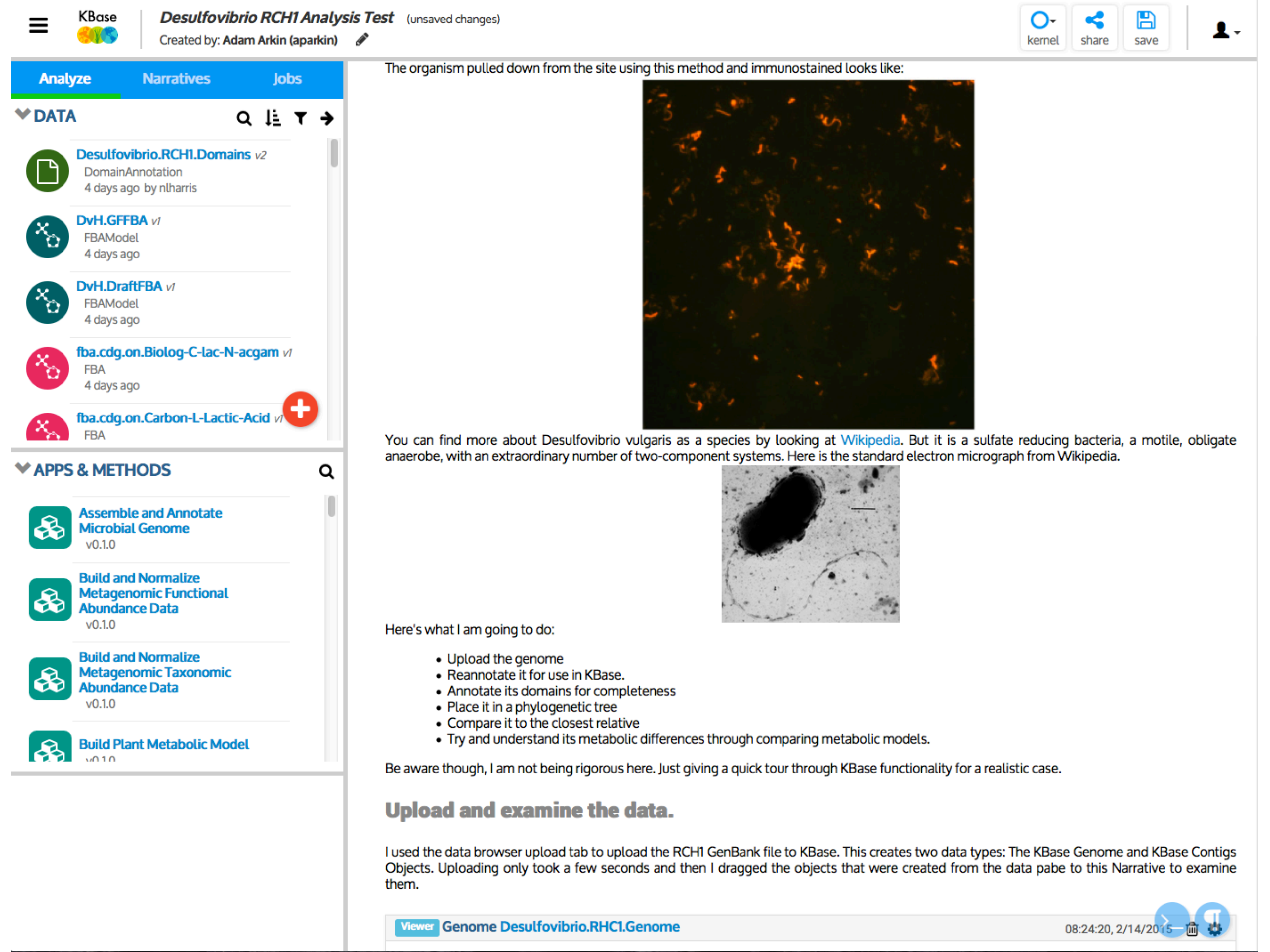
What's the problem?

The following is based on an article called "Use of immunomagnetic separation for the detection of Desulfovibrio vulgaris from environmental samples" that appeared in Journal of Microbial Methods in 2011.

Desulfovibrio vulgaris (Dv) is a well-characterized sulfate-reducer known to reduce metals, and has commonly been detected in DOE contaminated sites through genomic tools. D. vulgaris and closely related SRB have been routinely found at the uranium-contaminated groundwater at the Field Research Center (FRC) and the chromium- contaminated site at Hanford, WA (Chakraborty R. ncbi genome). To better comprehend the presence and activity of Dv or Dv- like microorganisms under these non-optimal conditions in-situ,it is imperative to examine the gene expression of these cells separated from their environment with minimal disruption or interference caused by cell processing. As part of our ongoing investigations on the stress and survival of SRB (namely Dv) in the environment (see more at Enigma), we developed and tested a non-destructive method that uses immunomagnetic separation (IMS) of the model sulfate-reducing bacterium, D.vulgaris. Our ultimate goal is to develop a field-deployable version of IMS that enables the detection of target microorganisms from often low biomass environmental samples to be then further processed in various -omics (e.g., transcriptomics and metabolomics) studies to better characterize the metabolic properties.

In this study, using an antibody raised against *Desulfovibrio vugaris Hildenborough* cells were pulled down from a Hanford Groundwater sample taken from the 100H region of the Hanford Reach National Monument.

The organism pulled down from the site using this method and immunostained looks like:



Desulfovibrio RCH1 Analysis Test (unsaved changes)

Created by: Adam Arkin (aparkin)

Desulfovibrio.RCH1.Domains v2
DomainAnnotation
4 days ago by nlharris

DvH.GFFBA v1
FBAModel
4 days ago

DvH.DraftFBA v1
FBAModel
4 days ago

fba.cdg.on.Biolog-C-lac-N-acgam v1
FBA
4 days ago

fba.cdg.on.Carbon-L-Lactic-Acid v1
FBA

Assemble and Annotate Microbial Genome
v0.1.0

Build and Normalize Metagenomic Functional Abundance Data
v0.1.0

Build and Normalize Metagenomic Taxonomic Abundance Data
v0.1.0

Build Plant Metabolic Model

The organism pulled down from the site using this method and immunostained looks like:

You can find more about Desulfovibrio vulgaris as a species by looking at Wikipedia. But it is a sulfate reducing bacteria, a motile, obligate anaerobe, with an extraordinary number of two-component systems. Here is the standard electron micrograph from Wikipedia.

Here's what I am going to do:

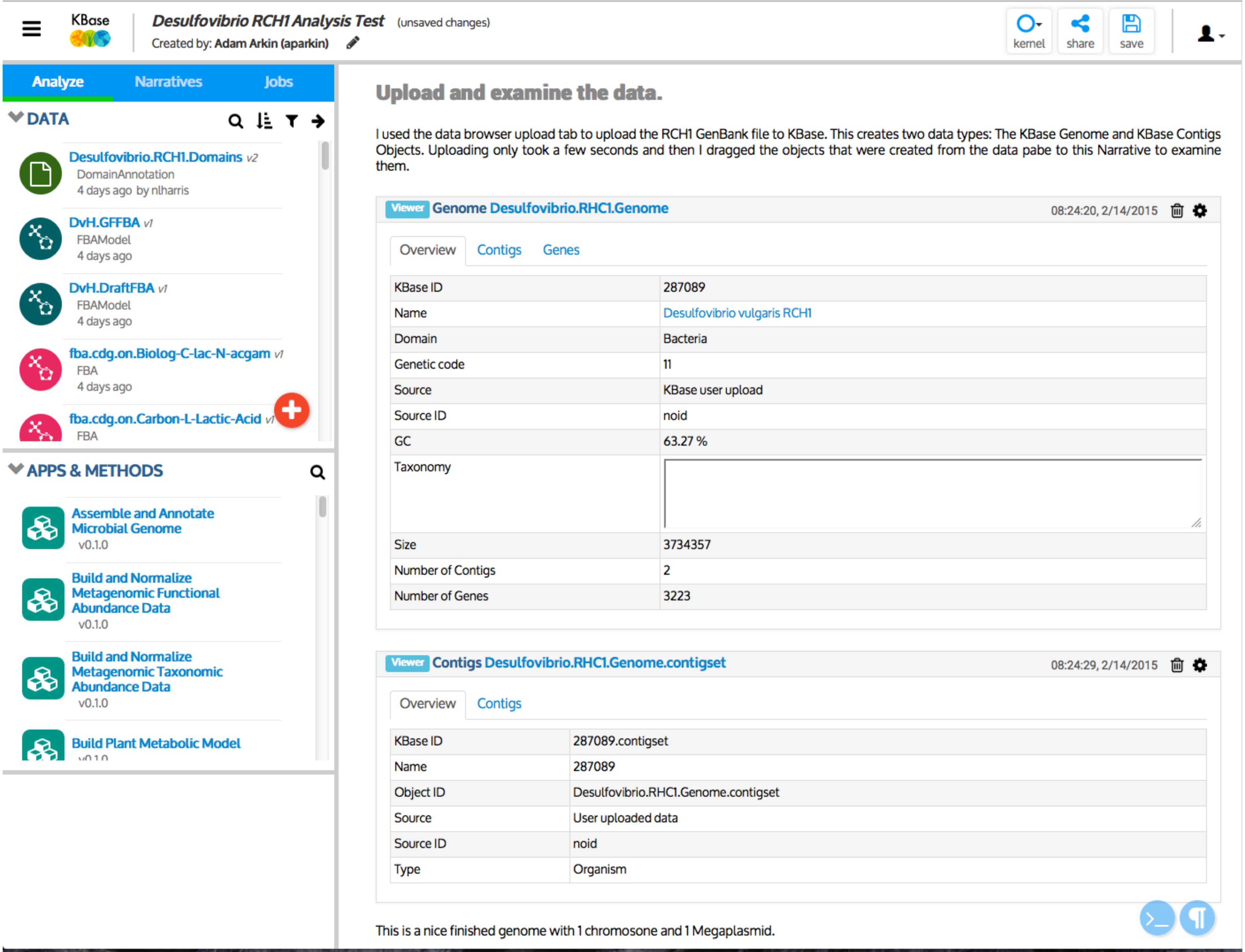
- Upload the genome
- Reannotate it for use in KBase.
- Annotate its domains for completeness
- Place it in a phylogenetic tree
- Compare it to the closest relative
- Try and understand its metabolic differences through comparing metabolic models.

Be aware though, I am not being rigorous here. Just giving a quick tour through KBase functionality for a realistic case.

Upload and examine the data.

I used the data browser upload tab to upload the RCH1 GenBank file to KBase. This creates two data types: The KBase Genome and KBase Contigs Objects. Uploading only took a few seconds and then I dragged the objects that were created from the data pabe to this Narrative to examine them.

Viewer Genome Desulfovibrio.RHC1.Genome 08:24:20, 2/14/2015



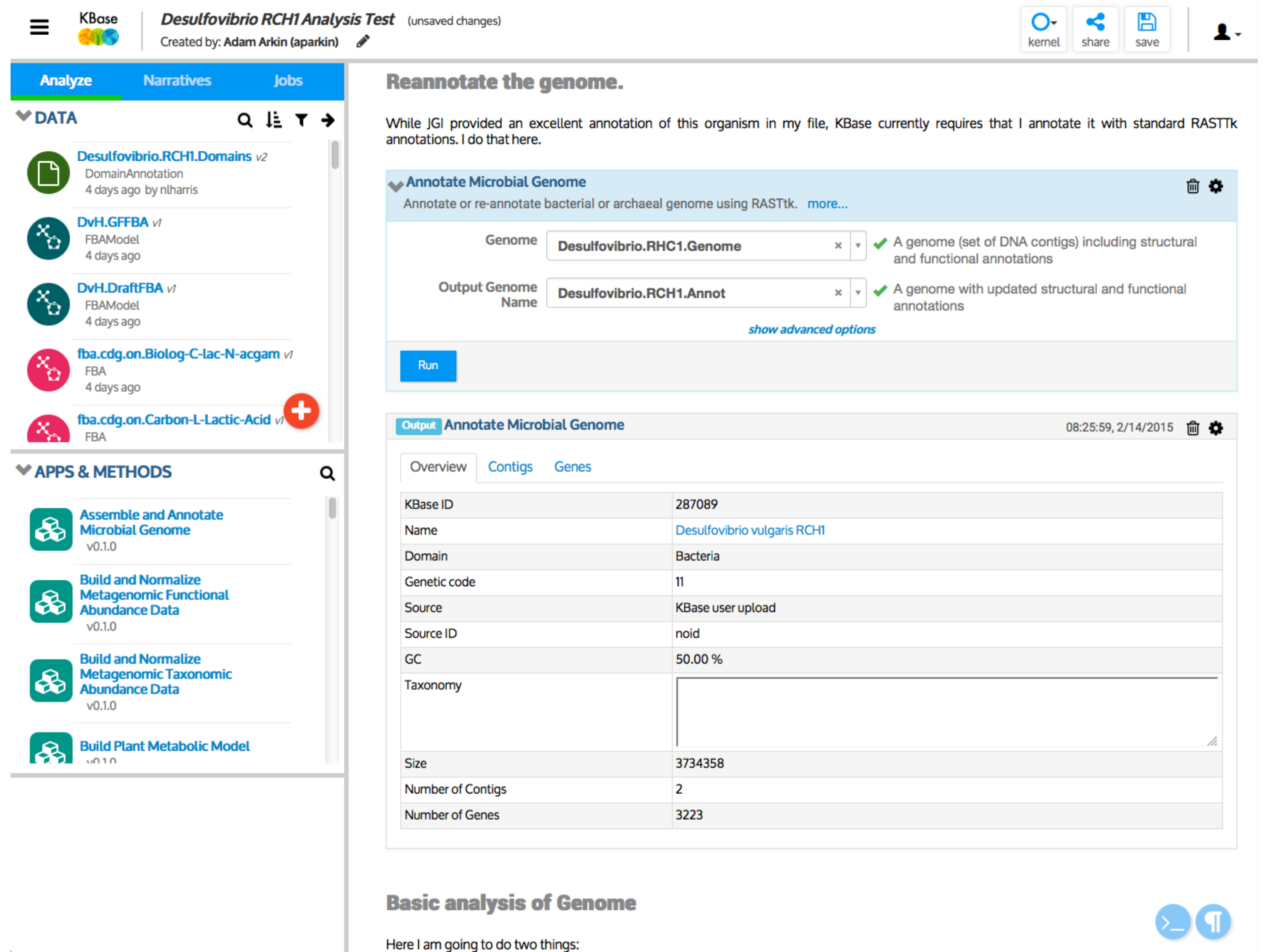
Upload and examine the data.

I used the data browser upload tab to upload the RCH1 GenBank file to KBase. This creates two data types: The KBase Genome and KBase Contigs Objects. Uploading only took a few seconds and then I dragged the objects that were created from the data pabe to this Narrative to examine them.

KBase ID	287089
Name	Desulfovibrio vulgaris RCH1
Domain	Bacteria
Genetic code	11
Source	KBase user upload
Source ID	noid
GC	63.27 %
Taxonomy	
Size	3734357
Number of Contigs	2
Number of Genes	3223

KBase ID	287089.contigset
Name	287089
Object ID	Desulfovibrio.RHC1.Genome.contigset
Source	User uploaded data
Source ID	noid
Type	Organism

This is a nice finished genome with 1 chromosone and 1 Megaplasmid.



Reannotate the genome.

While JGI provided an excellent annotation of this organism in my file, KBase currently requires that I annotate it with standard RASTTk annotations. I do that here.

Annotate Microbial Genome

Annotate or re-annotate bacterial or archaeal genome using RASTtk. more...

Genome: Desulfovibrio.RHC1.Genome ✓ A genome (set of DNA contigs) including structural and functional annotations

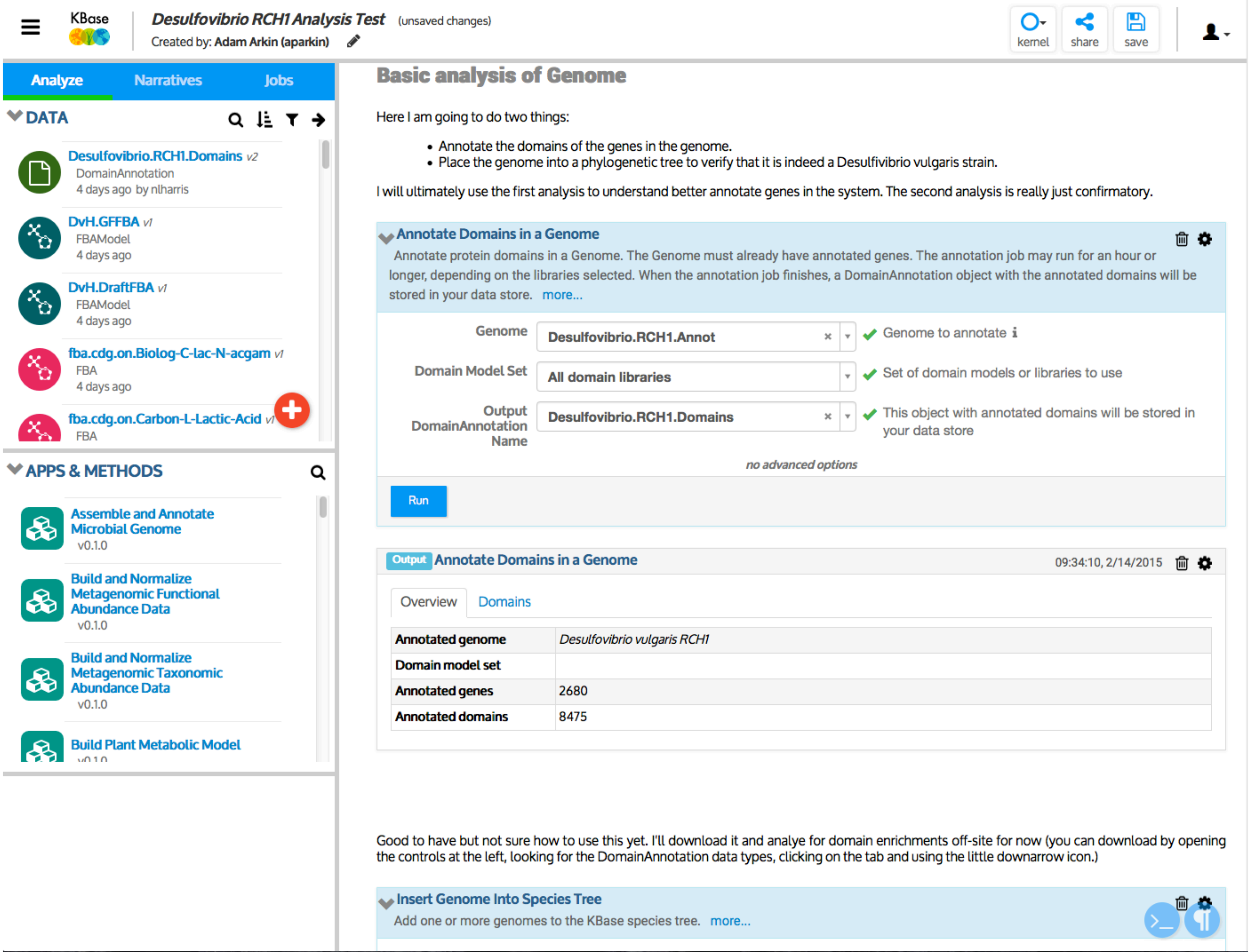
Output Genome Name: Desulfovibrio.RCH1.Annot ✓ A genome with updated structural and functional annotations

Output **Annotate Microbial Genome** 08:25:59, 2/14/2015

KBase ID	287089
Name	Desulfovibrio vulgaris RCH1
Domain	Bacteria
Genetic code	11
Source	KBase user upload
Source ID	noid
GC	50.00 %
Taxonomy	
Size	3734358
Number of Contigs	2
Number of Genes	3223

Basic analysis of Genome

Here I am going to do two things:



Desulfovibrio RCH1 Analysis Test (unsaved changes)

Created by: Adam Arkin (aparkin)

Basic analysis of Genome

Here I am going to do two things:

- Annotate the domains of the genes in the genome.
- Place the genome into a phylogenetic tree to verify that it is indeed a Desulfivibrio vulgaris strain.

I will ultimately use the first analysis to understand better annotate genes in the system. The second analysis is really just confirmatory.

Annotate Domains in a Genome

Annotate protein domains in a Genome. The Genome must already have annotated genes. The annotation job may run for an hour or longer, depending on the libraries selected. When the annotation job finishes, a DomainAnnotation object with the annotated domains will be stored in your data store. more...

Genome	Desulfovibrio.RCH1.Annot	✓ Genome to annotate
Domain Model Set	All domain libraries	✓ Set of domain models or libraries to use
Output DomainAnnotation Name	Desulfovibrio.RCH1.Domains	✓ This object with annotated domains will be stored in your data store

no advanced options

Run

Output **Annotate Domains in a Genome** 09:34:10, 2/14/2015

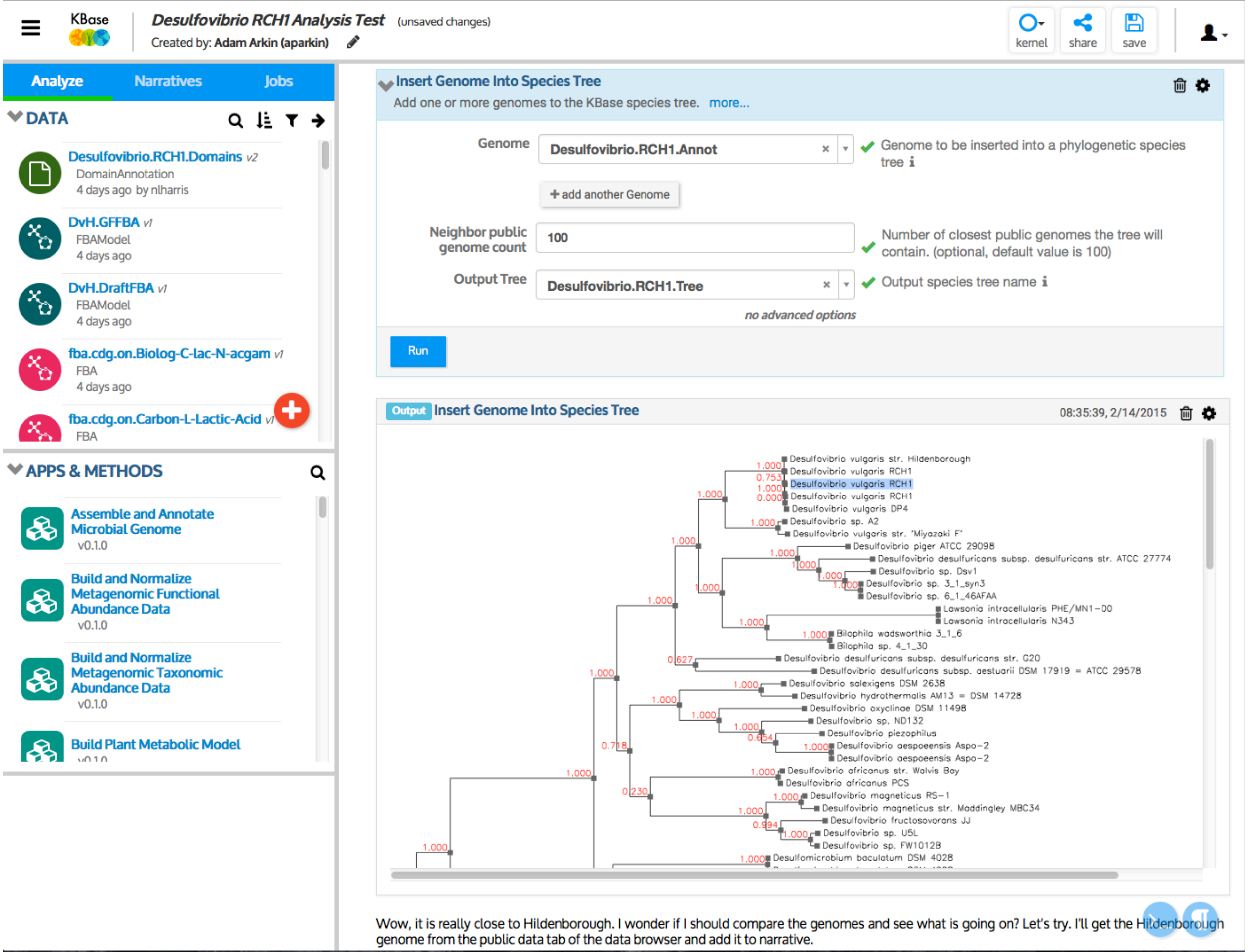
Overview Domains

Annotated genome	*Desulfovibrio vulgaris RCH1*
Domain model set	
Annotated genes	2680
Annotated domains	8475

Good to have but not sure how to use this yet. I'll download it and analye for domain enrichments off-site for now (you can download by opening the controls at the left, looking for the DomainAnnotation data types, clicking on the tab and using the little downarrow icon.)

Insert Genome Into Species Tree

Add one or more genomes to the KBase species tree. more...



Insert Genome Into Species Tree

Add one or more genomes to the KBase species tree. more...

Genome: Desulfovibrio.RCH1.Annot — Genome to be inserted into a phylogenetic species tree

Neighbor public genome count: 100 — Number of closest public genomes the tree will contain. (optional, default value is 100)

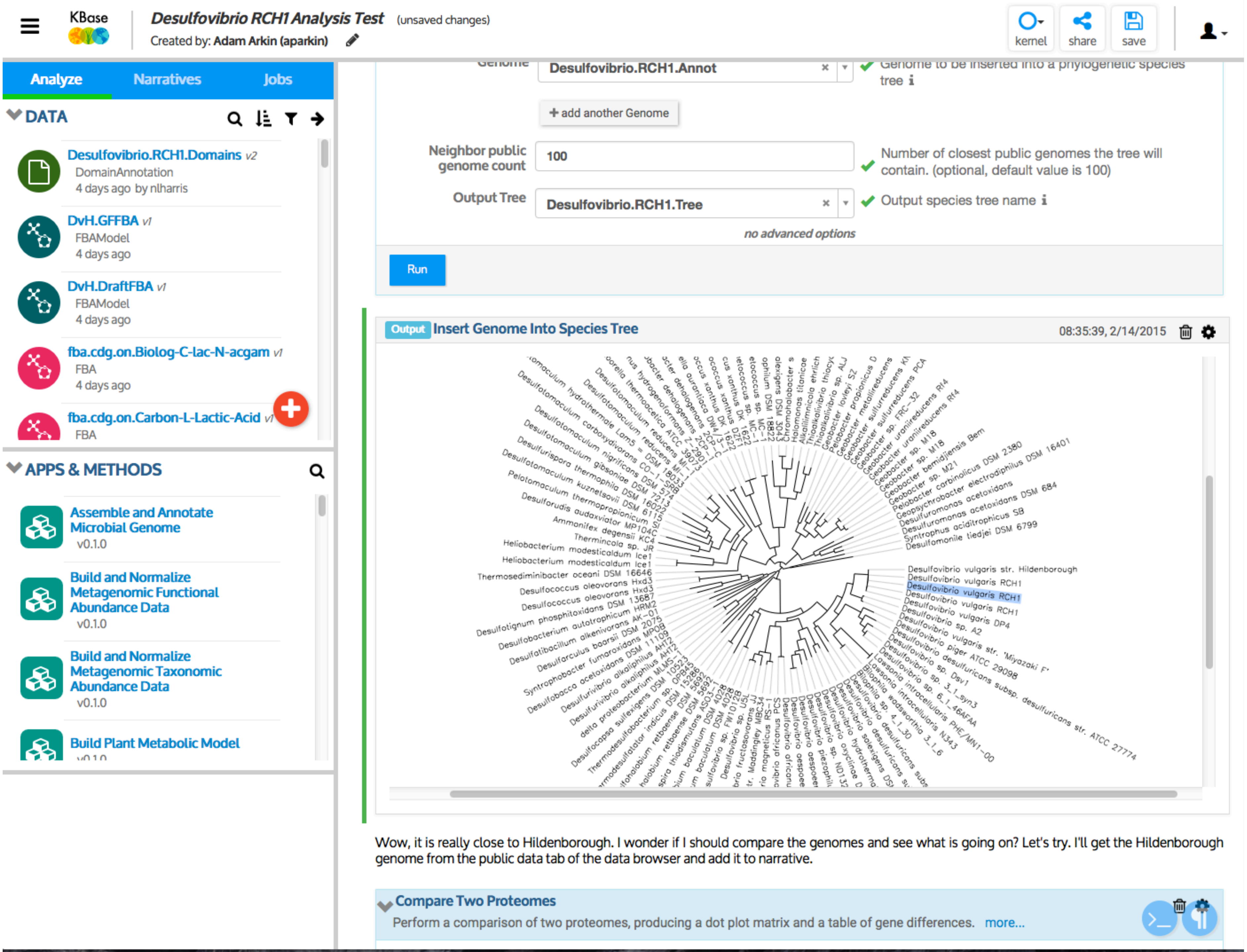
Output Tree: Desulfovibrio.RCH1.Tree — Output species tree name

no advanced options

Insert Genome Into Species Tree

08:35:39, 2/14/2015

Wow, it is really close to Hildenborough. I wonder if I should compare the genomes and see what is going on? Let's try. I'll get the Hildenborough genome from the public data tab of the data browser and add it to narrative.



Desulfovibrio RCH1 Analysis Test (unsaved changes)

Created by: Adam Arkin (aparkin)

Desulfovibrio.RCH1.Domains v2
DomainAnnotation
4 days ago by nlharris

DvH.GFFBA v1
FBAModel
4 days ago

DvH.DraftFBA v1
FBAModel
4 days ago

fba.cdg.on.Biolog-C-lac-N-acgam v1
FBA
4 days ago

fba.cdg.on.Carbon-L-Lactic-Acid v1
FBA

Assemble and Annotate Microbial Genome
v0.1.0

Build and Normalize Metagenomic Functional Abundance Data
v0.1.0

Build and Normalize Metagenomic Taxonomic Abundance Data
v0.1.0

Build Plant Metabolic Model

Genome: Desulfovibrio.RCH1.Annot — Genome to be inserted into a phylogenetic species tree

Neighbor public genome count: 100 — Number of closest public genomes the tree will contain. (optional, default value is 100)

Output Tree: Desulfovibrio.RCH1.Tree — Output species tree name

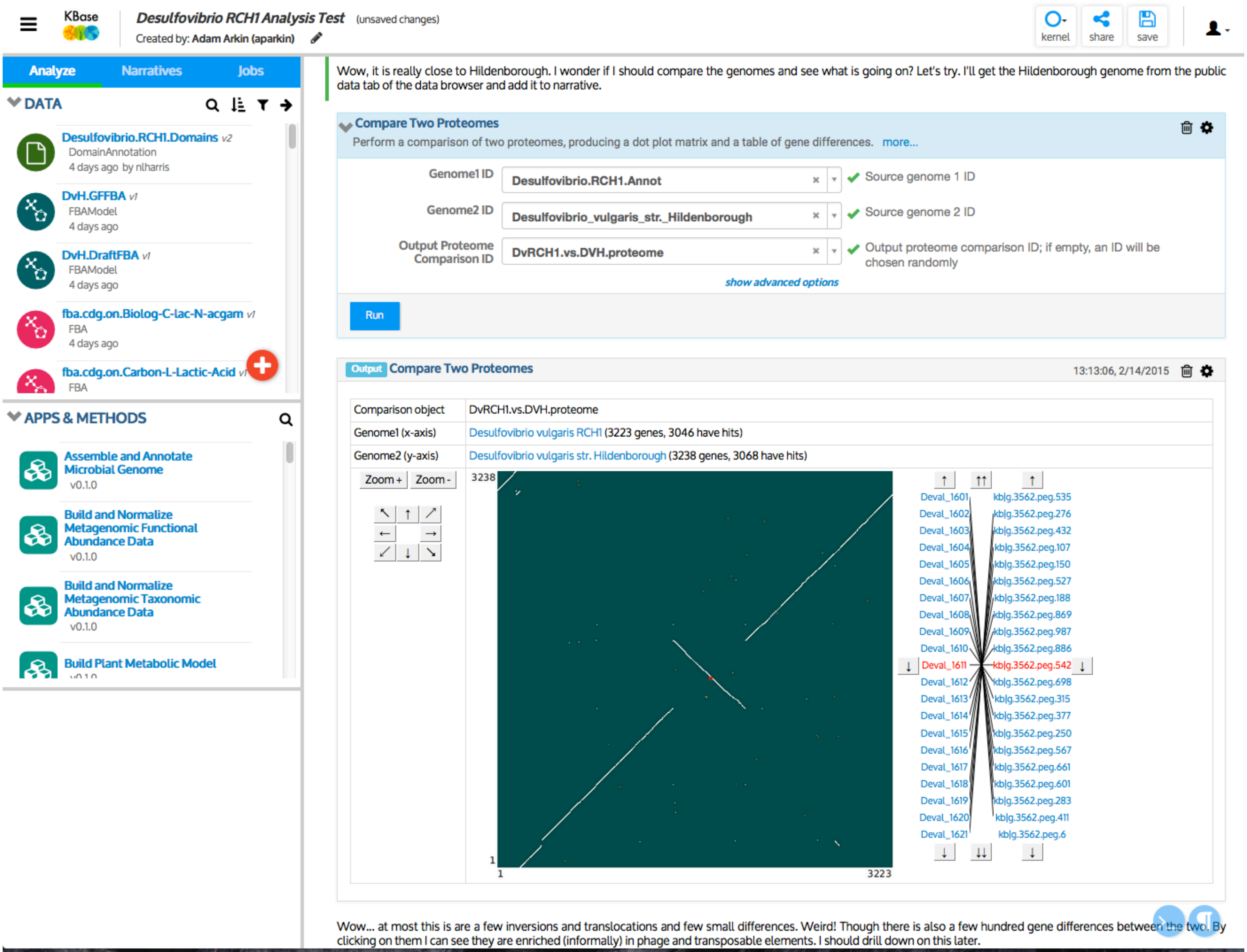
no advanced options

Output **Insert Genome Into Species Tree** 08:35:39, 2/14/2015

Wow, it is really close to Hildenborough. I wonder if I should compare the genomes and see what is going on? Let's try. I'll get the Hildenborough genome from the public data tab of the data browser and add it to narrative.

Compare Two Proteomes

Perform a comparison of two proteomes, producing a dot plot matrix and a table of gene differences. more...



Desulfovibrio RCH1 Analysis Test (unsaved changes)
Created by: Adam Arkin (aparkin)

Desulfovibrio.RCH1.Domains v2
DomainAnnotation
4 days ago by nlharris

DvH.GFFBA v1
FBAModel
4 days ago

DvH.DraftFBA v1
FBAModel
4 days ago

fba.cdg.on.Biolog-C-lac-N-acgam v1
FBA
4 days ago

fba.cdg.on.Carbon-L-Lactic-Acid v1
FBA

Assemble and Annotate Microbial Genome
v0.1.0

Build and Normalize Metagenomic Functional Abundance Data
v0.1.0

Build and Normalize Metagenomic Taxonomic Abundance Data
v0.1.0

Build Plant Metabolic Model

Wow, it is really close to Hildenborough. I wonder if I should compare the genomes and see what is going on? Let's try. I'll get the Hildenborough genome from the public data tab of the data browser and add it to narrative.

Compare Two Proteomes
Perform a comparison of two proteomes, producing a dot plot matrix and a table of gene differences. more...

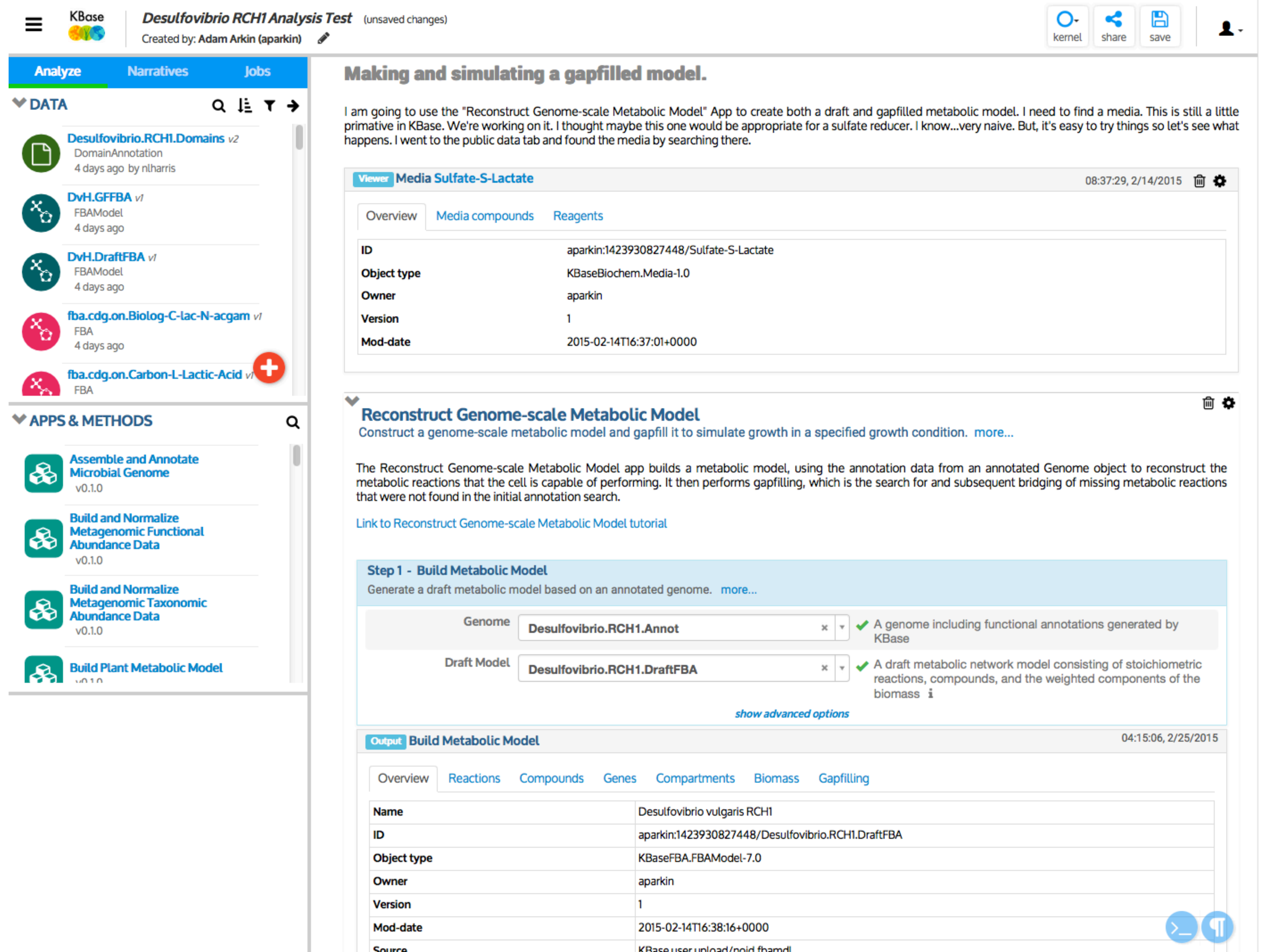
Genome1 ID: Desulfovibrio.RCH1.Annot ✓ Source genome 1 ID
Genome2 ID: Desulfovibrio_vulgaris_str._Hildenborough ✓ Source genome 2 ID
Output Proteome Comparison ID: DvRCH1.vs.DVH.proteome ✓ Output proteome comparison ID; if empty, an ID will be chosen randomly

show advanced options

Compare Two Proteomes 13:13:06, 2/14/2015

Comparison object	DvRCH1.vs.DVH.proteome
Genome1 (x-axis)	Desulfovibrio vulgaris RCH1 (3223 genes, 3046 have hits)
Genome2 (y-axis)	Desulfovibrio vulgaris str. Hildenborough (3238 genes, 3068 have hits)

Wow... at most this is are a few inversions and translocations and few small differences. Weird! Though there is also a few hundred gene differences between the two. By clicking on them I can see they are enriched (informally) in phage and transposable elements. I should drill down on this later.



Making and simulating a gapfilled model.

I am going to use the "Reconstruct Genome-scale Metabolic Model" App to create both a draft and gapfilled metabolic model. I need to find a media. This is still a little primative in KBase. We're working on it. I thought maybe this one would be appropriate for a sulfate reducer. I know...very naive. But, it's easy to try things so let's see what happens. I went to the public data tab and found the media by searching there.

Viewer **Media Sulfate-S-Lactate** 08:37:29, 2/14/2015

ID	aparkin:1423930827448/Sulfate-S-Lactate
Object type	KBaseBiochem.Media-1.0
Owner	aparkin
Version	1
Mod-date	2015-02-14T16:37:01+0000

Reconstruct Genome-scale Metabolic Model

Construct a genome-scale metabolic model and gapfill it to simulate growth in a specified growth condition. more...

The Reconstruct Genome-scale Metabolic Model app builds a metabolic model, using the annotation data from an annotated Genome object to reconstruct the metabolic reactions that the cell is capable of performing. It then performs gapfilling, which is the search for and subsequent bridging of missing metabolic reactions that were not found in the initial annotation search.

Link to Reconstruct Genome-scale Metabolic Model tutorial

Step 1 - Build Metabolic Model

Generate a draft metabolic model based on an annotated genome. more...

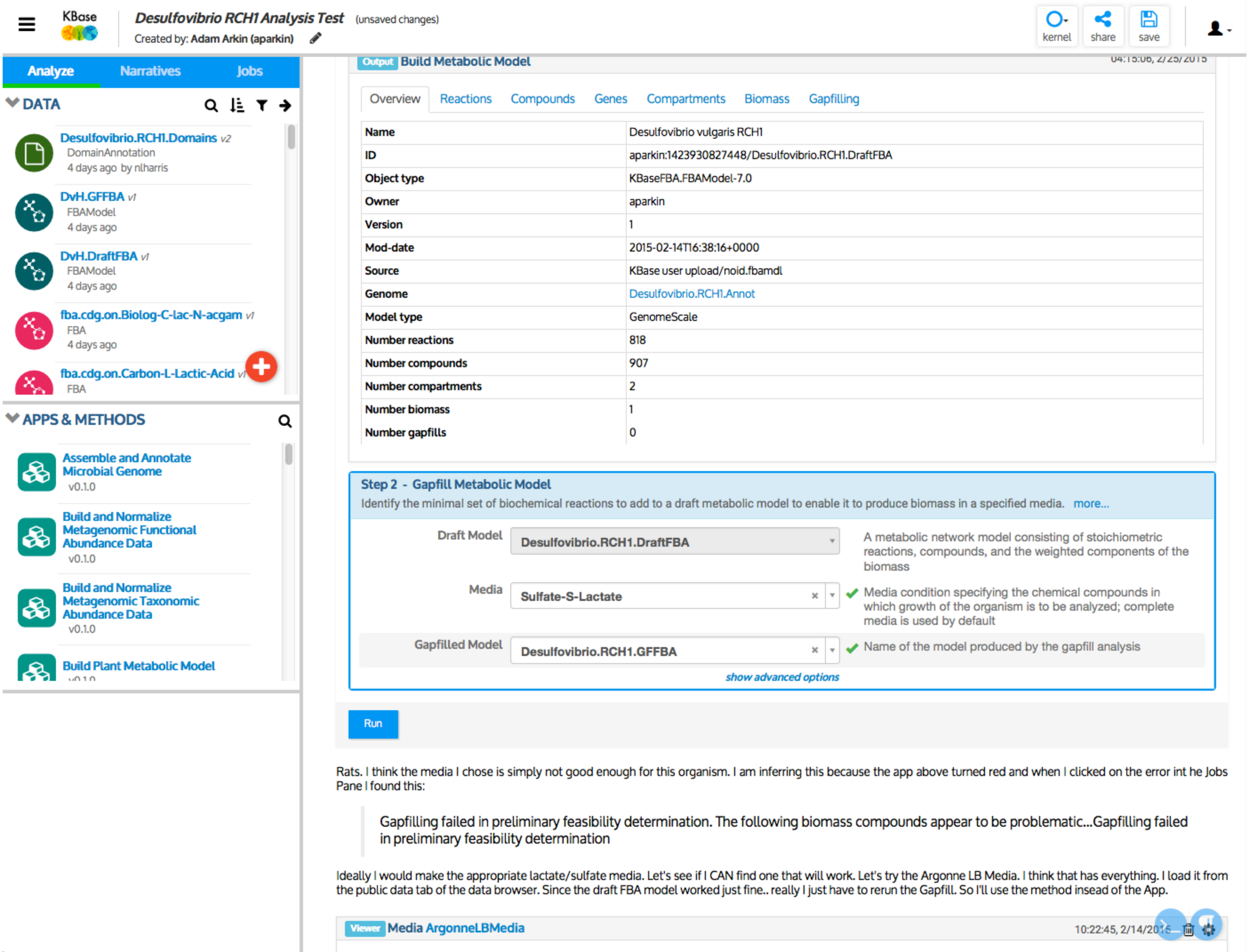
Genome: **Desulfovibrio.RCH1.Annot** ✓ A genome including functional annotations generated by KBase

Draft Model: **Desulfovibrio.RCH1.DraftFBA** ✓ A draft metabolic network model consisting of stoichiometric reactions, compounds, and the weighted components of the biomass

show advanced options

Output **Build Metabolic Model** 04:15:06, 2/25/2015

Name	Desulfovibrio vulgaris RCH1
ID	aparkin:1423930827448/Desulfovibrio.RCH1.DraftFBA
Object type	KBaseFBA.FBAModel-7.0
Owner	aparkin
Version	1
Mod-date	2015-02-14T16:38:16+0000
Source	KBase user upload/noid fbamdl



Name	Desulfovibrio vulgaris RCH1
ID	aparkin:1423930827448/Desulfovibrio.RCH1.DraftFBA
Object type	KBaseFBA.FBAModel-7.0
Owner	aparkin
Version	1
Mod-date	2015-02-14T16:38:16+0000
Source	KBase user upload/noid.fbamdl
Genome	Desulfovibrio.RCH1.Annot
Model type	GenomeScale
Number reactions	818
Number compounds	907
Number compartments	2
Number biomass	1
Number gapfills	0

Step 2 - Gapfill Metabolic Model

Identify the minimal set of biochemical reactions to add to a draft metabolic model to enable it to produce biomass in a specified media. more...

Draft Model: Desulfovibrio.RCH1.DraftFBA
A metabolic network model consisting of stoichiometric reactions, compounds, and the weighted components of the biomass

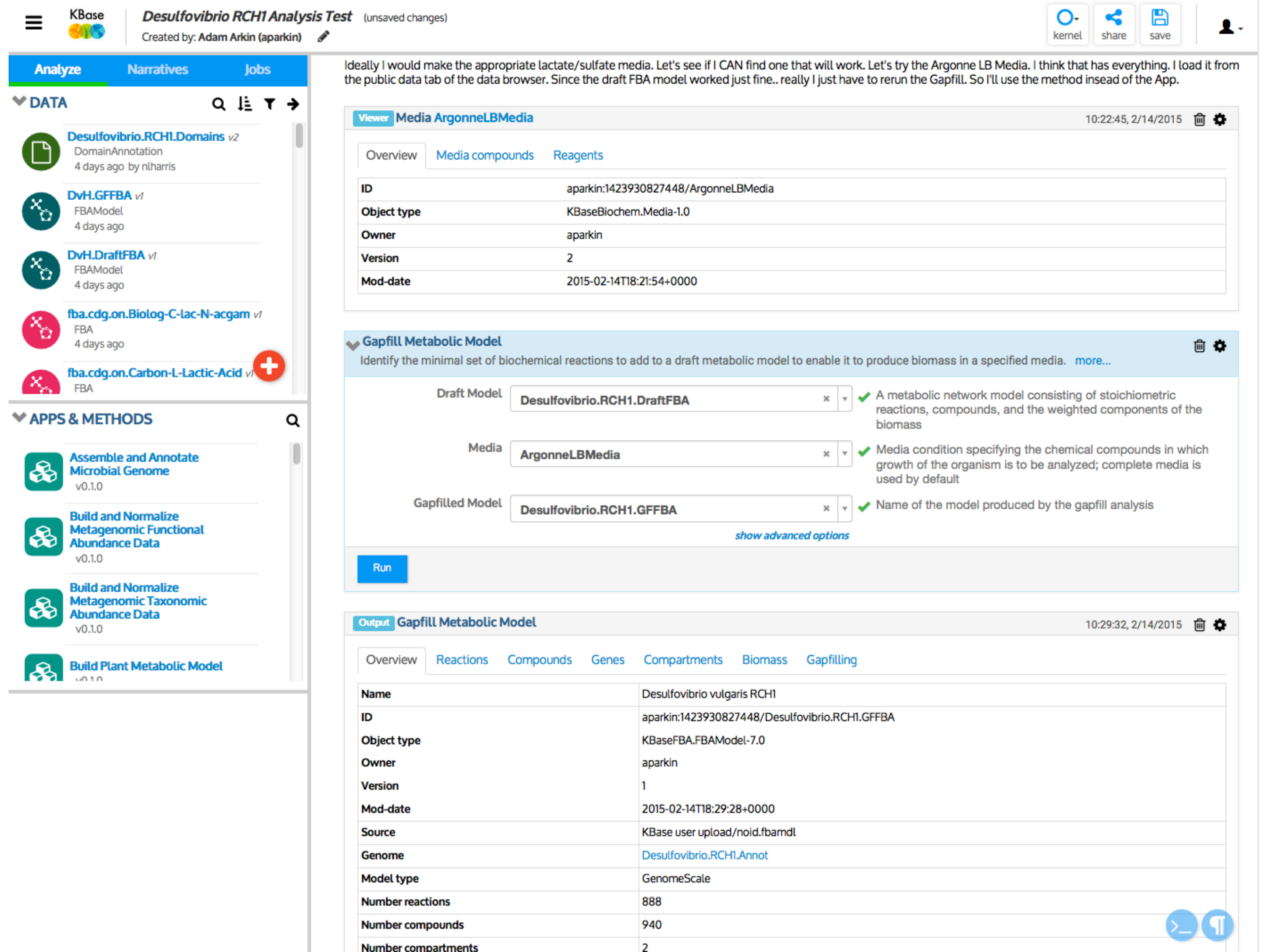
Media: Sulfate-S-Lactate
✓ Media condition specifying the chemical compounds in which growth of the organism is to be analyzed; complete media is used by default

Gapfilled Model: Desulfovibrio.RCH1.GFFBA
✓ Name of the model produced by the gapfill analysis

Rats. I think the media I chose is simply not good enough for this organism. I am inferring this because the app above turned red and when I clicked on the error int he Jobs Pane I found this:

> Gapfilling failed in preliminary feasibility determination. The following biomass compounds appear to be problematic...Gapfilling failed in preliminary feasibility determination

Ideally I would make the appropriate lactate/sulfate media. Let's see if I CAN find one that will work. Let's try the Argonne LB Media. I think that has everything. I load it from the public data tab of the data browser. Since the draft FBA model worked just fine.. really I just have to rerun the Gapfill. So I'll use the method insead of the App.



Desulfovibrio RCH1 Analysis Test (unsaved changes)
Created by: Adam Arkin (aparkin)

Desulfovibrio.RCH1.Domains v2
DomainAnnotation
4 days ago by nlharris

DvH.GFFBA v1
FBAModel
4 days ago

DvH.DraftFBA v1
FBAModel
4 days ago

fba.cdg.on.Biolog-C-lac-N-acgam v1
FBA
4 days ago

fba.cdg.on.Carbon-L-Lactic-Acid v1
FBA

Assemble and Annotate Microbial Genome
v0.1.0

Build and Normalize Metagenomic Functional Abundance Data
v0.1.0

Build and Normalize Metagenomic Taxonomic Abundance Data
v0.1.0

Build Plant Metabolic Model
v0.1.0

Ideally I would make the appropriate lactate/sulfate media. Let's see if I CAN find one that will work. Let's try the Argonne LB Media. I think that has everything. I load it from the public data tab of the data browser. Since the draft FBA model worked just fine.. really I just have to rerun the Gapfill. So I'll use the method insead of the App.

Viewer **Media ArgonneLBMedia** 10:22:45, 2/14/2015

Overview | Media compounds | Reagents

ID	aparkin:1423930827448/ArgonneLBMedia
Object type	KBaseBiochem.Media-1.0
Owner	aparkin
Version	2
Mod-date	2015-02-14T18:21:54+0000

Gapfill Metabolic Model

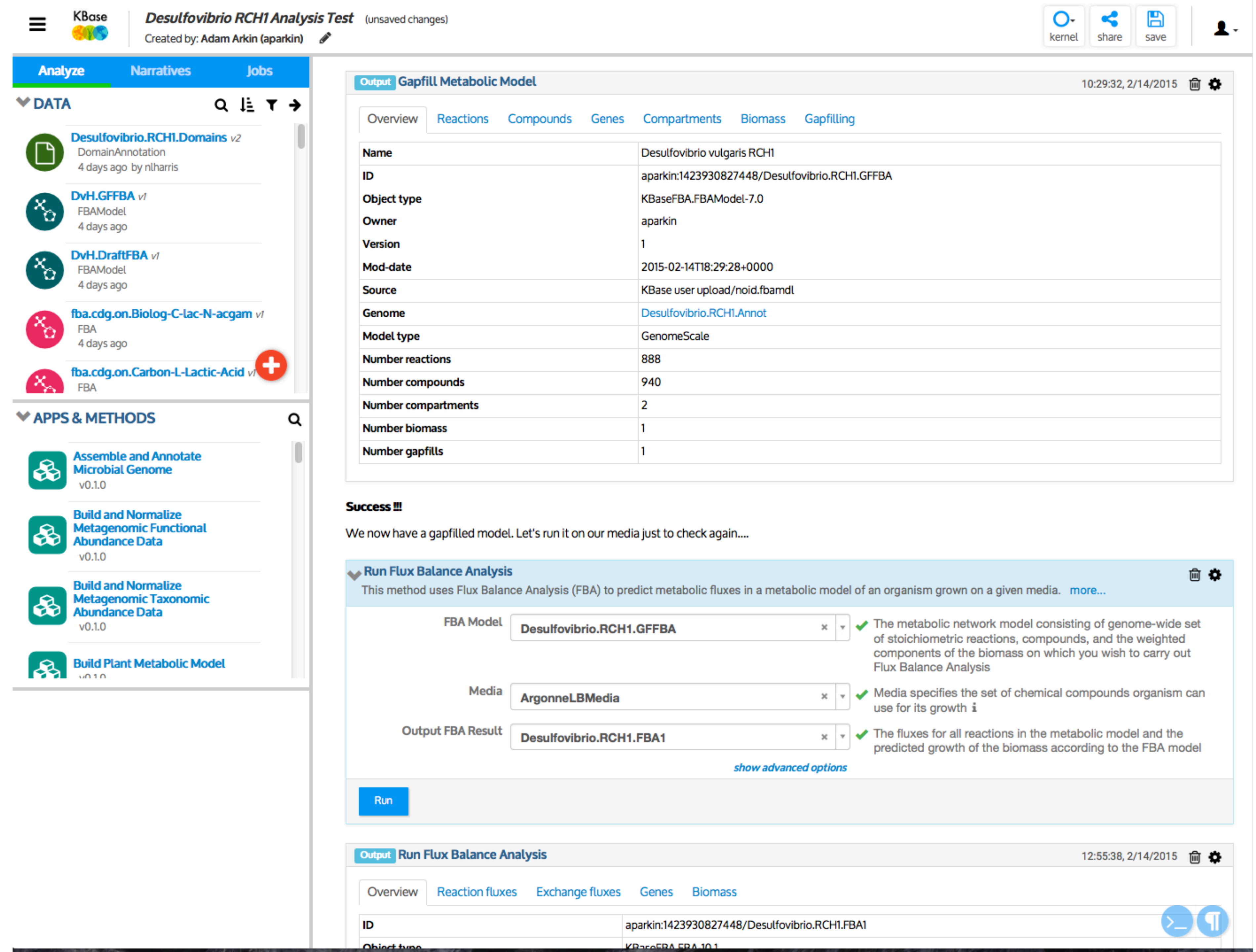
Identify the minimal set of biochemical reactions to add to a draft metabolic model to enable it to produce biomass in a specified media. more...

Draft Model	Desulfovibrio.RCH1.DraftFBA	✓ A metabolic network model consisting of stoichiometric reactions, compounds, and the weighted components of the biomass
Media	ArgonneLBMedia	✓ Media condition specifying the chemical compounds in which growth of the organism is to be analyzed; complete media is used by default
Gapfilled Model	Desulfovibrio.RCH1.GFFBA	✓ Name of the model produced by the gapfill analysis

Output **Gapfill Metabolic Model** 10:29:32, 2/14/2015

Overview | Reactions | Compounds | Genes | Compartments | Biomass | Gapfilling

Name	Desulfovibrio vulgaris RCH1
ID	aparkin:1423930827448/Desulfovibrio.RCH1.GFFBA
Object type	KBaseFBA.FBAModel-7.0
Owner	aparkin
Version	1
Mod-date	2015-02-14T18:29:28+0000
Source	KBase user upload/noid.fbamdl
Genome	Desulfovibrio.RCH1.Annot
Model type	GenomeScale
Number reactions	888
Number compounds	940
Number compartments	2



Name	Desulfovibrio vulgaris RCH1
ID	aparkin:1423930827448/Desulfovibrio.RCH1.GFFBA
Object type	KBaseFBA.FBAModel-7.0
Owner	aparkin
Version	1
Mod-date	2015-02-14T18:29:28+0000
Source	KBase user upload/noid.fbamdl
Genome	Desulfovibrio.RCH1.Annot
Model type	GenomeScale
Number reactions	888
Number compounds	940
Number compartments	2
Number biomass	1
Number gapfills	1

Success !!!

We now have a gapfilled model. Let's run it on our media just to check again....

Run Flux Balance Analysis

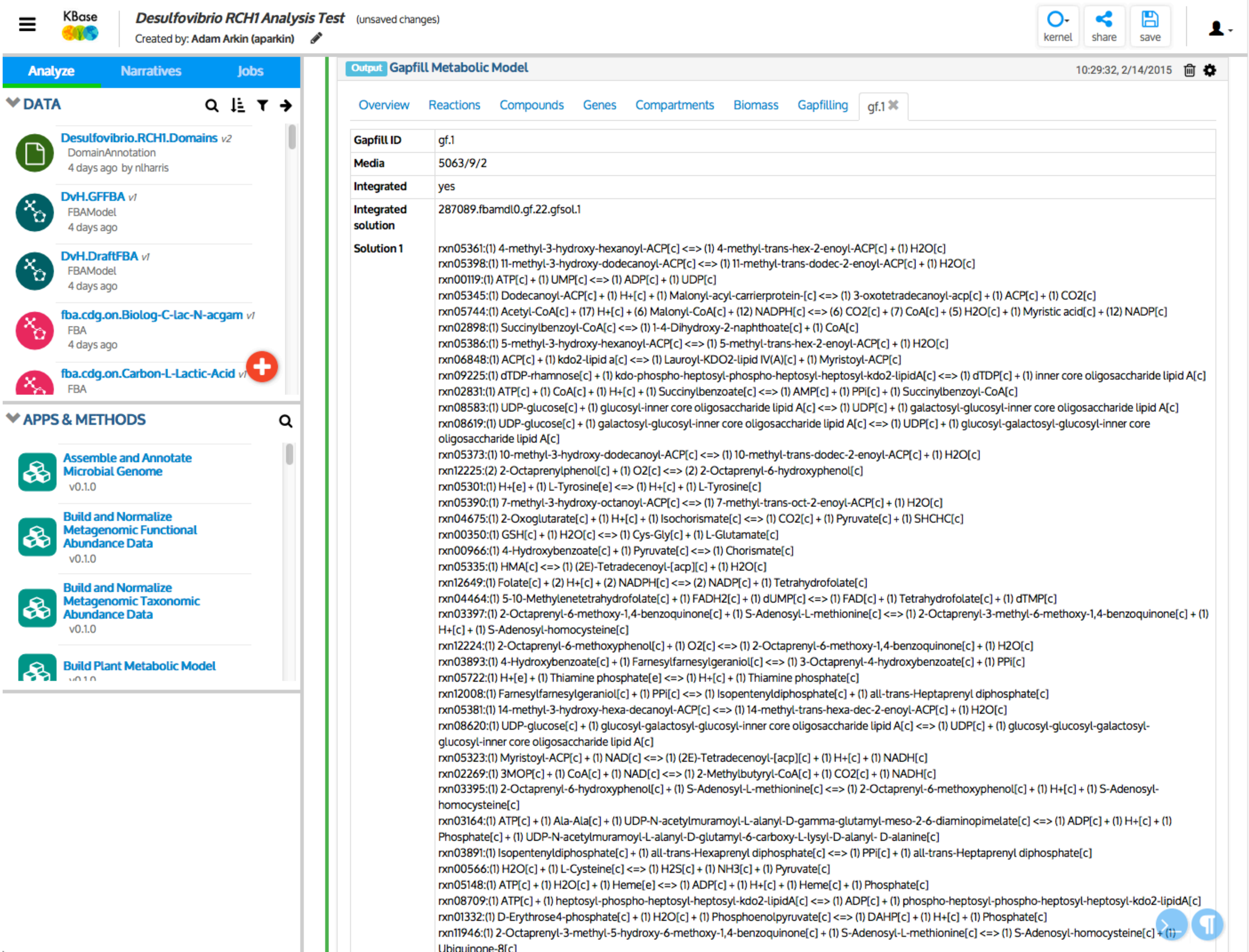
This method uses Flux Balance Analysis (FBA) to predict metabolic fluxes in a metabolic model of an organism grown on a given media. more...

FBA Model: **Desulfovibrio.RCH1.GFFBA**
✓ The metabolic network model consisting of genome-wide set of stoichiometric reactions, compounds, and the weighted components of the biomass on which you wish to carry out Flux Balance Analysis

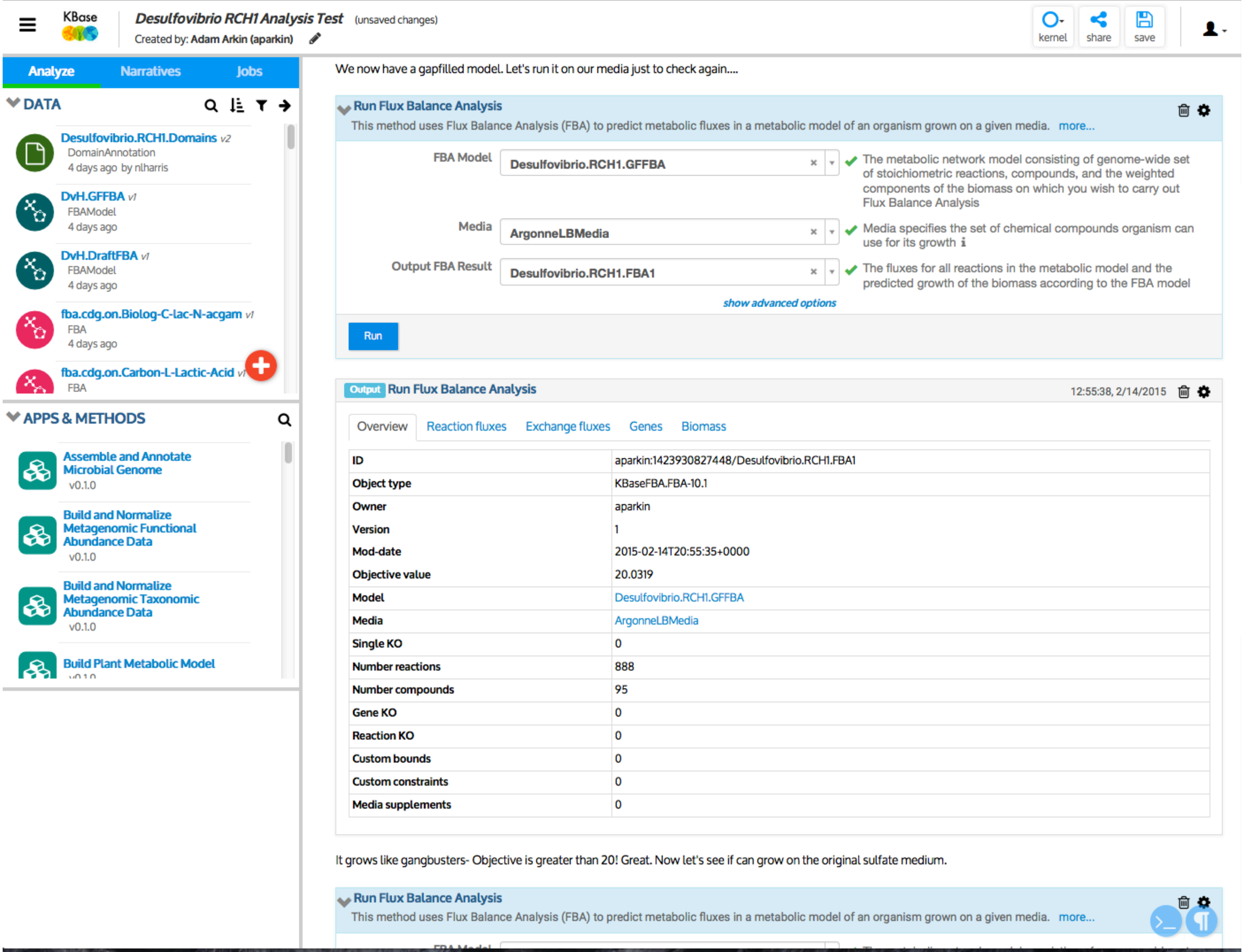
Media: **ArgonneLBMedia**
✓ Media specifies the set of chemical compounds organism can use for its growth

Output FBA Result: **Desulfovibrio.RCH1.FBA1**
✓ The fluxes for all reactions in the metabolic model and the predicted growth of the biomass according to the FBA model

ID	aparkin:1423930827448/Desulfovibrio.RCH1.FBA1
Object type	KBaseFBA.FBA-10.1



Gapfill ID	gf.1
Media	5063/9/2
Integrated	yes
Integrated solution	287089.fbamdl0.gf.22.gfsol.1
Solution 1	rxn05361:(1) 4-methyl-3-hydroxy-hexanoyl-ACP[c] <=> (1) 4-methyl-trans-hex-2-enoyl-ACP[c] + (1) H2O[c] rxn05398:(1) 11-methyl-3-hydroxy-dodecanoyl-ACP[c] <=> (1) 11-methyl-trans-dodec-2-enoyl-ACP[c] + (1) H2O[c] rxn00119:(1) ATP[c] + (1) UMP[c] <=> (1) ADP[c] + (1) UDP[c] rxn05345:(1) Dodecanoyl-ACP[c] + (1) H+[c] + (1) Malonyl-acyl-carrierprotein-[c] <=> (1) 3-oxotetradecanoyl-acp[c] + (1) ACP[c] + (1) CO2[c] rxn05744:(1) Acetyl-CoA[c] + (17) H+[c] + (6) Malonyl-CoA[c] + (12) NADPH[c] <=> (6) CO2[c] + (7) CoA[c] + (5) H2O[c] + (1) Myristic acid[c] + (12) NADP[c] rxn02898:(1) Succinylbenzoyl-CoA[c] <=> (1) 1-4-Dihydroxy-2-naphthoate[c] + (1) CoA[c] rxn05386:(1) 5-methyl-3-hydroxy-hexanoyl-ACP[c] <=> (1) 5-methyl-trans-hex-2-enoyl-ACP[c] + (1) H2O[c] rxn06848:(1) ACP[c] + (1) kdo2-lipid a[c] <=> (1) Lauroyl-KDO2-lipid IV(A)[c] + (1) Myristoyl-ACP[c] rxn09225:(1) dTDP-rhamnose[c] + (1) kdo-phospho-heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA[c] <=> (1) dTDP[c] + (1) inner core oligosaccharide lipid A[c] rxn02831:(1) ATP[c] + (1) CoA[c] + (1) H+[c] + (1) Succinylbenzoate[c] <=> (1) AMP[c] + (1) PPi[c] + (1) Succinylbenzoyl-CoA[c] rxn08583:(1) UDP-glucose[c] + (1) glucosyl-inner core oligosaccharide lipid A[c] <=> (1) UDP[c] + (1) galactosyl-glucosyl-inner core oligosaccharide lipid A[c] rxn08619:(1) UDP-glucose[c] + (1) galactosyl-glucosyl-inner core oligosaccharide lipid A[c] <=> (1) UDP[c] + (1) glucosyl-galactosyl-glucosyl-inner core oligosaccharide lipid A[c] rxn05373:(1) 10-methyl-3-hydroxy-dodecanoyl-ACP[c] <=> (1) 10-methyl-trans-dodec-2-enoyl-ACP[c] + (1) H2O[c] rxn12225:(2) 2-Octaprenylphenol[c] + (1) O2[c] <=> (2) 2-Octaprenyl-6-hydroxyphenol[c] rxn05301:(1) H+[e] + (1) L-Tyrosine[e] <=> (1) H+[c] + (1) L-Tyrosine[c] rxn05390:(1) 7-methyl-3-hydroxy-octanoyl-ACP[c] <=> (1) 7-methyl-trans-oct-2-enoyl-ACP[c] + (1) H2O[c] rxn04675:(1) 2-Oxoglutarate[c] + (1) H+[c] + (1) Isochorismate[c] <=> (1) CO2[c] + (1) Pyruvate[c] + (1) SHCHC[c] rxn00350:(1) GSH[c] + (1) H2O[c] <=> (1) Cys-Gly[c] + (1) L-Glutamate[c] rxn00966:(1) 4-Hydroxybenzoate[c] + (1) Pyruvate[c] <=> (1) Chorismate[c] rxn05335:(1) HMA[c] <=> (1) (2E)-Tetradecenoyl-[acp][c] + (1) H2O[c] rxn12649:(1) Folate[c] + (2) H+[c] + (2) NADPH[c] <=> (2) NADP[c] + (1) Tetrahydrofolate[c] rxn04464:(1) 5-10-Methylenetetrahydrofolate[c] + (1) FADH2[c] + (1) dUMP[c] <=> (1) FAD[c] + (1) Tetrahydrofolate[c] + (1) dTMP[c] rxn03397:(1) 2-Octaprenyl-6-methoxy-1,4-benzoquinone[c] + (1) S-Adenosyl-L-methionine[c] <=> (1) 2-Octaprenyl-3-methyl-6-methoxy-1,4-benzoquinone[c] + (1) H+[c] + (1) S-Adenosyl-homocysteine[c] rxn12224:(1) 2-Octaprenyl-6-methoxyphenol[c] + (1) O2[c] <=> (1) 2-Octaprenyl-6-methoxy-1,4-benzoquinone[c] + (1) H2O[c] rxn03893:(1) 4-Hydroxybenzoate[c] + (1) Farnesylfarnesylgeraniol[c] <=> (1) 3-Octaprenyl-4-hydroxybenzoate[c] + (1) PPi[c] rxn05722:(1) H+[e] + (1) Thiamine phosphate[e] <=> (1) H+[c] + (1) Thiamine phosphate[c] rxn12008:(1) Farnesylfarnesylgeraniol[c] + (1) PPi[c] <=> (1) Isopentenyldiphosphate[c] + (1) all-trans-Heptaprenyl diphosphate[c] rxn05381:(1) 14-methyl-3-hydroxy-hexa-decanoyl-ACP[c] <=> (1) 14-methyl-trans-hexa-dec-2-enoyl-ACP[c] + (1) H2O[c] rxn08620:(1) UDP-glucose[c] + (1) glucosyl-galactosyl-glucosyl-inner core oligosaccharide lipid A[c] <=> (1) UDP[c] + (1) glucosyl-glucosyl-galactosyl-glucosyl-inner core oligosaccharide lipid A[c] rxn05323:(1) Myristoyl-ACP[c] + (1) NAD[c] <=> (1) (2E)-Tetradecenoyl-[acp][c] + (1) H+[c] + (1) NADH[c] rxn02269:(1) 3MOP[c] + (1) CoA[c] + (1) NAD[c] <=> (1) 2-Methylbutyryl-CoA[c] + (1) CO2[c] + (1) NADH[c] rxn03395:(1) 2-Octaprenyl-6-hydroxyphenol[c] + (1) S-Adenosyl-L-methionine[c] <=> (1) 2-Octaprenyl-6-methoxyphenol[c] + (1) H+[c] + (1) S-Adenosyl-homocysteine[c] rxn03164:(1) ATP[c] + (1) Ala-Ala[c] + (1) UDP-N-acetylmuramoyl-L-alanyl-D-gamma-glutamyl-meso-2-6-diaminopimelate[c] <=> (1) ADP[c] + (1) H+[c] + (1) Phosphate[c] + (1) UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-carboxy-L-lysyl-D-alanyl- D-alanine[c] rxn03891:(1) Isopentenyldiphosphate[c] + (1) all-trans-Hexaprenyl diphosphate[c] <=> (1) PPi[c] + (1) all-trans-Heptaprenyl diphosphate[c] rxn00566:(1) H2O[c] + (1) L-Cysteine[c] <=> (1) H2S[c] + (1) NH3[c] + (1) Pyruvate[c] rxn05148:(1) ATP[c] + (1) H2O[c] + (1) Heme[e] <=> (1) ADP[c] + (1) H+[c] + (1) Heme[c] + (1) Phosphate[c] rxn08709:(1) ATP[c] + (1) heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA[c] <=> (1) ADP[c] + (1) phospho-heptosyl-phospho-heptosyl-heptosyl-kdo2-lipidA[c] rxn01332:(1) D-Erythrose4-phosphate[c] + (1) H2O[c] + (1) Phosphoenolpyruvate[c] <=> (1) DAHP[c] + (1) H+[c] + (1) Phosphate[c] rxn11946:(1) 2-Octaprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinone[c] + (1) S-Adenosyl-L-methionine[c] <=> (1) S-Adenosyl-homocysteine[c] + (1) Ubiquinone-8[c]



We now have a gapfilled model. Let's run it on our media just to check again....

Run Flux Balance Analysis

This method uses Flux Balance Analysis (FBA) to predict metabolic fluxes in a metabolic model of an organism grown on a given media. more...

FBA Model	Desulfovibrio.RCH1.GFFBA	✓ The metabolic network model consisting of genome-wide set of stoichiometric reactions, compounds, and the weighted components of the biomass on which you wish to carry out Flux Balance Analysis
Media	ArgonneLBMedia	✓ Media specifies the set of chemical compounds organism can use for its growth
Output FBA Result	Desulfovibrio.RCH1.FBA1	✓ The fluxes for all reactions in the metabolic model and the predicted growth of the biomass according to the FBA model

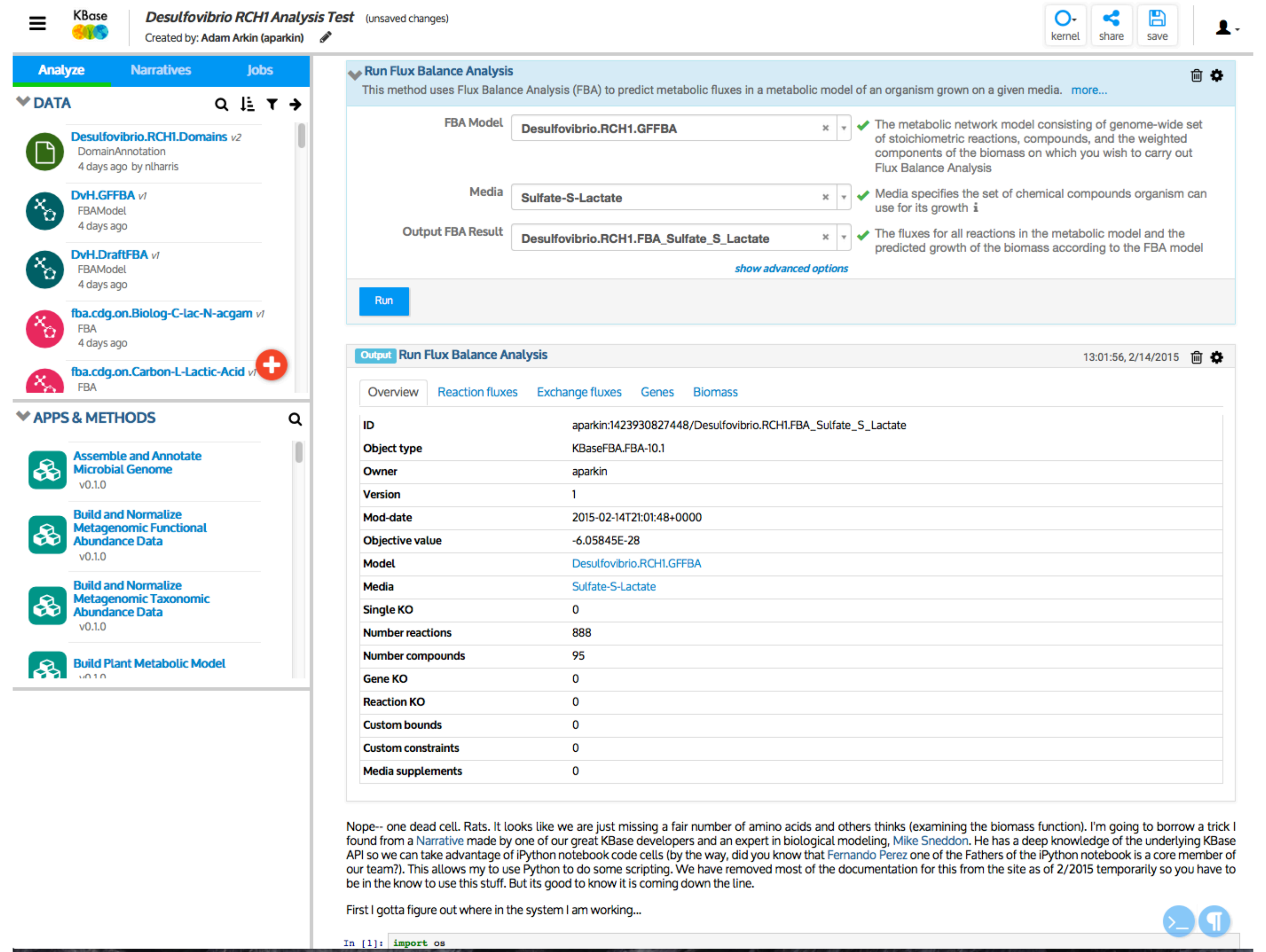
Output **Run Flux Balance Analysis** 12:55:38, 2/14/2015

ID	aparkin:1423930827448/Desulfovibrio.RCH1.FBA1
Object type	KBaseFBA.FBA-10.1
Owner	aparkin
Version	1
Mod-date	2015-02-14T20:55:35+0000
Objective value	20.0319
Model	Desulfovibrio.RCH1.GFFBA
Media	ArgonneLBMedia
Single KO	0
Number reactions	888
Number compounds	95
Gene KO	0
Reaction KO	0
Custom bounds	0
Custom constraints	0
Media supplements	0

It grows like gangbusters- Objective is greater than 20! Great. Now let's see if can grow on the original sulfate medium.

Run Flux Balance Analysis

This method uses Flux Balance Analysis (FBA) to predict metabolic fluxes in a metabolic model of an organism grown on a given media. more...



Run Flux Balance Analysis

This method uses Flux Balance Analysis (FBA) to predict metabolic fluxes in a metabolic model of an organism grown on a given media. more...

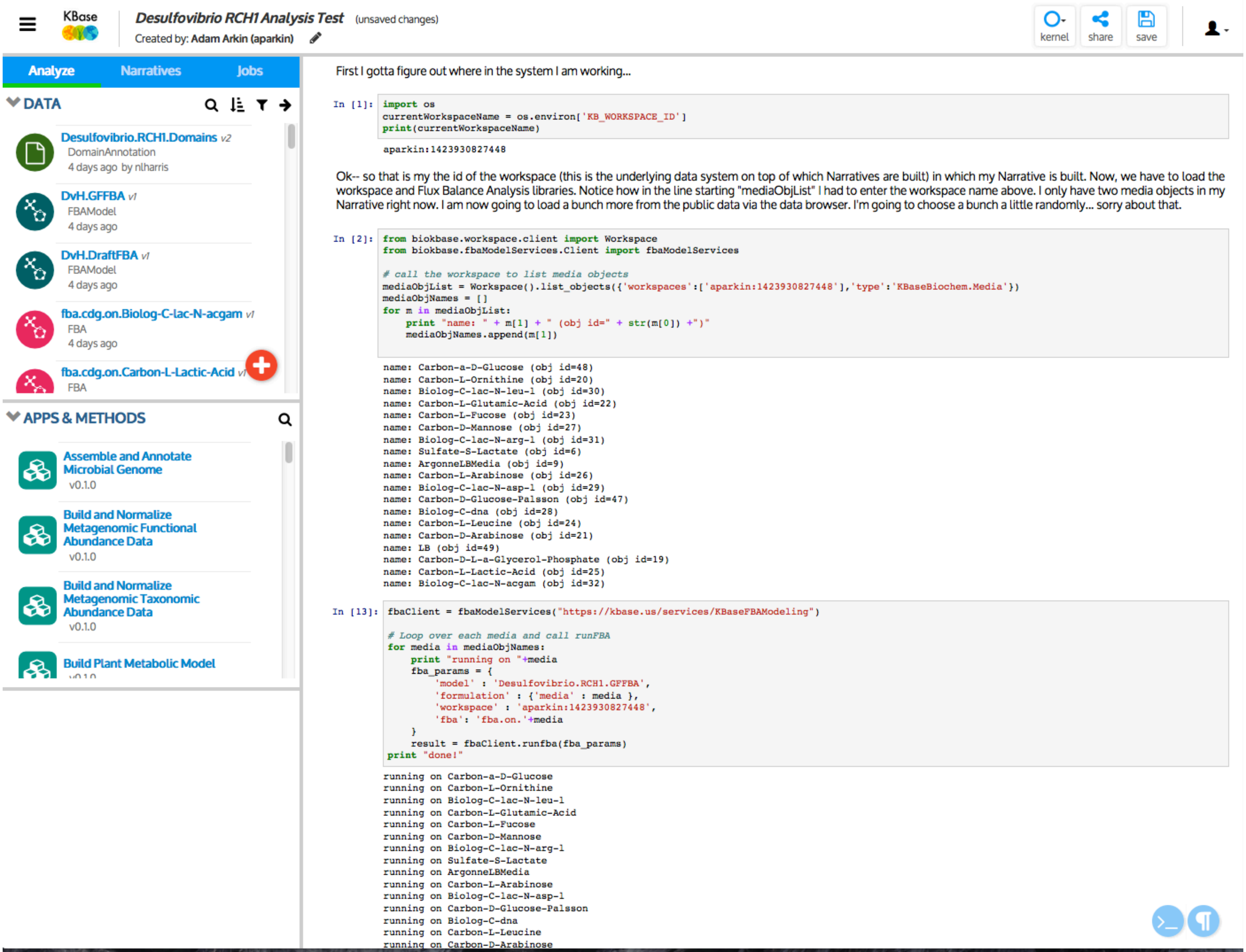
FBA Model	Desulfovibrio.RCH1.GFFBA	✓ The metabolic network model consisting of genome-wide set of stoichiometric reactions, compounds, and the weighted components of the biomass on which you wish to carry out Flux Balance Analysis
Media	Sulfate-S-Lactate	✓ Media specifies the set of chemical compounds organism can use for its growth
Output FBA Result	Desulfovibrio.RCH1.FBA_Sulfate_S_Lactate	✓ The fluxes for all reactions in the metabolic model and the predicted growth of the biomass according to the FBA model

ID	aparkin:1423930827448/Desulfovibrio.RCH1.FBA_Sulfate_S_Lactate
Object type	KBaseFBA.FBA-10.1
Owner	aparkin
Version	1
Mod-date	2015-02-14T21:01:48+0000
Objective value	-6.05845E-28
Model	Desulfovibrio.RCH1.GFFBA
Media	Sulfate-S-Lactate
Single KO	0
Number reactions	888
Number compounds	95
Gene KO	0
Reaction KO	0
Custom bounds	0
Custom constraints	0
Media supplements	0

Nope-- one dead cell. Rats. It looks like we are just missing a fair number of amino acids and others thinks (examining the biomass function). I'm going to borrow a trick I found from a Narrative made by one of our great KBase developers and an expert in biological modeling, Mike Sneddon. He has a deep knowledge of the underlying KBase API so we can take advantage of iPython notebook code cells (by the way, did you know that Fernando Perez one of the Fathers of the iPython notebook is a core member of our team?). This allows my to use Python to do some scripting. We have removed most of the documentation for this from the site as of 2/2015 temporarily so you have to be in the know to use this stuff. But its good to know it is coming down the line.

First I gotta figure out where in the system I am working...

```
In [1]: import os
```



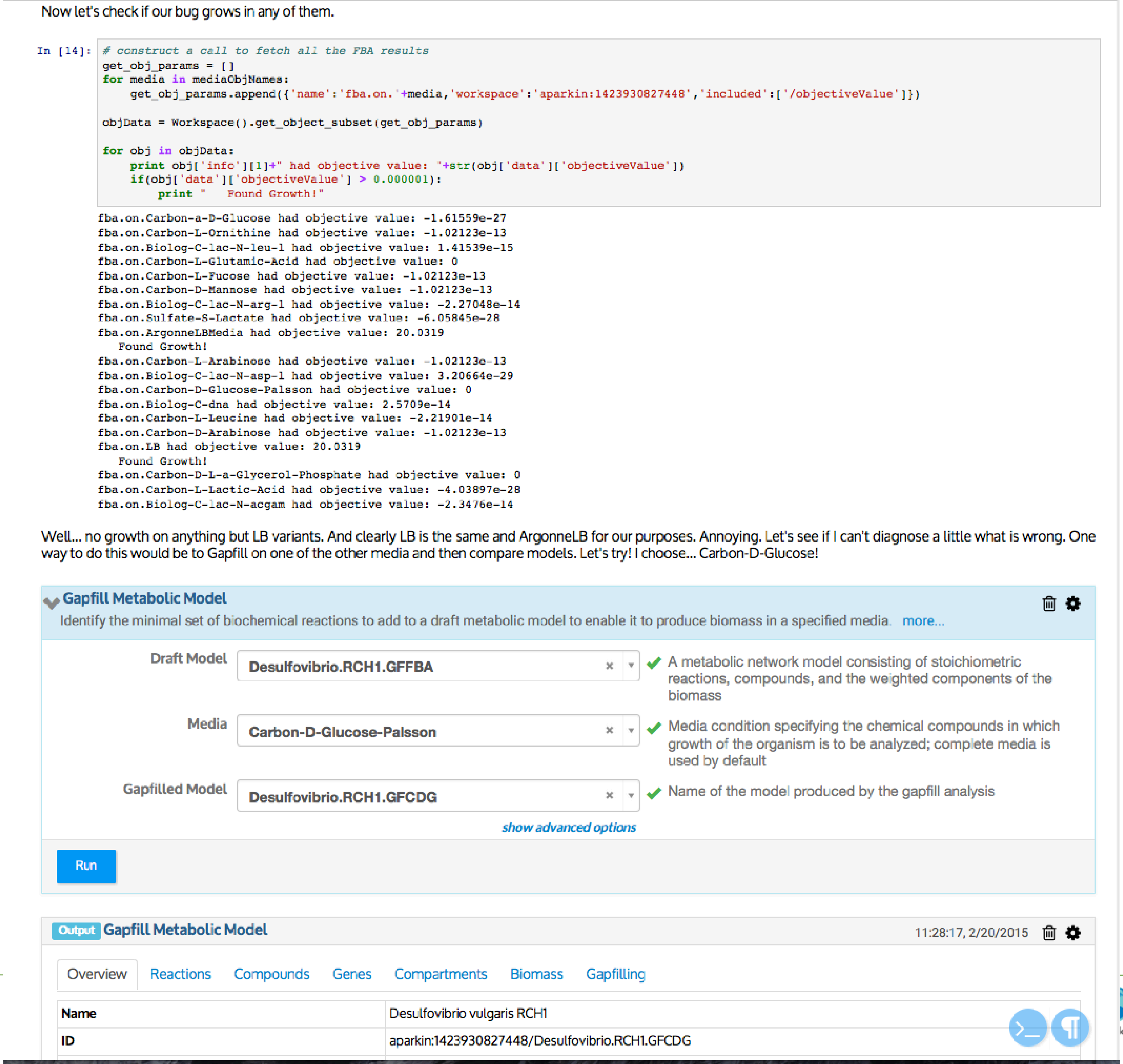
First I gotta figure out where in the system I am working...

```
In [1]: import os
        currentWorkspaceName = os.environ['KB_WORKSPACE_ID']
        print(currentWorkspaceName)

        aparkin:1423930827448
```

Ok-- so that is my the id of the workspace (this is the underlying data system on top of which Narratives are built) in which my Narrative is built. Now, we have to load the workspace and Flux Balance Analysis libraries. Notice how in the line starting "mediaObjList" I had to enter the workspace name above. I only have two media objects in my Narrative right now. I am now going to load a bunch more from the public data via the data browser. I'm going to choose a bunch a little randomly... sorry about that.

```
In [2]: from biokbase.workspace.client import Workspace
        from biokbase.fbaModelServices.Client import fbaModelServices

        # call the workspace to list media objects
        mediaObjList = Workspace().list_objects({'workspaces':['aparkin:1423930827448'],'type':'KBaseBiochem.Media'})
        mediaObjNames = []
        for m in mediaObjList:
            print "name: " + m[1] + " (obj id=" + str(m[0]) +")"
            mediaObjNames.append(m[1])

        name: Carbon-a-D-Glucose (obj id=48)
        name: Carbon-L-Ornithine (obj id=20)
        name: Biolog-C-lac-N-leu-l (obj id=30)
        name: Carbon-L-Glutamic-Acid (obj id=22)
        name: Carbon-L-Fucose (obj id=23)
        name: Carbon-D-Mannose (obj id=27)
        name: Biolog-C-lac-N-arg-l (obj id=31)
        name: Sulfate-S-Lactate (obj id=6)
        name: ArgonneLBMedia (obj id=9)
        name: Carbon-L-Arabinose (obj id=26)
        name: Biolog-C-lac-N-asp-l (obj id=29)
        name: Carbon-D-Glucose-Palsson (obj id=47)
        name: Biolog-C-dna (obj id=28)
        name: Carbon-L-Leucine (obj id=24)
        name: Carbon-D-Arabinose (obj id=21)
        name: LB (obj id=49)
        name: Carbon-D-L-a-Glycerol-Phosphate (obj id=19)
        name: Carbon-L-Lactic-Acid (obj id=25)
        name: Biolog-C-lac-N-acgam (obj id=32)
```

```
In [13]: fbaClient = fbaModelServices("https://kbase.us/services/KBaseFBAModeling")

         # Loop over each media and call runFBA
         for media in mediaObjNames:
             print "running on "+media
             fba_params = {
                 'model' : 'Desulfovibrio.RCH1.GFFBA',
                 'formulation' : {'media' : media },
                 'workspace' : 'aparkin:1423930827448',
                 'fba': 'fba.on.'+media
             }
             result = fbaClient.runfba(fba_params)
         print "done!"

        running on Carbon-a-D-Glucose
        running on Carbon-L-Ornithine
        running on Biolog-C-lac-N-leu-l
        running on Carbon-L-Glutamic-Acid
        running on Carbon-L-Fucose
        running on Carbon-D-Mannose
        running on Biolog-C-lac-N-arg-l
        running on Sulfate-S-Lactate
        running on ArgonneLBMedia
        running on Carbon-L-Arabinose
        running on Biolog-C-lac-N-asp-l
        running on Carbon-D-Glucose-Palsson
        running on Biolog-C-dna
        running on Carbon-L-Leucine
        running on Carbon-D-Arabinose
```



Now let's check if our bug grows in any of them.

In [14]:

```
# construct a call to fetch all the FBA results
get_obj_params = []
for media in mediaObjNames:
    get_obj_params.append({'name':'fba.on.'+media,'workspace':'aparkin:1423930827448','included':['/objectiveValue']})

objData = Workspace().get_object_subset(get_obj_params)

for obj in objData:
    print obj['info'][1]+" had objective value: "+str(obj['data']['objectiveValue'])
    if(obj['data']['objectiveValue'] > 0.000001):
        print "   Found Growth!"
```

```
fba.on.Carbon-a-D-Glucose had objective value: -1.61559e-27
fba.on.Carbon-L-Ornithine had objective value: -1.02123e-13
fba.on.Biolog-C-lac-N-leu-l had objective value: 1.41539e-15
fba.on.Carbon-L-Glutamic-Acid had objective value: 0
fba.on.Carbon-L-Fucose had objective value: -1.02123e-13
fba.on.Carbon-D-Mannose had objective value: -1.02123e-13
fba.on.Biolog-C-lac-N-arg-l had objective value: -2.27048e-14
fba.on.Sulfate-S-Lactate had objective value: -6.05845e-28
fba.on.ArgonneLBMedia had objective value: 20.0319
   Found Growth!
fba.on.Carbon-L-Arabinose had objective value: -1.02123e-13
fba.on.Biolog-C-lac-N-asp-l had objective value: 3.20664e-29
fba.on.Carbon-D-Glucose-Palsson had objective value: 0
fba.on.Biolog-C-dna had objective value: 2.5709e-14
fba.on.Carbon-L-Leucine had objective value: -2.21901e-14
fba.on.Carbon-D-Arabinose had objective value: -1.02123e-13
fba.on.LB had objective value: 20.0319
   Found Growth!
fba.on.Carbon-D-L-a-Glycerol-Phosphate had objective value: 0
fba.on.Carbon-L-Lactic-Acid had objective value: -4.03897e-28
fba.on.Biolog-C-lac-N-acgam had objective value: -2.3476e-14
```

Well... no growth on anything but LB variants. And clearly LB is the same and ArgonneLB for our purposes. Annoying. Let's see if I can't diagnose a little what is wrong. One way to do this would be to Gapfill on one of the other media and then compare models. Let's try! I choose... Carbon-D-Glucose!

Gapfill Metabolic Model

Identify the minimal set of biochemical reactions to add to a draft metabolic model to enable it to produce biomass in a specified media. more...

Draft Model: Desulfovibrio.RCH1.GFFBA — ✓ A metabolic network model consisting of stoichiometric reactions, compounds, and the weighted components of the biomass

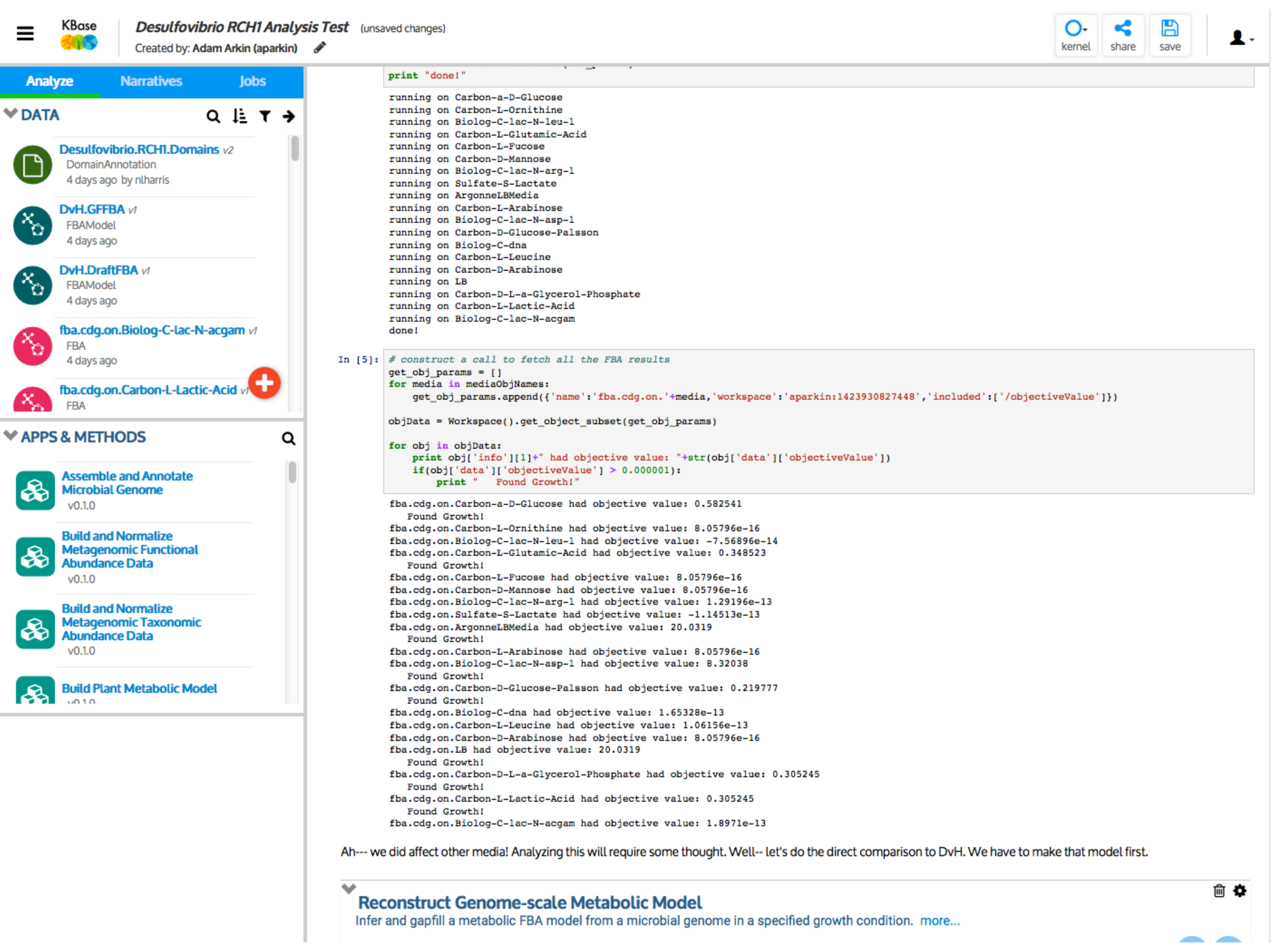
Media: Carbon-D-Glucose-Palsson — ✓ Media condition specifying the chemical compounds in which growth of the organism is to be analyzed; complete media is used by default

Gapfilled Model: Desulfovibrio.RCH1.GFCDG — ✓ Name of the model produced by the gapfill analysis

show advanced options

Output **Gapfill Metabolic Model** 11:28:17, 2/20/2015

Name	Desulfovibrio vulgaris RCH1
ID	aparkin:1423930827448/Desulfovibrio.RCH1.GFCDG



```
print "done!"
```

```
running on Carbon-a-D-Glucose
running on Carbon-L-Ornithine
running on Biolog-C-lac-N-leu-l
running on Carbon-L-Glutamic-Acid
running on Carbon-L-Fucose
running on Carbon-D-Mannose
running on Biolog-C-lac-N-arg-l
running on Sulfate-S-Lactate
running on ArgonneLBMedia
running on Carbon-L-Arabinose
running on Biolog-C-lac-N-asp-l
running on Carbon-D-Glucose-Palsson
running on Biolog-C-dna
running on Carbon-L-Leucine
running on Carbon-D-Arabinose
running on LB
running on Carbon-D-L-a-Glycerol-Phosphate
running on Carbon-L-Lactic-Acid
running on Biolog-C-lac-N-acgam
done!
```

In [5]:

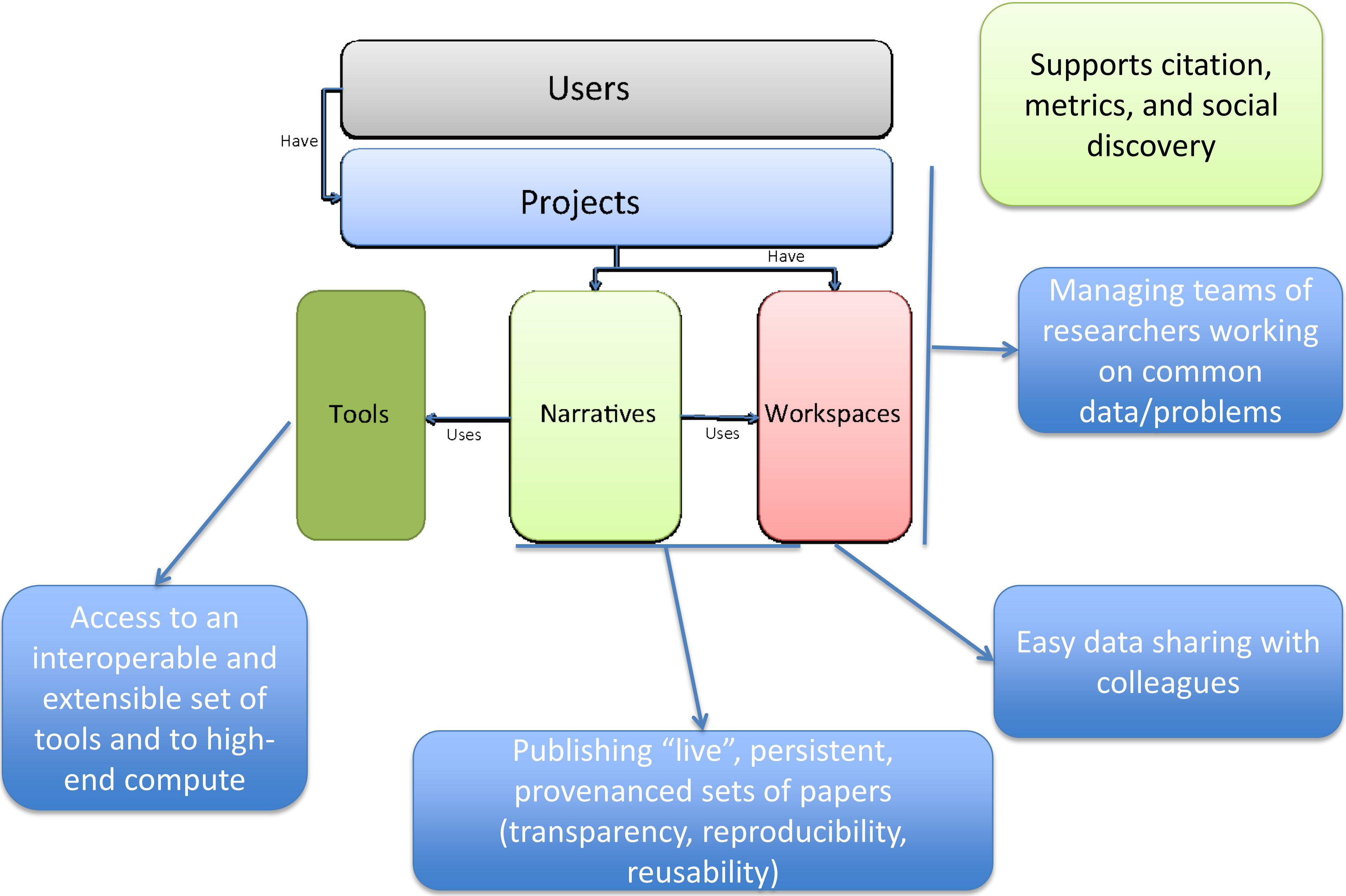
```
# construct a call to fetch all the FBA results
get_obj_params = []
for media in mediaObjNames:
    get_obj_params.append({'name':'fba.cdg.on.'+media,'workspace':'aparkin:1423930827448','included':['/objectiveValue']})

objData = Workspace().get_object_subset(get_obj_params)

for obj in objData:
    print obj['info'][1]+" had objective value: "+str(obj['data']['objectiveValue'])
    if(obj['data']['objectiveValue'] > 0.000001):
        print "   Found Growth!"
```

```
fba.cdg.on.Carbon-a-D-Glucose had objective value: 0.582541
   Found Growth!
fba.cdg.on.Carbon-L-Ornithine had objective value: 8.05796e-16
fba.cdg.on.Biolog-C-lac-N-leu-l had objective value: -7.56896e-14
fba.cdg.on.Carbon-L-Glutamic-Acid had objective value: 0.348523
   Found Growth!
fba.cdg.on.Carbon-L-Fucose had objective value: 8.05796e-16
fba.cdg.on.Carbon-D-Mannose had objective value: 8.05796e-16
fba.cdg.on.Biolog-C-lac-N-arg-l had objective value: 1.29196e-13
fba.cdg.on.Sulfate-S-Lactate had objective value: -1.14513e-13
fba.cdg.on.ArgonneLBMedia had objective value: 20.0319
   Found Growth!
fba.cdg.on.Carbon-L-Arabinose had objective value: 8.05796e-16
fba.cdg.on.Biolog-C-lac-N-asp-l had objective value: 8.32038
   Found Growth!
fba.cdg.on.Carbon-D-Glucose-Palsson had objective value: 0.219777
   Found Growth!
fba.cdg.on.Biolog-C-dna had objective value: 1.65328e-13
fba.cdg.on.Carbon-L-Leucine had objective value: 1.06156e-13
fba.cdg.on.Carbon-D-Arabinose had objective value: 8.05796e-16
fba.cdg.on.LB had objective value: 20.0319
   Found Growth!
fba.cdg.on.Carbon-D-L-a-Glycerol-Phosphate had objective value: 0.305245
   Found Growth!
fba.cdg.on.Carbon-L-Lactic-Acid had objective value: 0.305245
   Found Growth!
fba.cdg.on.Biolog-C-lac-N-acgam had objective value: 1.8971e-13
```

Ah--- we did affect other media! Analyzing this will require some thought. Well-- let's do the direct comparison to DvH. We have to make that model first.

Reconstruct Genome-scale Metabolic Model

Infer and gapfill a metabolic FBA model from a microbial genome in a specified growth condition. more...



Users

Have

Projects

Have

Tools

Uses

Narratives

Uses

Workspaces

Supports citation, metrics, and social discovery

Managing teams of researchers working on common data/problems

Access to an interoperable and extensible set of tools and to high-end compute

Publishing "live", persistent, provenanced sets of papers (transparency, reproducibility, reusability)

Easy data sharing with colleagues

What next?

- Way more functionality.
- Better and more complete data update cycle.
- Improved social interface
- Improved reliability and resource utilization
- Controlled reintroduction of routes to 3rd party software integration.
- Expanded and regularized data update cycle + improved data model
- Next-generation design for scalability and **meta-analysis**