



## DOE Systems Biology Knowledgebase

The KBase Platform for Dissemination of Tools and Analysis of Microbes, Plants and Their Communities: Examples from the users

---

INTEGRATION and  
MODELING *for*  
PREDICTIVE BIOLOGY

---



U.S. DEPARTMENT OF  
**ENERGY**


Office of  
Science

Office of Biological and Environmental Research




# KBase Team

*KBase is a multi-institutional collaboration with participation from these laboratories and universities*



Lawrence  
Berkeley  
National  
Laboratory

*Lead institution*



Argonne  
National  
Laboratory



Brookhaven  
National  
Laboratory



Oak Ridge  
National  
Laboratory



Chris Henry, ANL

Cold  
Spring  
Harbor  
Laboratory

University  
of  
Tennessee



Bob Cottingham, ORNL

# What is KBase?

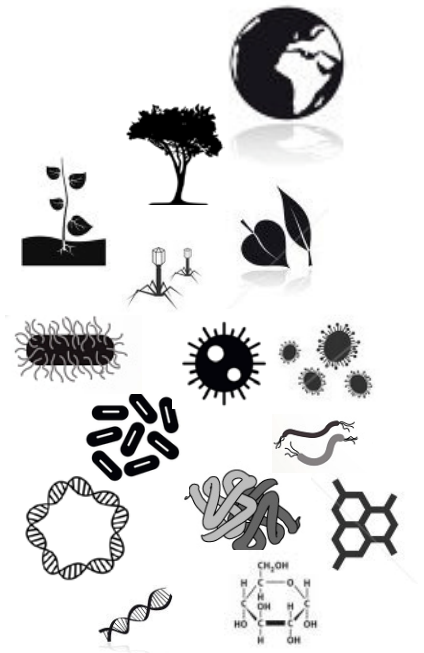
KBase is a knowledge creation and discovery environment

# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

## Data about...

- ❖ KBase is designed to accelerate research about , microorganisms, plants and their communities in environmental context with emphasis on DOE goals.

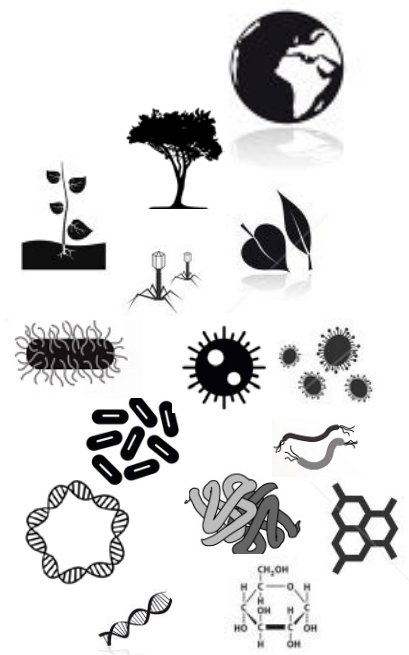


# KBase is a knowledge creation and discovery environment

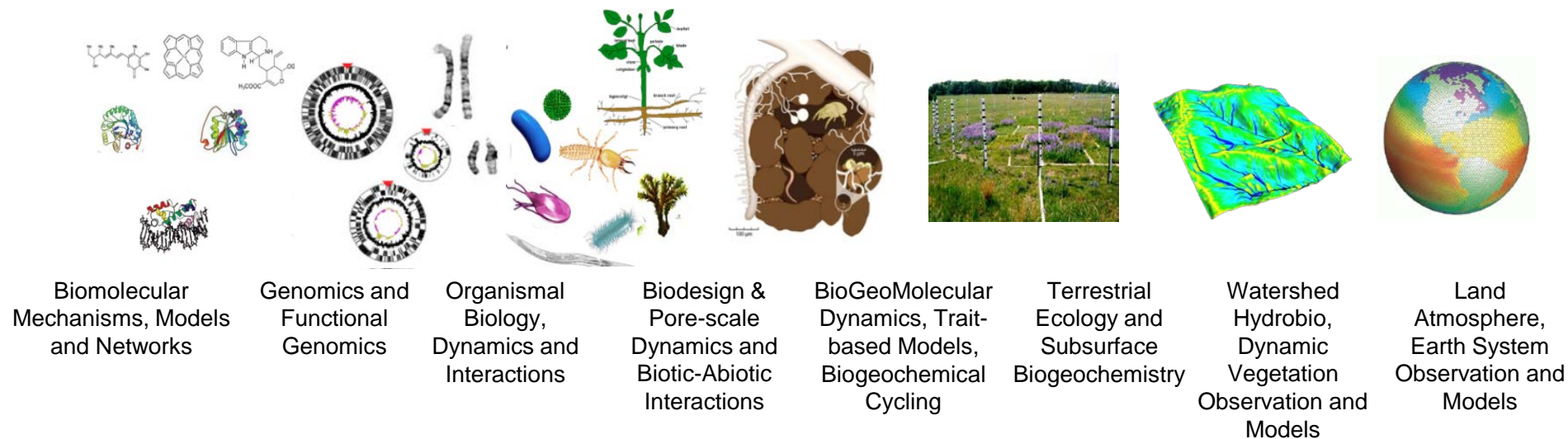
Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

## Data about...

❖ KBase is designed to accelerate research about microorganisms, plants and their communities in environmental context with emphasis on DOE goals.



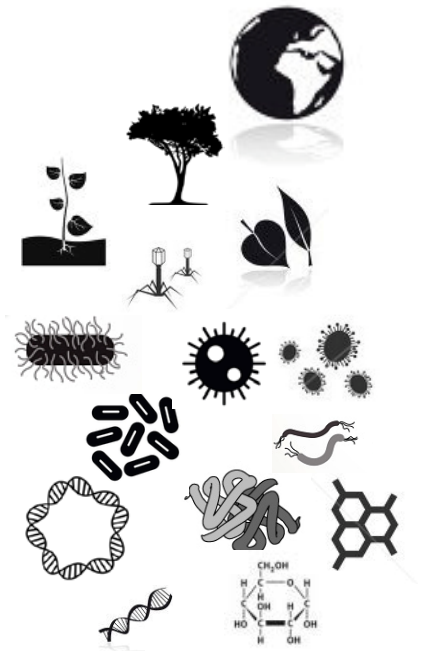
### KBase Scope of Operations



# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

## Data about...



❖ KBase is designed to accelerate research about microorganisms, plants and their communities in environmental context with emphasis on DOE goals.

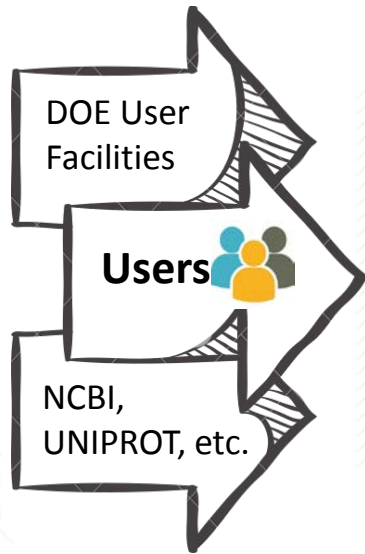
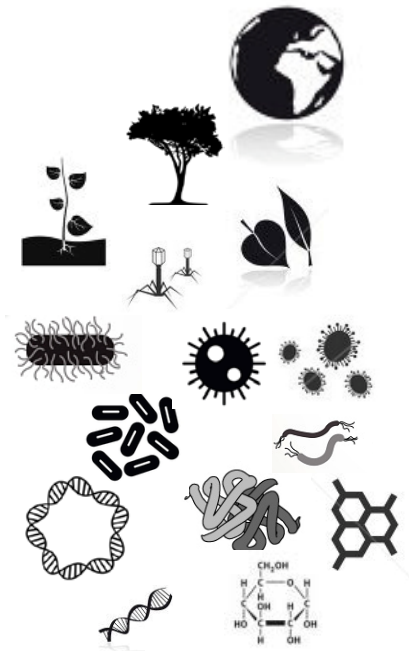
- Understand the biological complexity of plant and microbial metabolism and interfaces across scales spanning molecules to ecosystems.

- Scalable data processing, data analysis, machine learning, discrete algorithms, and multiscale multiphysical simulation are crucial for advancement of biological and environmental systems science.
- Innovations in representation, search, and visualization of large-scale, heterogeneous, ontologically rich primary and derived biological and contextual data (e.g., abiotic environmental information) are crucial for input to and validation of these methods.
- New architectures, data transport protocols, software libraries, and languages are necessary to create a platform for community tool development and use supporting interactive and seamless interoperation of both mid- and large-scale cluster resources and enterprise-class computing environments.

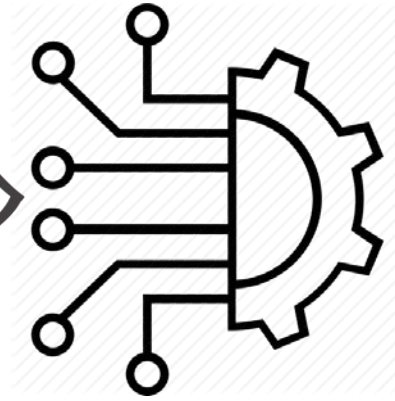
# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

Data about...

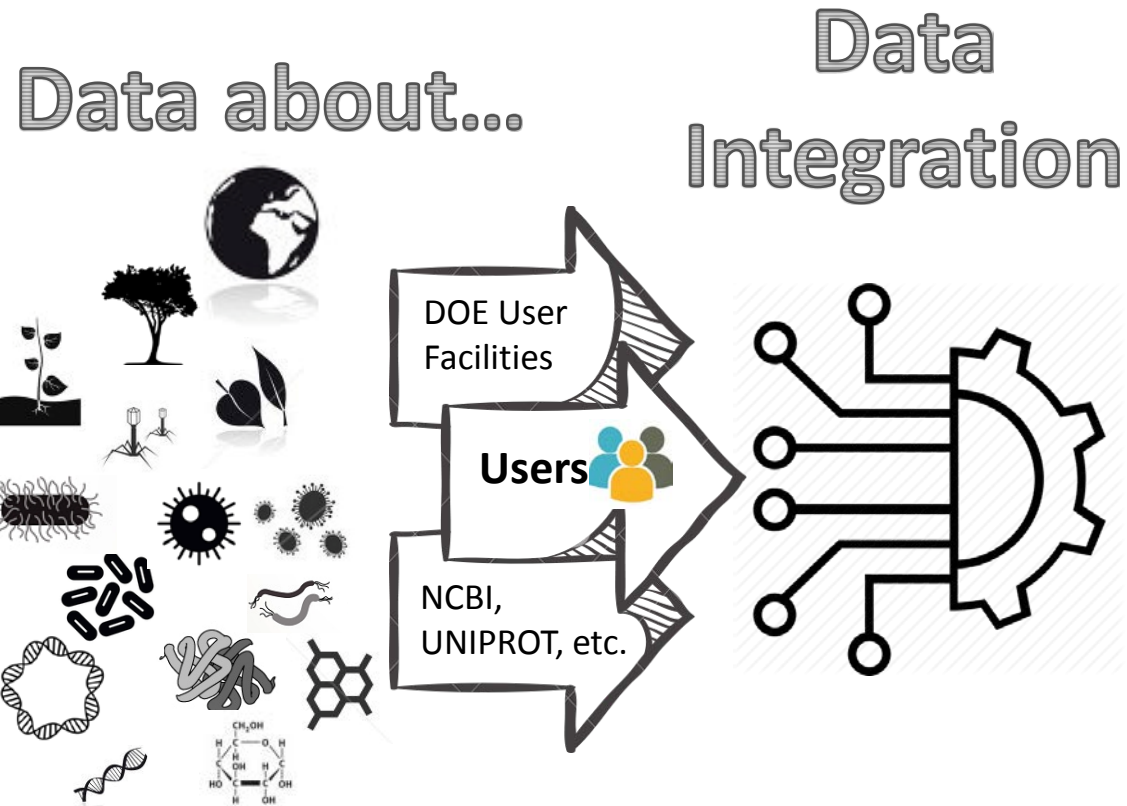


Data  
Integration



# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.



❖ KBase is designed to accelerate research through:

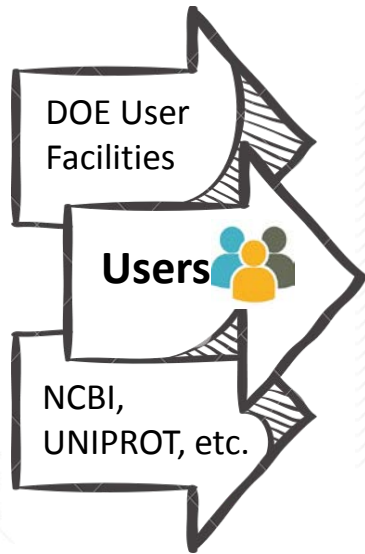
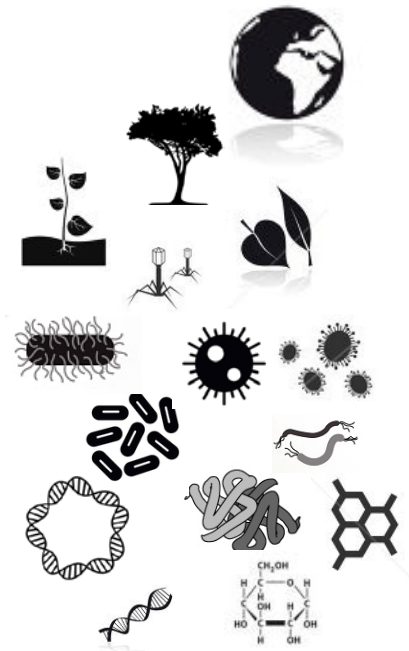
- ❖ Lowering the barrier to integration of diverse data from multiple sources
- ❖ Sharing data and analyses in a “persistent”, transparent, reusable, computable, and reproducible format.



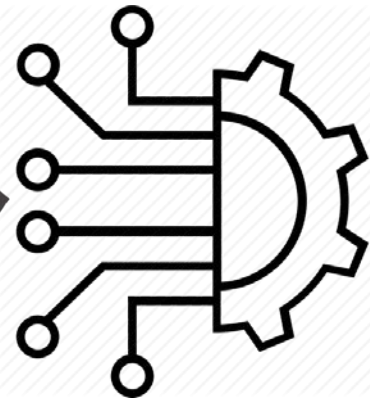
# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

## Data about...



## Data Integration



❖ KBase is designed to accelerate research through:

- ❖ Lowering the barrier to integration of diverse data from multiple sources
- ❖ Sharing data and analyses in a “persistent”, transparent, reusable, computable, and reproducible format.

**Science** AAAS  
Journals unite for reproducibility  
Marcia McNitt  
• See all authors and affiliations  
Science 07 Nov 2014  
Vol. 346, Issue 6201, pp. 679  
DOI: 10.1126/science.12574

**Article** Info & Metrics eLetters PDF

Reproducibility, rigor, transparency, and independent verification are cornerstones of the scientific method. Of course, just because a result is reproducible does not necessarily make it right, and just because it is not reproducible does not necessarily make it wrong. A transparent and rigorous approach, however, can almost always shine a light on issues of reproducibility. This light ensures that science moves forward, through independent verifications as well as the course corrections that come from refutations and the objective examination of the resulting data.

It was with the goal of strengthening such approaches in the biomedical sciences that a group of editors representing over 30 major journals, representatives from funding agencies, and scientific leaders assembled at the AAAS headquarters in June of 2014 to discuss principles and guidelines for preclinical biomedical research. The gathering was convened by the U.S. National Institutes of Health, Nature,\* and Science.

The discussion ranged from what journals were already doing to address reproducibility and the

...scientific journals are standing together in their conviction that reproducibility and transparency are important...

a nature research journal  
**SCIENTIFIC DATA**

**OPEN** Comment: The FAIR Guiding Principles for scientific data management and stewardship  
Mark D. Wilkinson *et al.*<sup>#</sup>

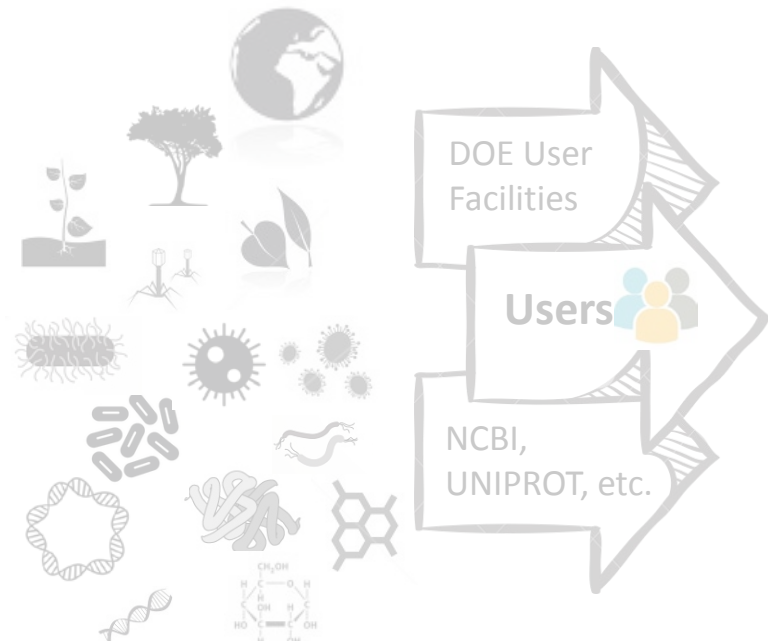
There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Received: 10 December 2015  
Accepted: 12 February 2016  
Published: 15 March 2016

# KBase is a knowledge creation and discovery environment

## Data Integration

Data about...



Data from all your Narratives

Data from collaborators

Public data from KBase's reference collection

Example data

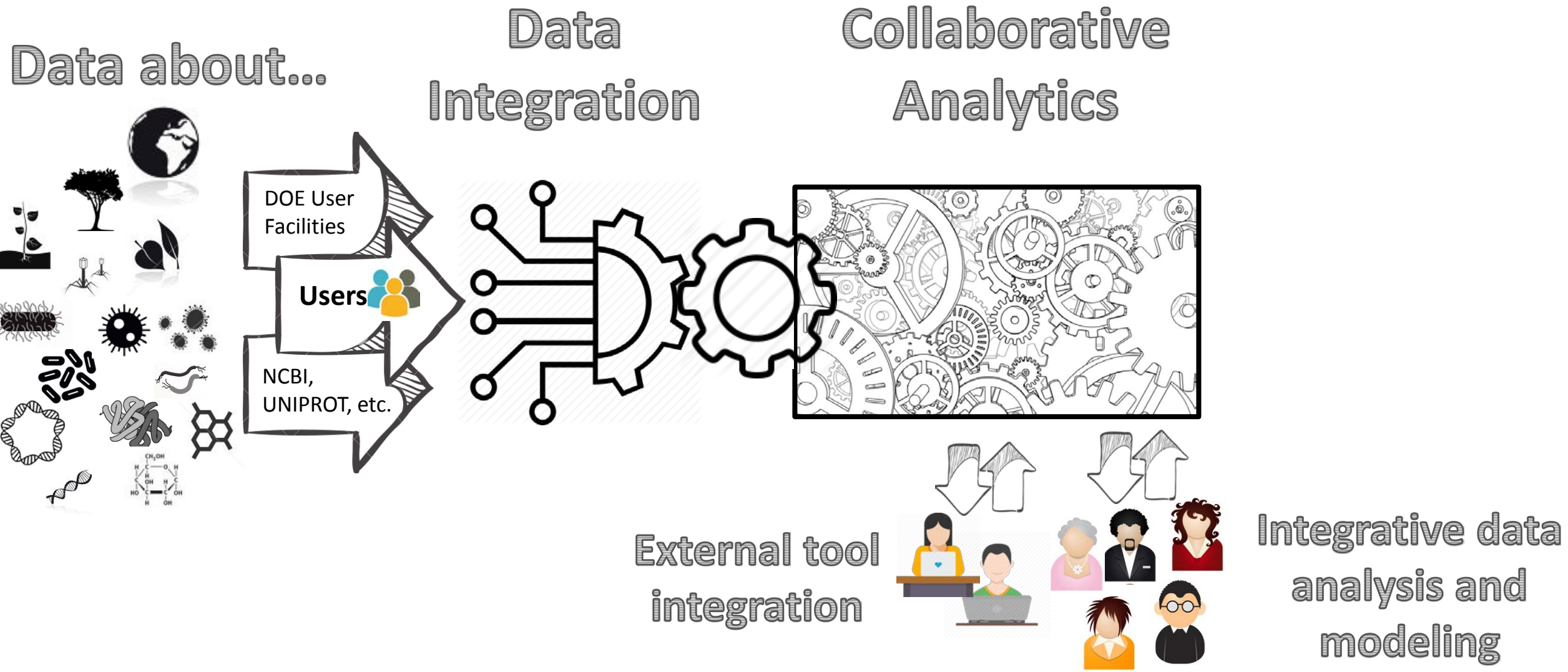
Import your own data

The screenshot shows the KBase interface with the following components:

- Navigation Bar:** Analyze (active), Narratives, Jobs, My Data, Shared With Me, Public, Example, Import.
- DATA Section:** A search bar and filters. A message states "This Narrative has no data yet." with a red "Add Data" button highlighted by a green box. A callout box says: "Click here to access the Data Browser and begin loading data into your Narrative".
- APPS & METHODS Section:** A list of tools including "Assemble and Annotate Microbial Genome v0.1.0", "Build and Normalize Metagenomic Functional Abundance Data v0.1.0", and "Build and Normalize Metagenomic Taxonomic Abundance Data v0.1.0".
- Data List:** A list of data items with icons and titles: "P. trichocarpa JGI3.0 draftModel v1", "PlantHeterotrophicMedia v3", "Populus\_trichocarpa v7", "P. trichocarpa JGI3.0 gapfilledModel v1", "P. trichocarpa JGI3.0 annotate v1", "Escherichia\_coli\_O157\_H7\_str.\_EC4045 v1", and "Clostridium\_thermocellum\_DSM\_4150 v1".
- Currently Uploadable Types:** A green box containing a list of data types: Short reads, Contigs, Genomes, Transcripts, Expression, FBA models, Media, and Phenotype sets.
- Close Button:** A "Close" button at the bottom right of the data list.

# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.



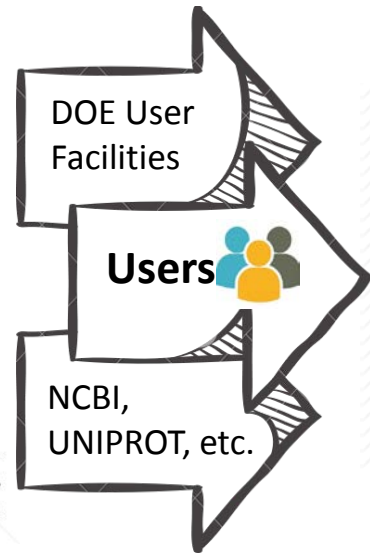
# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

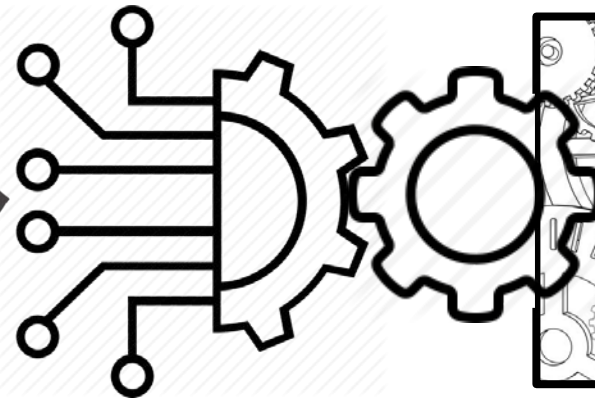
❖ KBase is designed to accelerate research through:

- ❖ Access to a user-extensible powerful suite of tools driving towards modeling biological function.
- ❖ Access to enterprise-level computational resources.
- ❖ A collaborative notebook based system for organizing data, analyses and conclusions.

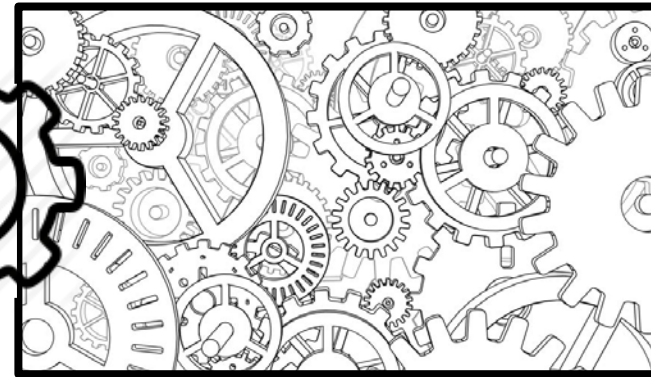
Data about...



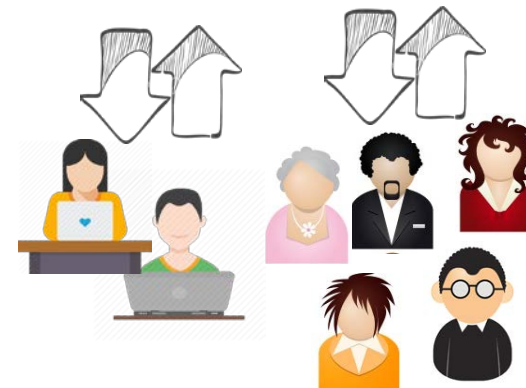
Data  
Integration



Collaborative  
Analytics



External tool  
integration



Integrative data  
analysis and  
modeling

# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

The screenshot shows the KBase Narrative Interface for a project titled "Shewanella Comparative Genomics". The interface is divided into several sections:

- Data:** A list of data objects including "Shewanella\_tree.v1", "Shewanella\_oneidensis\_MRI\_NCBI.v1", "Shewanella\_amazonensis\_SB2B\_NCBI.v1", "Shewanella\_orthologs.v1", and others.
- Apps:** A list of application categories such as Annotation, Assembly, Communities, Comparative Genomics, Expression, Metabolic Modeling, Reads, Sequence, Uncategorized, Upload, and Util.
- Analysis Steps:** A workflow diagram showing the execution of an app, "Insert Genome Into Species Tree", which takes "Shewanella\_unknown" as input and produces "Shewanella\_unknown\_tree" as output.
- Visuals:** A phylogenetic tree visualization showing the relationship between various Shewanella strains, with bootstrap values and branch lengths.
- Sharing:** A share icon in the top right corner.
- Comments:** A comment icon in the top right corner.
- Custom Scripts:** A code editor icon in the bottom right corner.

## The Narrative Interface

An interactive, dynamic, and persistent document created by users that promotes open, reproducible, and collaborative science

ed to  
earch through:  
a user-  
powerful suite  
iving towards  
biological  
enterprise-level  
onal resources.  
ative notebook  
em for  
data, analyses  
usions.

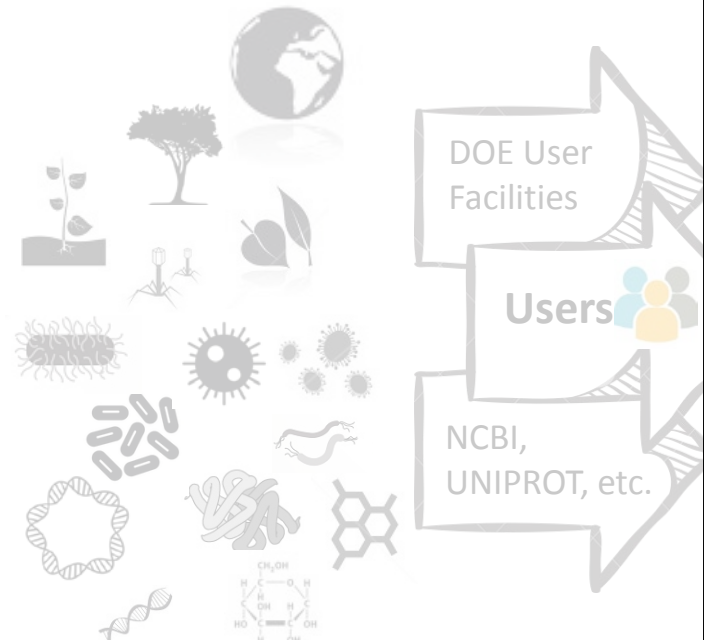


ata  
d

# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

Data about...



**Data**

- GW456A\_trim\_reads\_paired\_SPA
- GW456A\_trim\_reads\_paired\_Velvet
- GW456A\_trim\_reads\_MEGAHIT
- GW456A\_Velvet.contigs\_mj v1
- GW456A\_reads\_trim\_mj\_unpaired
- GW456A\_reads\_trim\_mj\_unpaired
- GW456A\_reads\_trim\_mj\_unpaired
- GW456A\_reads\_trim\_mj\_unpaired
- GW456A\_Prokka\_genome v2

**Apps**

- Build AssemblySet
- Build Feature Set from Genome
- Build GenomeSet

**FastQC Report**

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence QC content
- Per base N content
- Sequence Length Distribution

**Rich Markdown**

```
from biobase.narrative.jobs.appmanager import AppManager
AppManager().run_app(
    {
        "kb_trimmomatic/run_trimmomatic",
        {
            "read_type": "PE",
            "input_reads_ref": "GW456A_reads",
            "output_reads_name": "GW456A_trim_reads",
            "quality_encoding": "phred33",
            "adapter_clip": {
                "adapterPa": "TruSeq-PE.fa",
                "seed_mismatches": 2,
                "palindrome_clip_threshold": 3,
                "simple_clip_threshold": 10
            },
            "sliding_window": {
                "sliding_window_size": 4,
                "sliding_window_min_quality": 15
            },
            "crop_length": 0,
            "head_crop_length": 0,
            "leading_min_quality": 3,
            "trailing_min_quality": 3,
            "min_length": 36
        }
    }
)
```

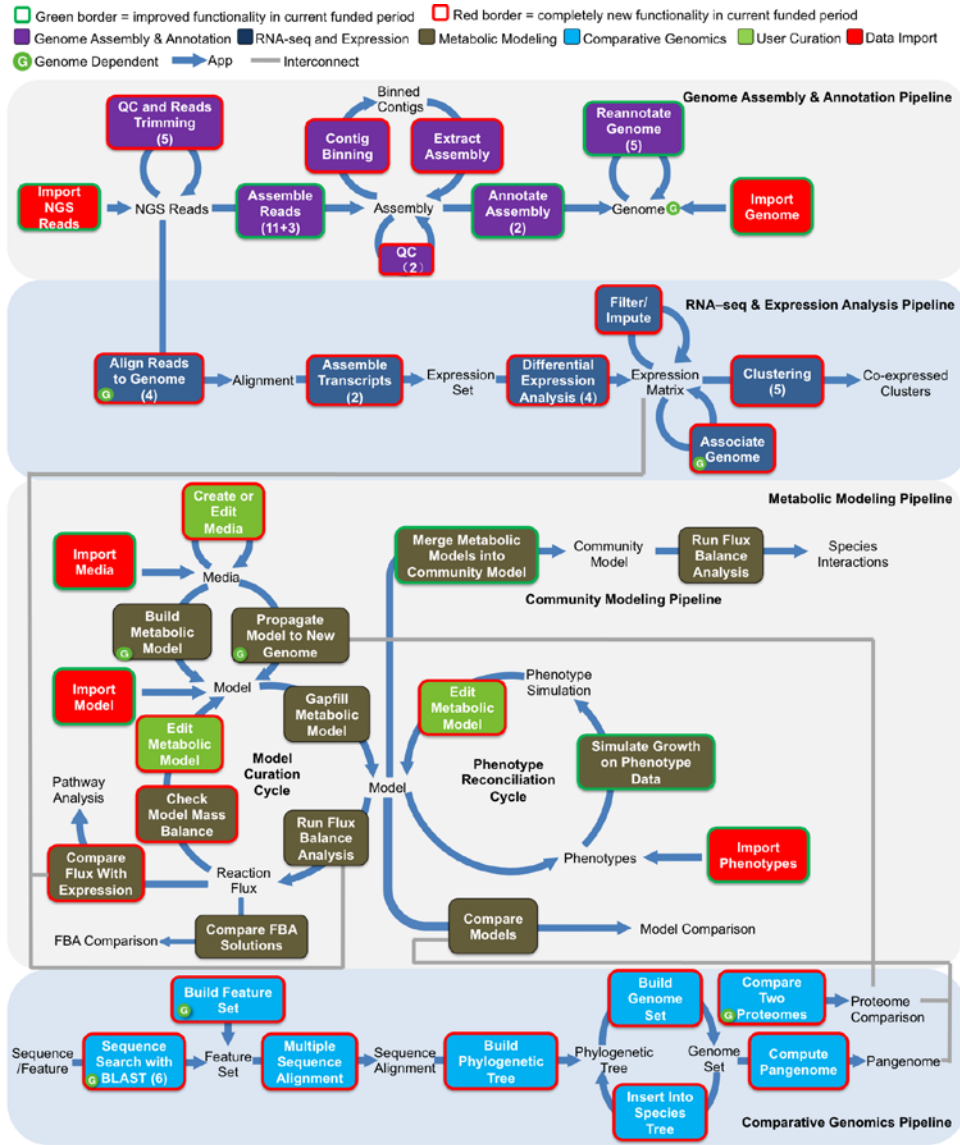
KBase is designed to accelerate research through:

- ❖ Access to a user-extensible powerful suite of tools driving towards modeling biological function.
- ❖ Access to enterprise-level computational resources.
- ❖ A collaborative notebook based system for organizing data, analyses and conclusions.

Integrative data analysis and modeling

# KBase is a knowledge creation and discovery environment

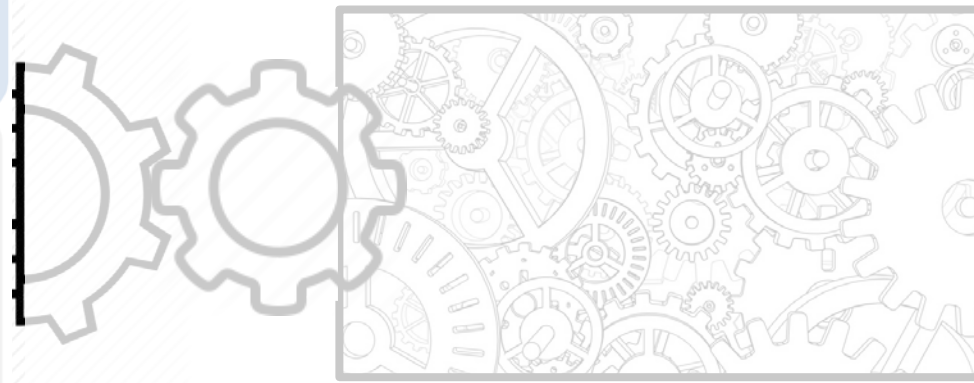
Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
 Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.



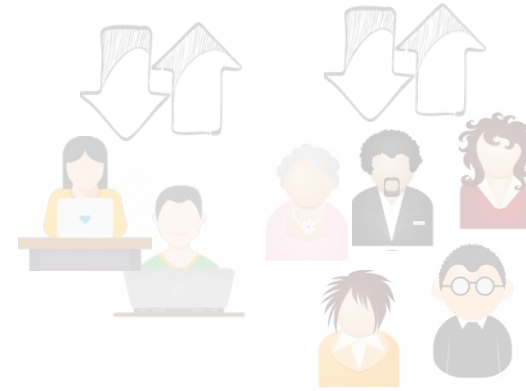
❖ KBase is designed to accelerate research through:

- ❖ Access to a user-extensible powerful suite of tools driving toward modeling biological function.
- ❖ Access to enterprise-level computational resources.
- ❖ A collaborative notebook based system for organizing data, analyses and conclusions.

Collaborative Analytics



External tool integration



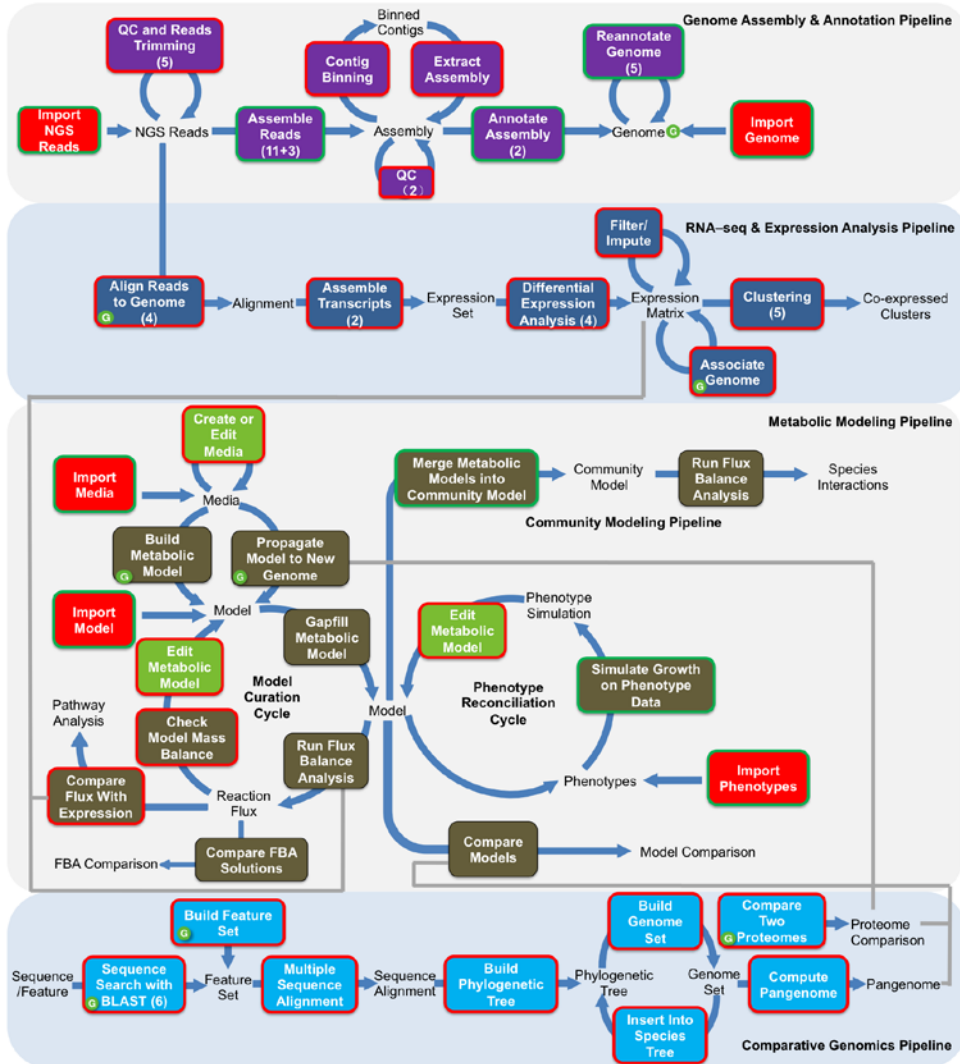
Integrative data analysis and modeling

# KBase is a knowledge creation and discovery environment

Seamless integration of multiply-sourced data and tools in a platform supporting learning, reproducibility and collaboration  
 Powerful scientific framework for predicting function of biomolecules, microorganisms, plants and their communities.

Green border = improved functionality in current funded period  
 Red border = completely new functionality in current funded period  
Genome Assembly & Annotation  
 RNA-seq and Expression  
 Metabolic Modeling  
 Comparative Genomics  
 User Curation  
Data Import  
G Genome Dependent  
 → App  
 — Interconnect

❖ KBase is designed to



o Go from reads to metabolic models of plants, microbes and their community in a matter of a few hours.

o Current tools(That you can add to) include:

- ✓ Bulk upload, download and execution
- ✓ reads management
- ✓ sequence quality assessment and control
- ✓ Comparative microbial assembly and annotation
- ✓ Basic metagenomic binning, assembly and annotation
- ✓ RNA-SEQ analysis for proks and euks
- ✓ Comparative genomics
- ✓ Metabolic modeling for isolates and communities.

External tool  
integration

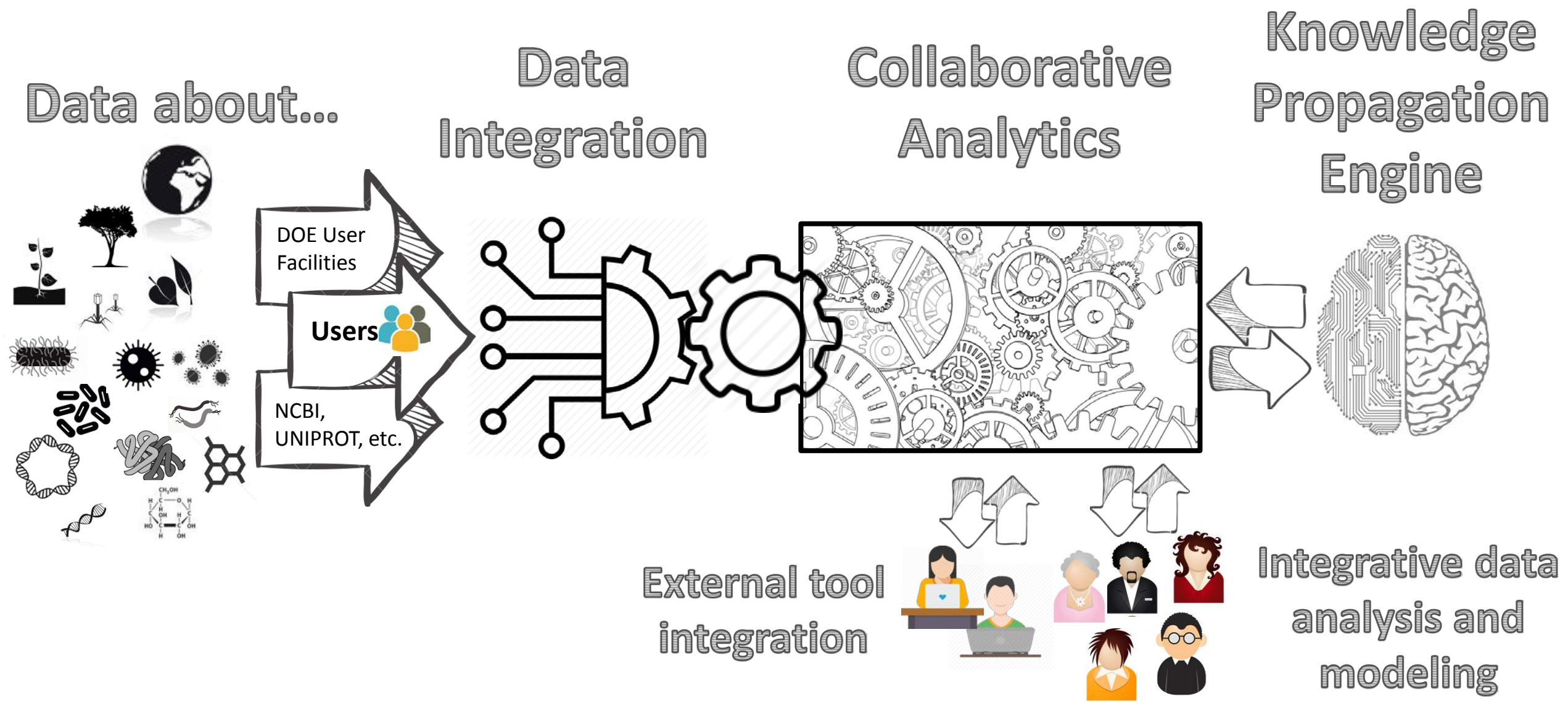


integrative and  
analysis and  
modeling



# KBase is a knowledge creation and discovery environment

- ❖ Integration of primary and derived products into a data model that supports human and machine learning analysis of all shared and published results across the system and automatic propagation of new results to biologically-related entities.



**What is new in KBase since 2016?**

# Platform Improvements since 2016

## Improved Ease of Use

- ✓ Streamlined sign-up and login process
- ✓ Fast and powerful search interface to public and user's private data in KBase
- ✓ Improved search and import of JGI public data
- ✓ Easily import large data and collections of data via Globus or drag-and-drop interface

## Increased Stability and Scalability

- ✓ Narrative interface core built on Jupyter is kept current with upstream
  - ✓ All previous apps converted to the SDK to improve stability
  - ✓ "Docker-ized" the platform to simplify updates, improve reliability, and reduce hardware requirements
  - ✓ Added initial support for HPC and parallel execution
- 
- ✓ Introduced the KBase SDK to dramatically lower the barrier to adding new apps
  - ✓ Used the SDK to develop 100+ new Apps and fill in gaps in reads management, isolate analysis, microbiome analysis, expression analysis, and modeling
  - ✓ Revamped reference data, which now includes all of RefSeq and Phytozome and portions of MycoCosm

# Interface to search reference, public and shared user data

The screenshot displays the KBase Data Search (BETA) interface. At the top, the search bar contains 'pseudomonas'. Below the search bar, there are filters for 'User Data 4,433' and 'Reference Data 4,126'. The search results show 'Found 2,235 Genomes, 2,175 FBA Models, 23 Narratives'. The interface includes a sidebar with navigation options: Dashboard, Catalog, Search (beta), Jobs, and Account. The main content area shows a table of search results. The first result is an FBA Model named 'ISB1139' by Gyorgy Babnigg, modified on 02/23/2018. Below this, there are two detailed views for the model. Each view includes a 'Matches' section with scientific name and taxonomy, and a 'Detail' section with various attributes.

Type	Name	Modified
IAA Pathway Modeling	Gyorgy Babnigg	02/23/2018
FBA Model	ISB1139	02/22/2018

Matches		Detail	
Scientific Name	<i>Pseudomonas</i> fluorescens SBW25	ID	ISB1139
Taxonomy	; Pseudomonadaceae; <i>Pseudomonas</i> ; <i>Pseudomonas</i> fluorescens group; <i>Pseudomonas</i> fluorescens	Name	ISB1139
		Source	External
		Type	SBML Model
		Model Compartments	3
		Model Compounds	1,194
		Model Reactions	1,190
		Genome Reference	24121/138/3
		Scientific name	Pseudomonas fluorescens SBW25
		Taxonomy	cellular organisms > Bacteria > Proteobacteria > Gammaproteobacteria > Pseudomonadales > Pseudomonadaceae > Pseudomonas > Pseudomonas fluorescens group > Pseudomonas fluorescens
		Genome Name	GCF_000009225.2

Matches		Detail	
Scientific Name	<i>Pseudomonas</i> fluorescens SBW25	ID	ISB1139
Taxonomy	; Pseudomonadaceae; <i>Pseudomonas</i> ; <i>Pseudomonas</i> fluorescens group; <i>Pseudomonas</i> fluorescens	Name	ISB1139
		Source	External
		Type	SBML Model

- New high-speed interface to search all public data and data shared by users in KBase
- Currently limited to top-level objects (genomes, models, media, Narratives)
- Will soon expand to lower-level objects (genes, proteins, functions, reactions, compounds)
- Valuable for data/collaboration discovery (e.g., search for a species name)



- Dashboard
- Catalog
- <sup>beta</sup> Search
- Jobs
- Account

Search: JGI KBase - User Data, Reference Data, Features

Q ENIGMA| [x] [↶] [?]

Copy Selected...

View Detail ▾

User Data 14 Reference Data 0

Page 1 of 1

Found 14 Narratives

Access:  Private  Public

Type	Name	Modified
Narrative	<a href="#">ENIGMA Mercury Mystery Plasmids</a>	04/20/2018
	<a href="#">Narrative.1523576735739</a>	04/20/2018
<b>Matches</b> Title <b>ENIGMA</b> Mercury Mystery Plasmids		<b>Detail</b> Title ENIGMA Mercury Mystery Plasmids

Narrative	<a href="#">ENIGMA FW-306 Core-Sediment Campaign</a>	02/21/2018
	<a href="#">Narrative.1519257520913</a>	02/21/2018
<b>Matches</b> Source <b>**This Narrative is part of a series of narratives from the ENIGMA project**</b> ; to see all of them Title <b>ENIGMA</b> FW-306 Core-Sediment Campaign		<b>Detail</b> Title ENIGMA FW-306 Core-Sediment Campaign

Narrative	<a href="#">ENIGMA project master narrative</a>	02/21/2018
	<a href="#">Narrative.1516747247669</a>	02/21/2018
<b>Matches</b> Source <b># ENIGMA Project ## This Narrative links to other ENIGMA narratives in KBase, to help project in the cells below. ### More info about the ENIGMA project can be found on our website: https://enigma.lbl.gov/ ### Most ENIGMA data that can't currently be stored in KBase are stored in our Google</b> Title <b>ENIGMA</b> project master narrative		<b>Detail</b> Title ENIGMA project master narrative

Narrative	<a href="#">ENIGMA metagenomic sequence data</a>	03/09/2018
	<a href="#">Narrative.1513798166246</a>	03/09/2018
<b>Matches</b> Source <b>**This Narrative is part of a series of narratives from the ENIGMA project**</b> ; to see all of them <b>## ENIGMA Metagenomic Sequences This Narrative contains metagenomic sequence data sequenced by</b> Title <b>ENIGMA</b> metagenomic sequence data		<b>Detail</b> Title ENIGMA metagenomic sequence data



## ENIGMA Project

This Narrative links to other ENIGMA narratives in KBase, to help project members find data and analyses from across the project.

Other Narratives are described and linked in the cells below.

More info about the ENIGMA project can be found on our website: <https://enigma.lbl.gov/>

Most ENIGMA data that can't currently be stored in KBase are stored in our Google Drive data folder, here: <https://drive.google.com/drive/folders/0B62rJp3HQTPMbUdjOC1Nd3dkSWs>



### Samples and Wells

This spreadsheet contains a sample request log for samples taken since the 100 Well Survey (see below for those data), along with records of which samples were taken. Look on the 2nd and 3rd tabs of the spreadsheet for sample records and measurements taken on each sample: [https://docs.google.com/spreadsheets/d/1a\\_NC6vVY5Au6geVZj3Mn4eltdqpG4EppftA1uiyWkOg/edit#gid=1606165980](https://docs.google.com/spreadsheets/d/1a_NC6vVY5Au6geVZj3Mn4eltdqpG4EppftA1uiyWkOg/edit#gid=1606165980)

Metadata describing the wells (location, screen depths, etc) are in this folder: <https://drive.google.com/drive/folders/0B18gfpPD5aW8b0prTnUxd2dBa0E>



### 2017 Core Sample Pilot

Data and methods from the 2017 Core Sample Pilot Project are here: <https://narrative.kbase.us/narrative/ws.26612.obj.1>



### 100 Well Survey

Data from the 100 Well Survey and the resulting Smith et al mBio paper are here: <https://narrative.kbase.us/narrative/ws.26835.obj.1>



### Isolates

A narrative with all the sequenced isolates is here: <https://narrative.kbase.us/narrative/ws.24918.obj.1>

Note that you can access these sequenced isolates as "Genome" objects in KBase: just go to "Add Data" and look under "shared with me" to copy any of these into a new narrative so you can explore the data



# Fast Search and Import of JGI Data

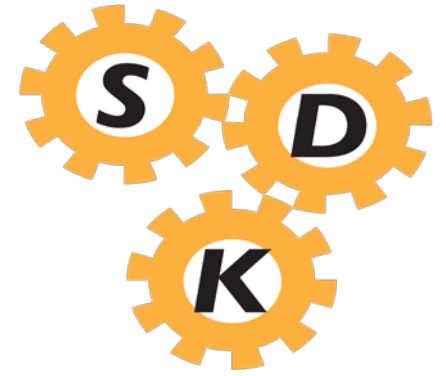
The screenshot displays the KBase JGI Search (BETA) interface. At the top left, the KBase logo and 'JGI Search (BETA)' are visible. On the right, there are icons for 'Feedback', 'CI', and a network diagram. A sidebar on the left contains navigation icons for Dashboard, Catalog, Search (beta), Jobs, Account, and Feeds. The main search area features a search bar with 'poplar' entered, a search button, and a close button. Below the search bar are filter options: 'Type' (Select one or more file types), 'PI' (Filter by PI last), 'Proposal' (Filter by propc), and 'Project' (Filter by seq. p). Navigation controls show '1 to 7 of 3,656' results, 'page 1 of 523', and a 'View Staging Jobs' button. A table of search results is shown with columns: Title, PI, Prop..., Proj..., Date, Scientific Name, T..., S1, S2, Size, In..., C..., and St... The first two rows of results are visible, both for 'Poplar trichocarpa Nisqually-1 Gene A...' by 'Stacey, Gary' with proposal '1097' and project '1074...'. A green overlay box at the bottom contains the following text:

- Co-developed with the JGI team using JAMO
- Currently ports raw reads (FASTQ) and assembled contigs (FASTA) – more datatypes coming
- Currently supports public data – the ability to search private data is coming

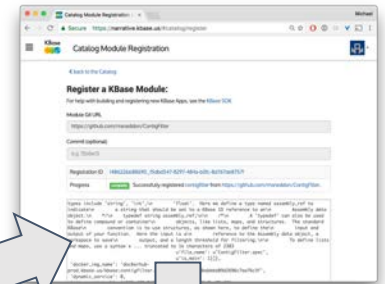
# KBase Software Development Kit

A tool and framework for dynamically adding new Apps to KBase:

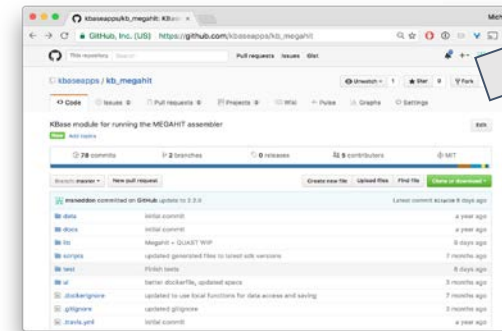
- All scientific tools available to KBase users run as Apps
- Apps can be updated and deployed directly by any developer
- App Catalog tracks and manages all versions of Apps
- Built-in Dev/Beta/Release lifecycle management
- Opens the door for 3rd party developers
- Improved data access and file handling
- Ability to call other KBase modules from your App
- HTML Reports with linked files
- Support for loading versioned Reference Data in Apps
- Web services built with the SDK



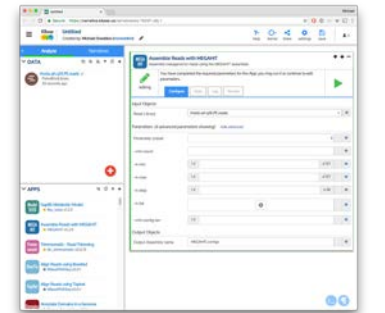
**Register with KBase**



**Push to GitHub**

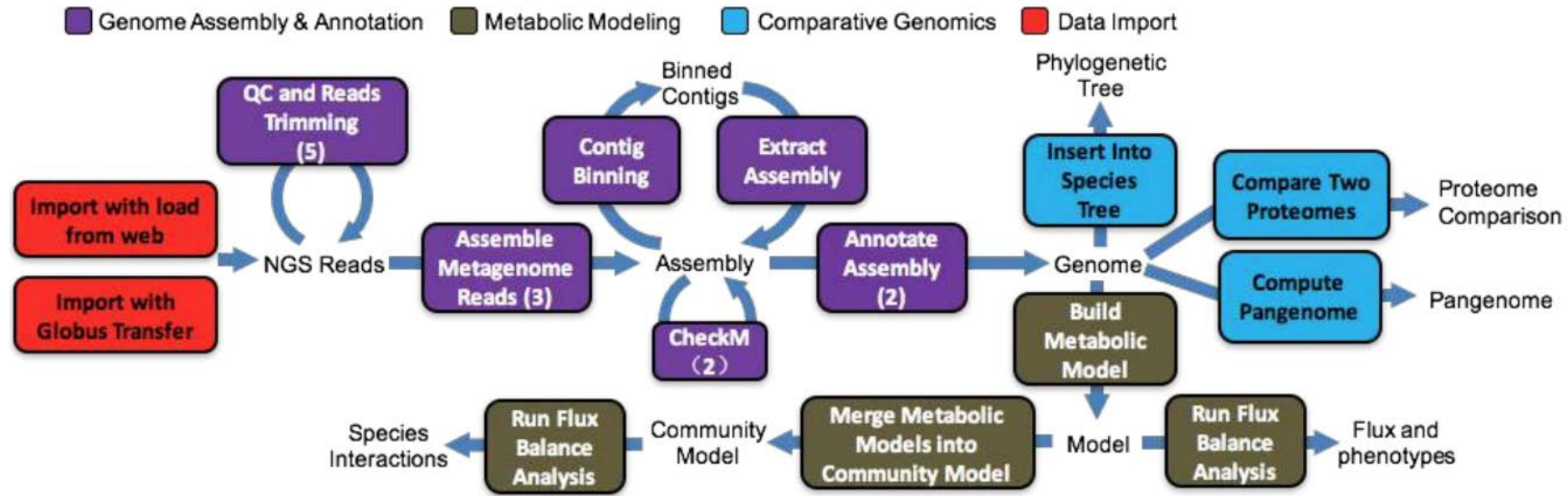


**App is Available!**

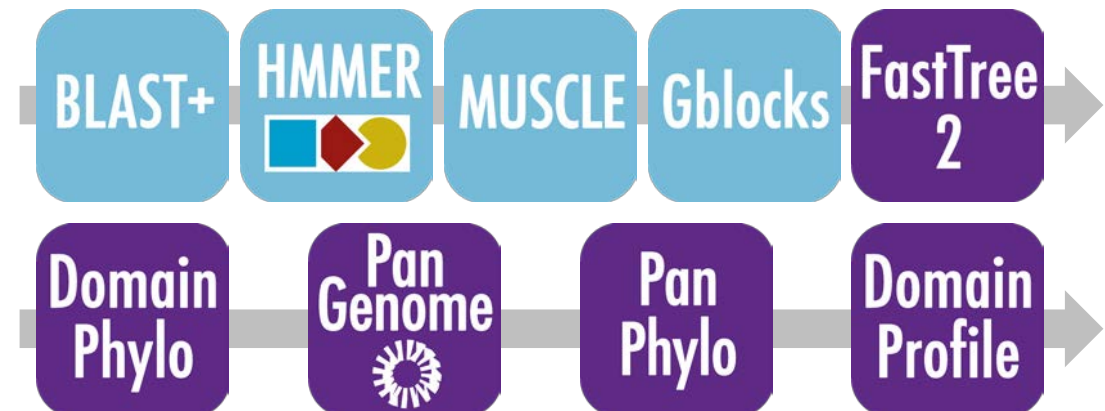




# New workflows in microbiome and comparative genomics



- Assemblers for metagenomics reads
- Binning of contigs into species
- Genome quality assessment
- Multiple algorithms for homology search
- Algorithms for multi-sequence alignment
- Tools for pangenome and phylogenetic analysis



# Better App through 3rd party dev and SDK

## KBase FBA Model Import (Current default - internally developed)

Import TSV/XLS/SBML File as an FBAModel from Staging Area  
Import a file in TSV, XLS (Excel) or SBML format from your staging area into your Narrative as an FBAModel

Reset Finished with success 3m 4s ago View Configure Job Status Result

**Input Objects**

Genome: GCF\_000005845.2

**Parameters**

Model file type: SBML

Model file path (reactions if TSV): ijOI366.xml

Biomass: R\_BIOMASS\_Ec\_ijOI366\_WT\_53p95M

Compounds file path (if uploading TSV):

**Output Objects**

FBA Model object name: ijOI366

## Import Report (current)

**Objects**

Created Object Name	Type	Description
ijOI366	FBAModel	Imported FBAModel

**Summary**

Import Finished  
FBAModel Object Name: ijOI366  
Imported File: ijOI366.xml

3rd party developer constructed a far better version of tool based on 5 years of thesis work on biochemistry integration



## SBML Tools Model Import (3rd party)

Integrate Imported Model into KBase Namespace  
none

Run Configure Job Status Result

**Input Objects**

Model ID: ijOI366

Genome ID: GCF\_000005845.2

**Parameters (6 advanced parameters hidden)** show advanced

Compartment mappings:

Translate Metabolite and Reaction identifiers: KEGG

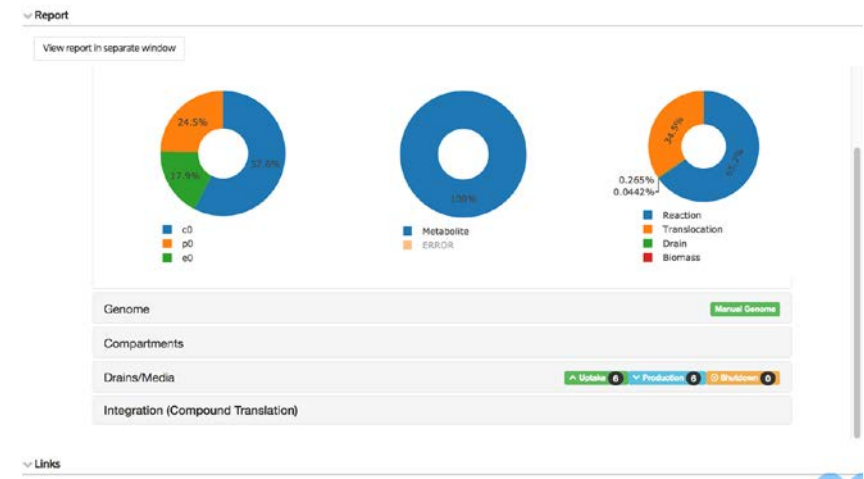
New biomass reactions: Available Items - filtering 2259 of 2259

**Output Objects**

Export Default media:

Output ID:

## Import Report (3rd party)

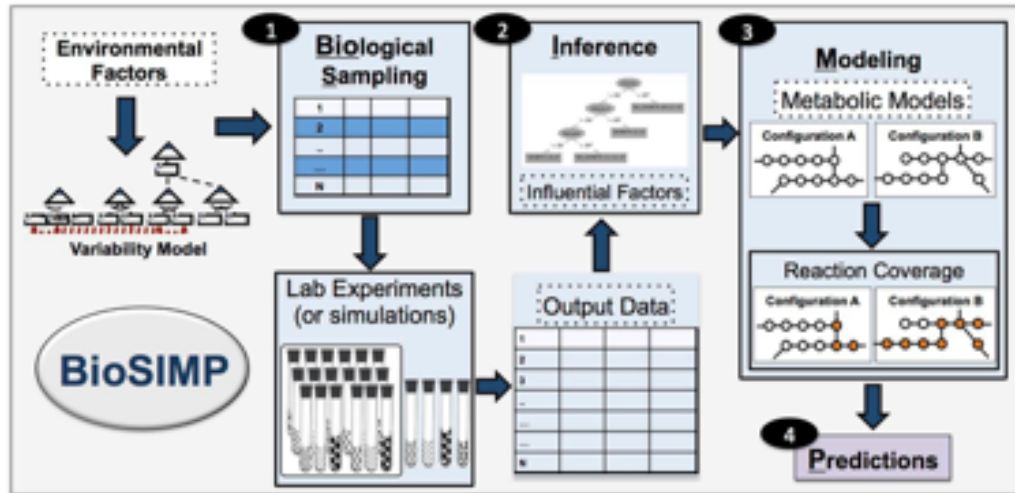


3rd party developer also generated far better visualizations of models



# SDK enables external developers to wrap tools as KBase apps

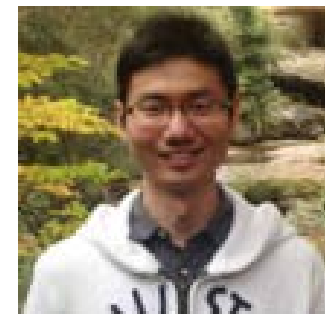
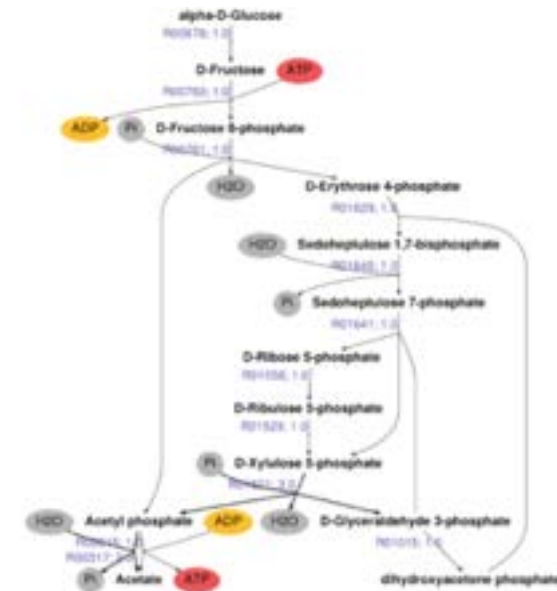
**BioSIMP:** ML-based prediction of growth of new microbes



Mikaela  
Cashman &  
Jennie L. Catlett



**OptStoic:** Predict and design metabolic pathways

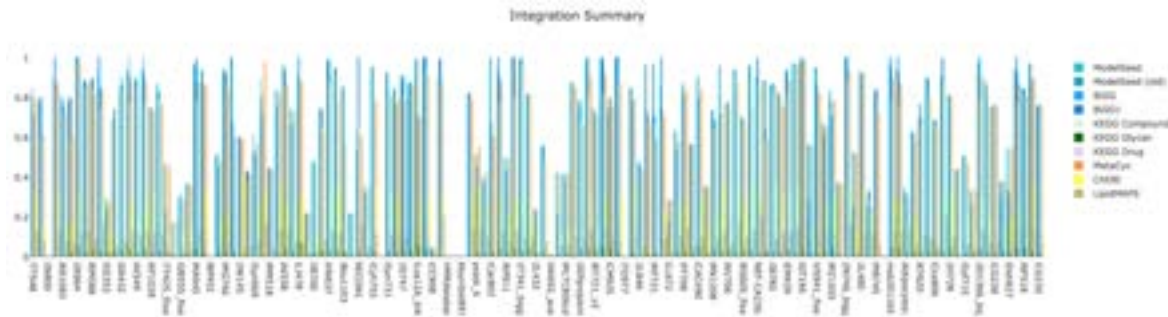


Lin Wang &  
Costas Maranas



# SDK enables external developers to wrap tools as KBase apps

## SBML Importer: Bulk import and integration of metabolic models

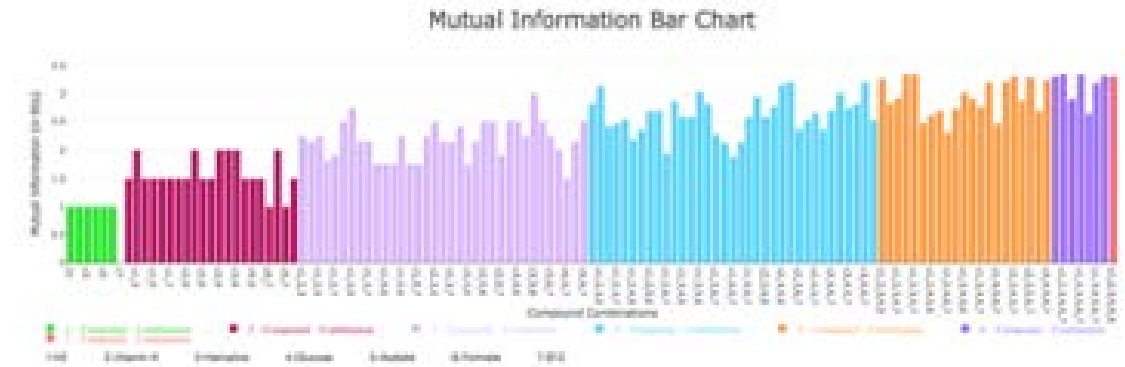


Filipe Liu & Isabel Rocha



- Automatically selects the correct genome to associate with each model by searching the reference genomes for genes in the models
- Enables a curated integration of models with the KBase namespace for compounds, reactions, compartments, and genes
- Applied to import 118 published metabolic models into KBase

## Flux Mutual Information Analysis



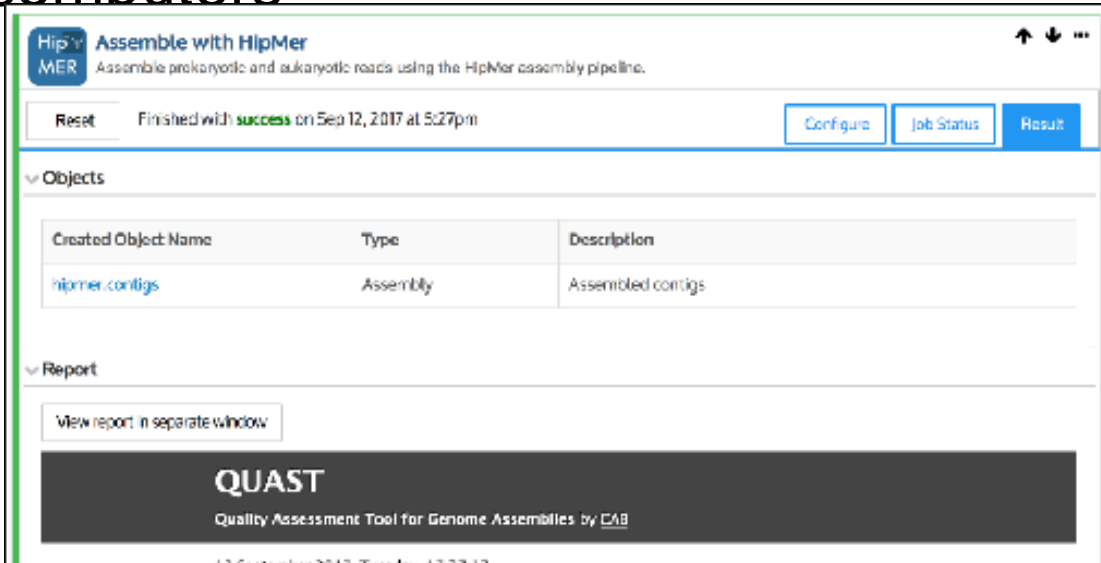
Zahmeeth Sakkaff & Massimiliano Pierobon



- Computes mutual information between reaction flux / cell interactome and input nutrient profile
- Aids experimentalists in identifying the combination of nutrients that reveal the greatest information about the cell by causing variations in pathways, by-products, and growth rate

# Facilities are Using SDK to Add Tools

HipMer: assembly of large genomes and metagenomics on enterprise-class computers



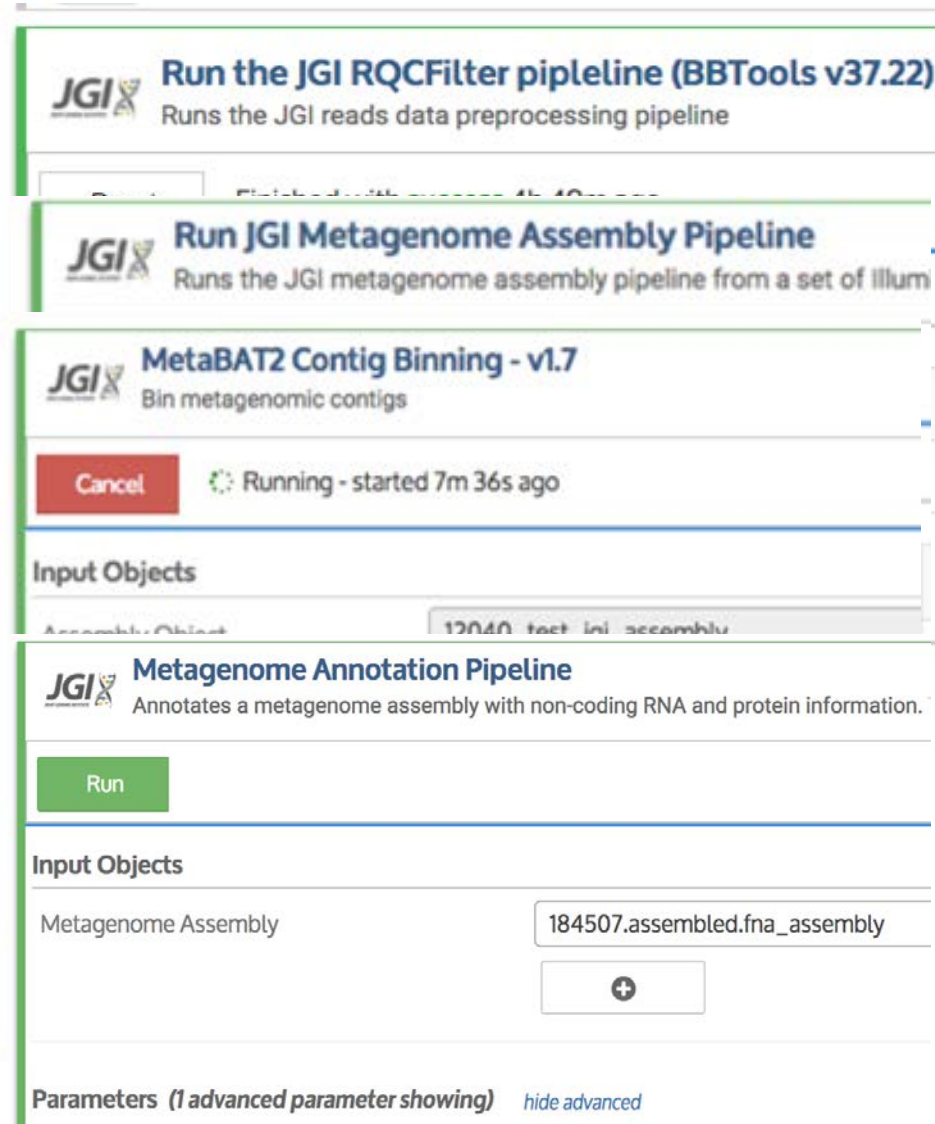
Job: Assemble with HipMer  
Finished with success on Sep 12, 2017 at 5:27pm

Created Object Name	Type	Description
hipmer.contigs	Assembly	Assembled contigs

Report: View report in separate window

**QUAST**  
Quality Assessment Tool for Genome Assemblies by CAB

BBTools: core sequencing quality control from JGI



Run the JGI RQCFilter pipeline (BBTools v37.22)  
Runs the JGI reads data preprocessing pipeline

Run JGI Metagenome Assembly Pipeline  
Runs the JGI metagenome assembly pipeline from a set of illum

MetaBAT2 Contig Binning - v1.7  
Bin metagenomic contigs

Cancel Running - started 7m 36s ago

Input Objects  
Assembly Object 12040 fast ini assembly

Run

Metagenome Annotation Pipeline  
Annotates a metagenome assembly with non-coding RNA and protein information.

Input Objects  
Metagenome Assembly 184507.assembled.fna\_assembly

Parameters (1 advanced parameter showing) hide advanced



**Kathy Yelick**  
CRD, NERSC, LBNL  
EECS, UCB

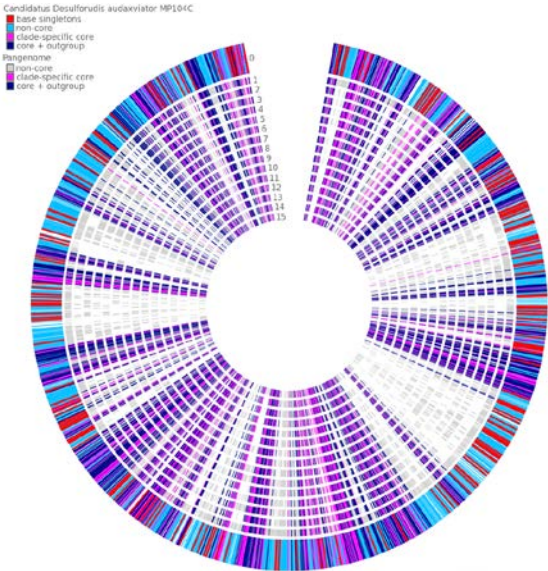


**Dan Rokhsar**  
EGSB, LBL  
MCB, UCB



**Jeff Froula**  
JGI, LBL

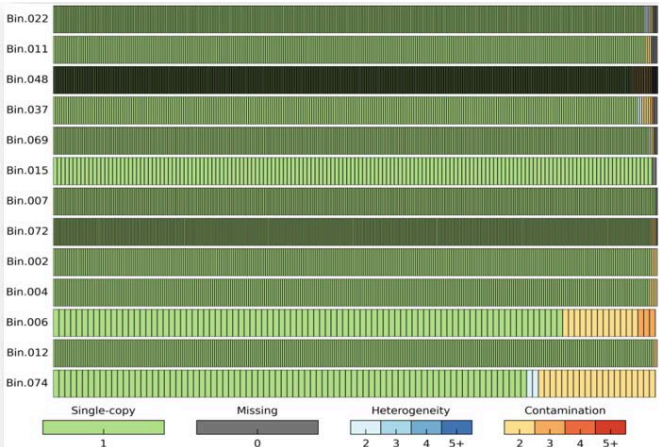
# KBase is providing access to increasingly sophisticated visualization: New this year



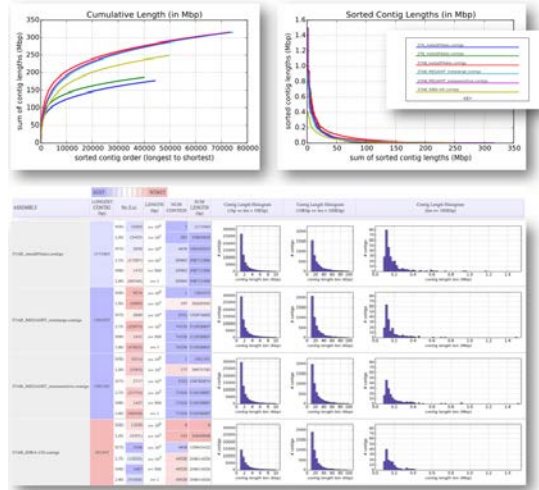
Multi genome alignment



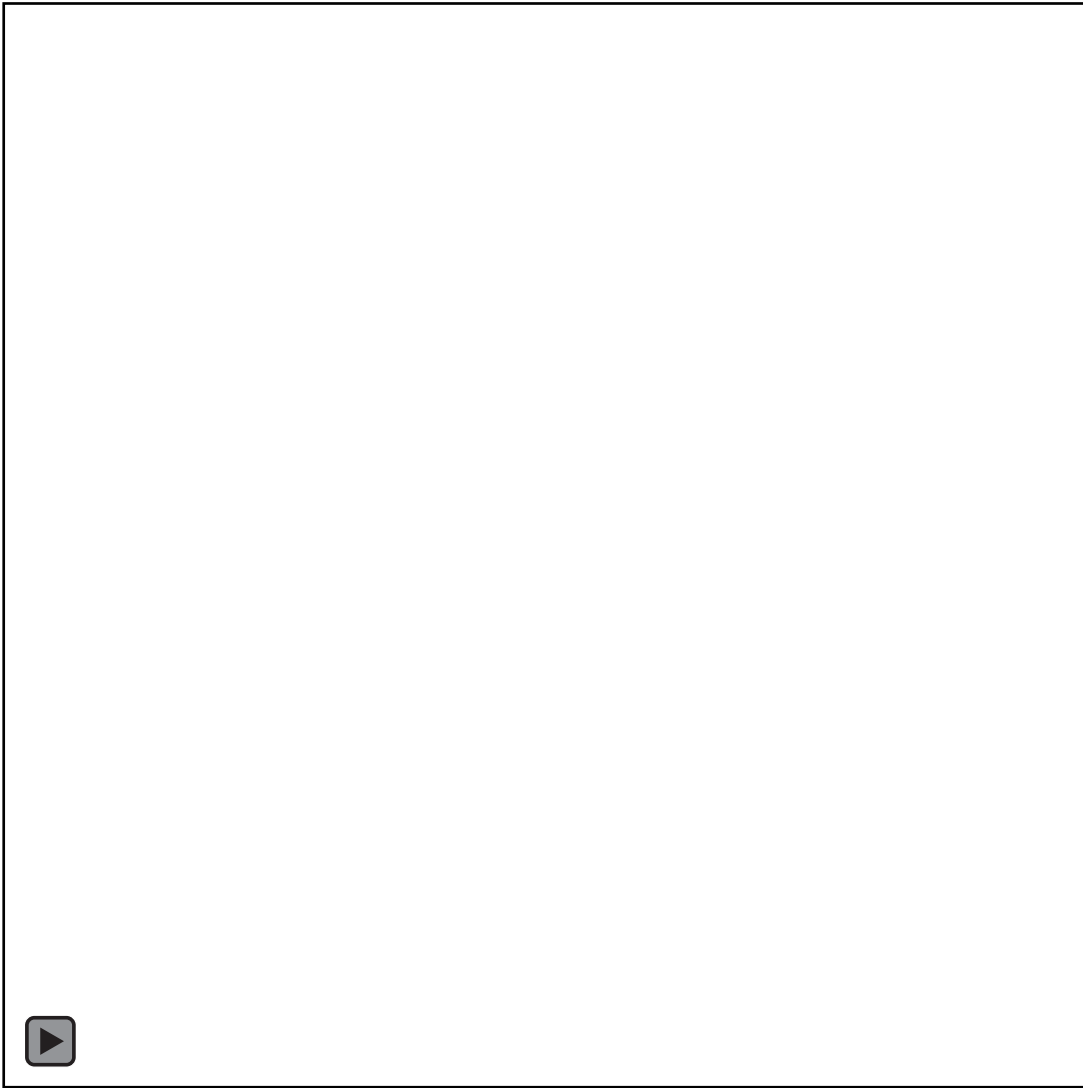
Pangenome Accumulation



QUAST assembly statistics



CheckM single copy gene analysis



Network visualization of metabolic interactions between species in a microbiome

# Growing number of analysis apps (>120 in Production, >55 in Beta, many in development)

The screenshot shows the KBase App Catalog interface. The top navigation bar includes the KBase logo, a search bar, and a user profile. The main content area is titled 'Read Processing' and displays a grid of app cards. A dropdown menu is open over the 'Organize by' filter, listing options: My Favorites, Favorites Count, Run Count, Name (a-z), Name (z-a), Category, Module, Developer, Input Types, and Output Types. The app cards include 'FastQC - Assess Read Quality', 'Trimomatic', 'Load Paired-End Reads From Web - v1.0.7', 'cutadapt - v1.14', 'KB Build Reads From Web', 'KB Build Reads From Web', 'ea-utils Demultiplex with ea-utils FASTQ-MULTX', 'ea-utils Join Overlapping mate Pairs with ea-utils FASTQ-JOIN', 'Load Sim v1.0.7', 'KB Merge Multiple ReadsSets to One ReadsSet - v1.0.1', 'ea-utils Compute Simple Read Library Stats with ea-utils', and 'PRINSEQ PRINSEQ Complex'. The bottom section is titled 'Genome Assembly' and shows 'MEGAHIT Assemble Reads with MEGAHIT v1.1.1', 'CheckM Assess Genome Quality with CheckM - v1.0.8', and 'SPAdes Assemble metagenomic reads using the SPAdes assembler'.

metaSPAdes Assemble with metaSPAdes - v3.11.1  
kb\_SPAdes by gaprice, dylan  
★ 9 ↻ 656 R B i

metaSPAdes Assemble with metaSPAdes - v3.11.1  
kb\_SPAdes v.1.1.1 by gaprice, dylan  
★ 9 R B ↻ 656 ✓ 81.6% ⌚ 5h 58m

Assemble metagenomic reads using the SPAdes assembler.

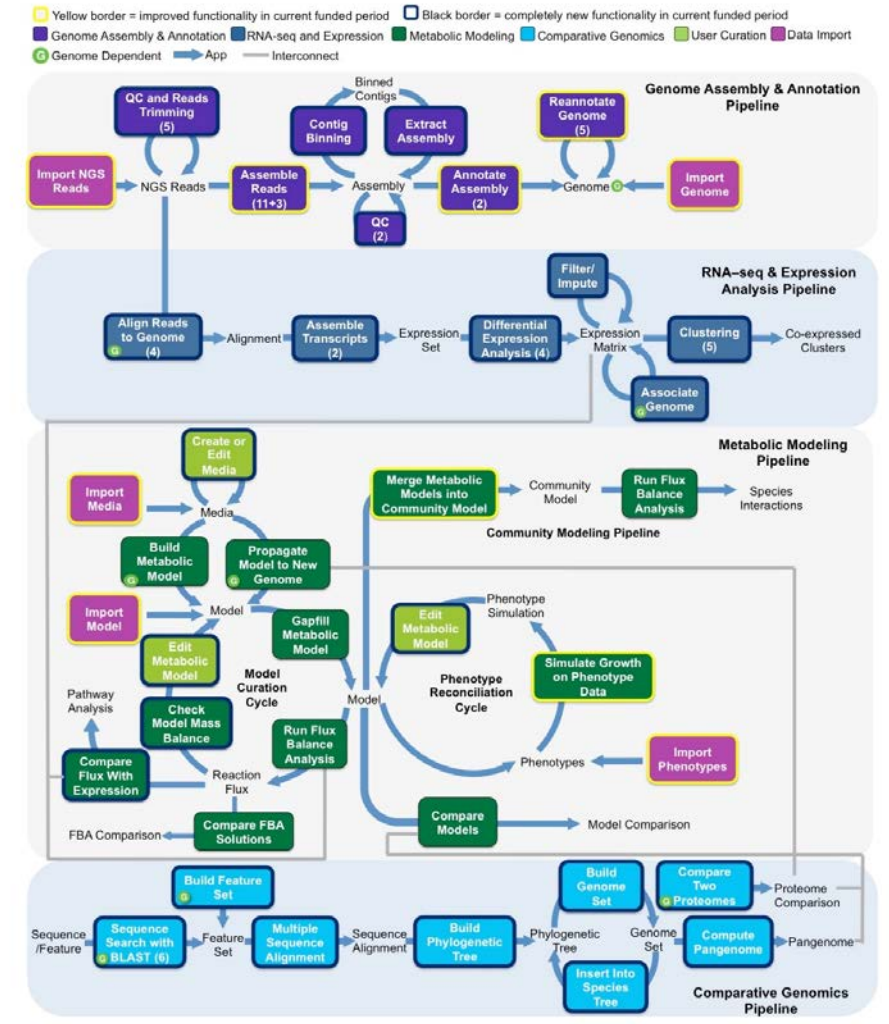
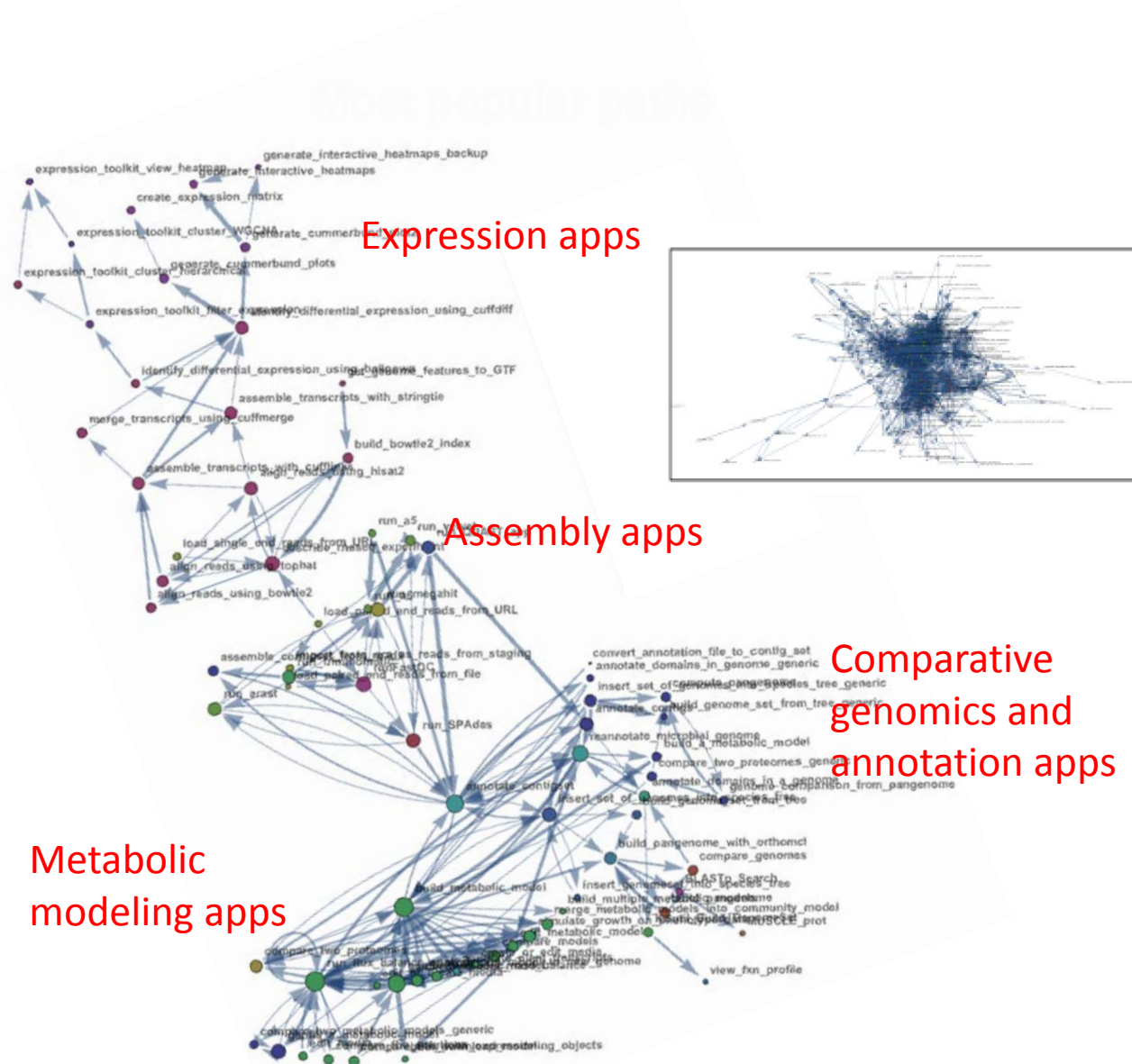
This is a KBase wrapper for the [metaSPAdes](#) genomic reads assembler.

metaSPAdes is designed for assembling shotgun metagenomic reads.

Operational notes:

- Currently the wrapper only supports Illumina, IonTorrent, PacBio CLR and PacBio CCS in FASTQ format, either uncompressed or gzipped.
- The `--careful` flag is always used, except for metagenomic assemblies where it is not allowed.

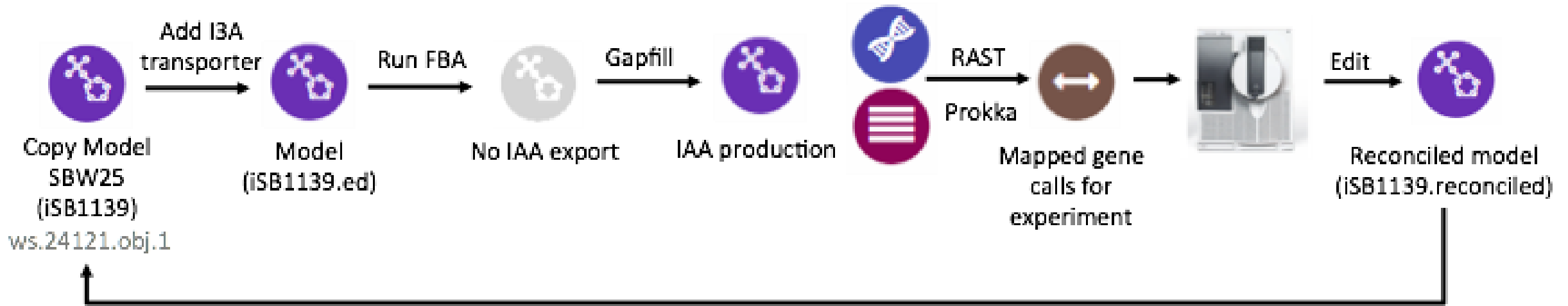
# Users are taking many paths through the system





# Discovery of genes associated with indole-acetate production in *Pseudomonas SBW25*

*Gyorgy Babnig, ANL*

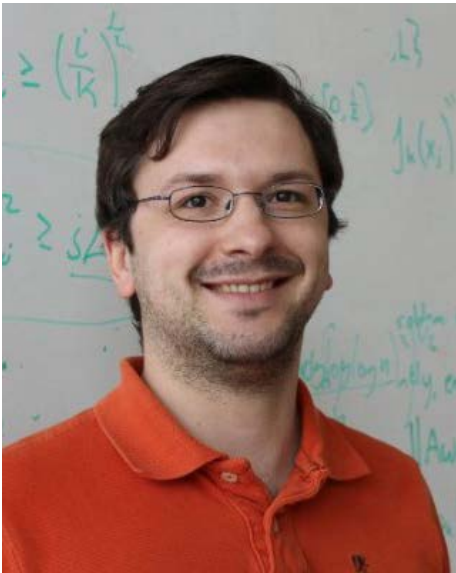


- KBase identified key roles involved in production of indole-acetate (IAA) in SBW25
- KBase enabled easy integration of multiple annotation sources (RAST, Prokka, BIGG) to propose gene candidates for each step of IAA synthesis
- KBase Narrative: <https://narrative.kbase.us/narrative/ws.27990.obj.1>

# Annotating and modeling variation in metabolic function among *Pseudomonas*

Collin Timm, Johns Hopkins University

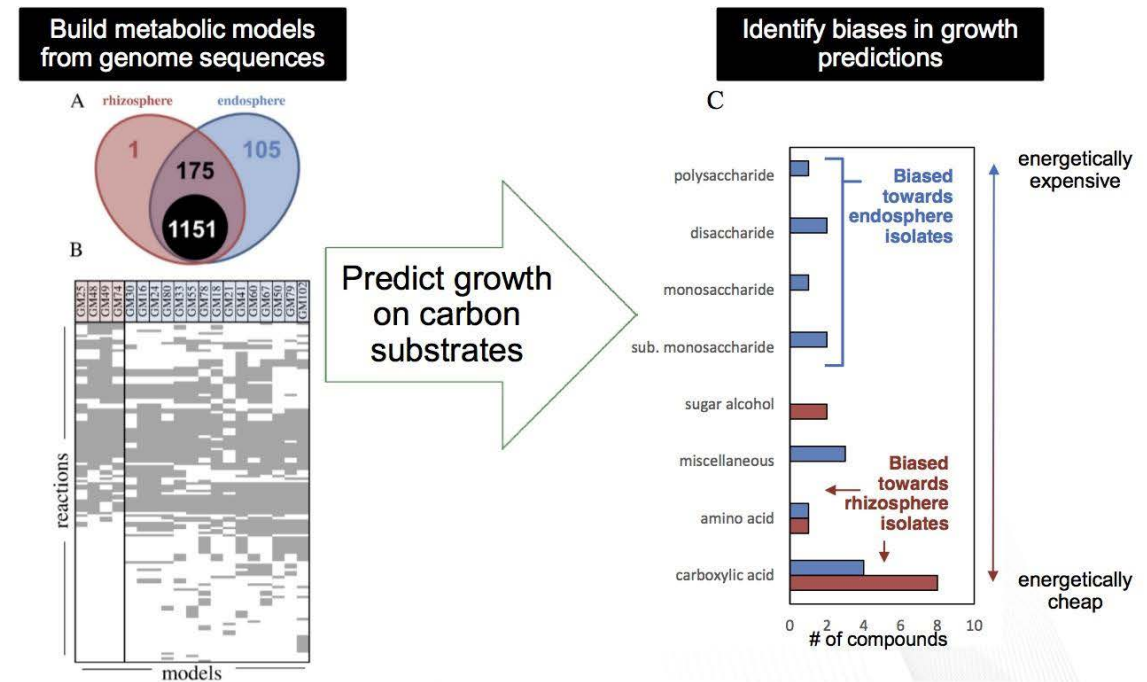
- KBase consistently annotated 19 strains of *Pseudomonas* from rhizosphere
- KBase models predicted how annotation variations lead to phenotype variations



“KBase let me focus on the science, not technical difficulties with implementation of tools”

“KBase enables rapid generation and comparison of metabolic models for genome-sequenced bacteria.”

## Metabolic models predict substrate bias based on isolation environment



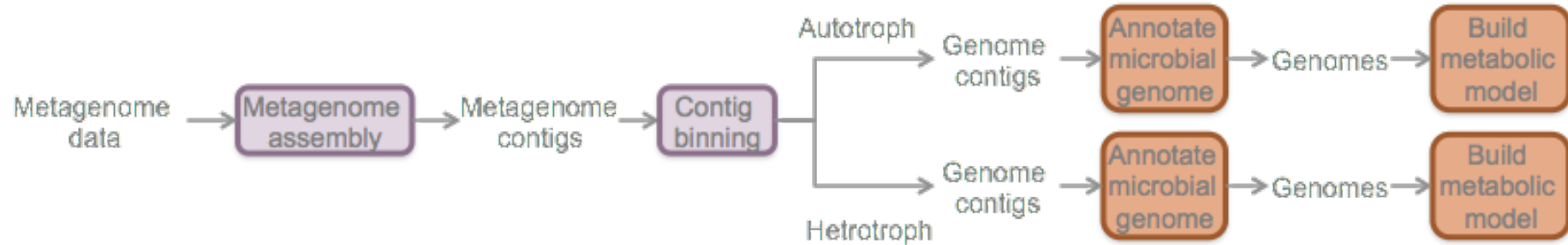
Model comparison summary and predictions for *Pseudomonas* isolates. A. Number of reactions in models. B. Reaction distribution in endosphere and rhizosphere isolates C. Compound groups bias

\*Timm *et al*, 2015, *Frontiers in Microbiology*

# Predicting metabolic interactions between an autotroph and heterotroph

Hyun-Seob Song, PNNL

- Unique integration of assembly, annotation, transcriptomics, and modeling all in one platform (KBase)



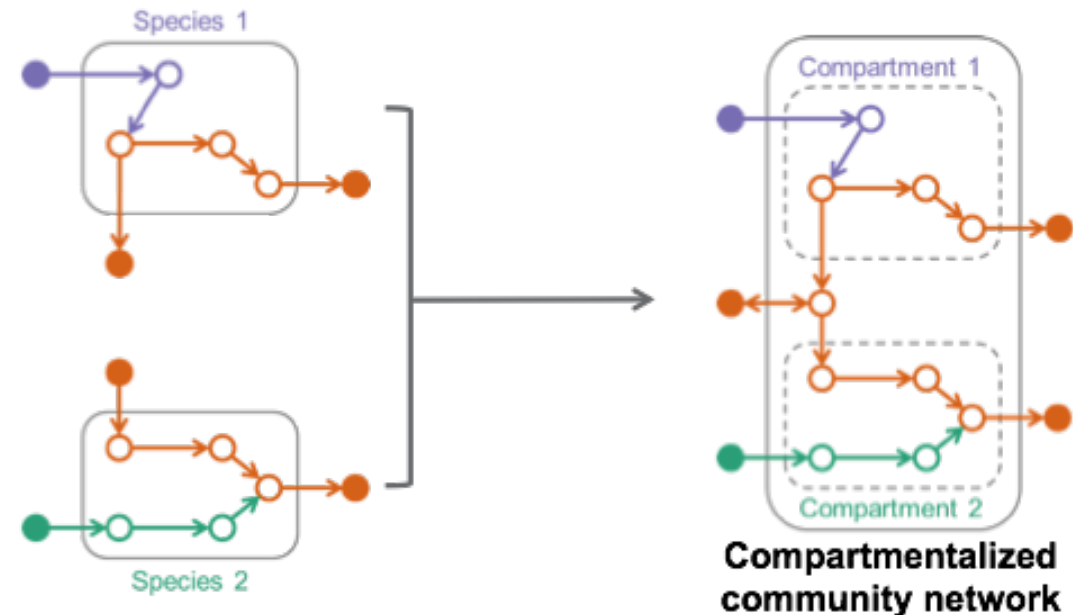
- Predicted metabolite exchanged between species in a simple microbiome
- Explored multiple modeling and gapfilling mechanisms and identified the best alternatives

## Cyanobacterium

Integration of a curated model/genome of a closely related species (*Synechocystis* sp. PCC 6803) (Nogales et al., 2012)

## Heterotroph

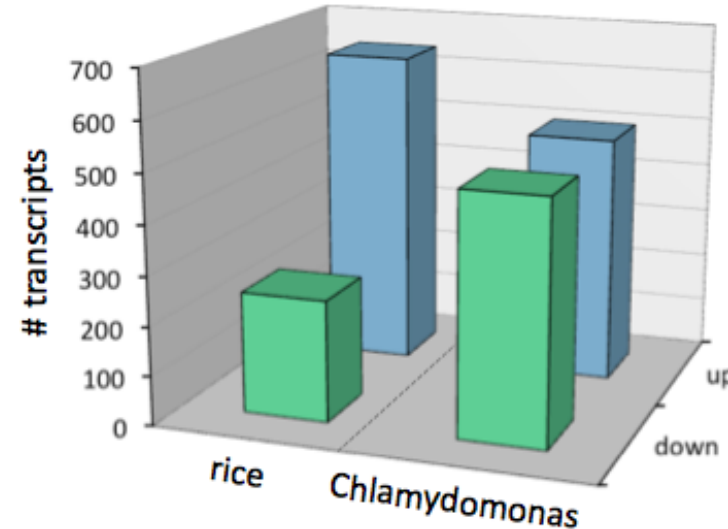
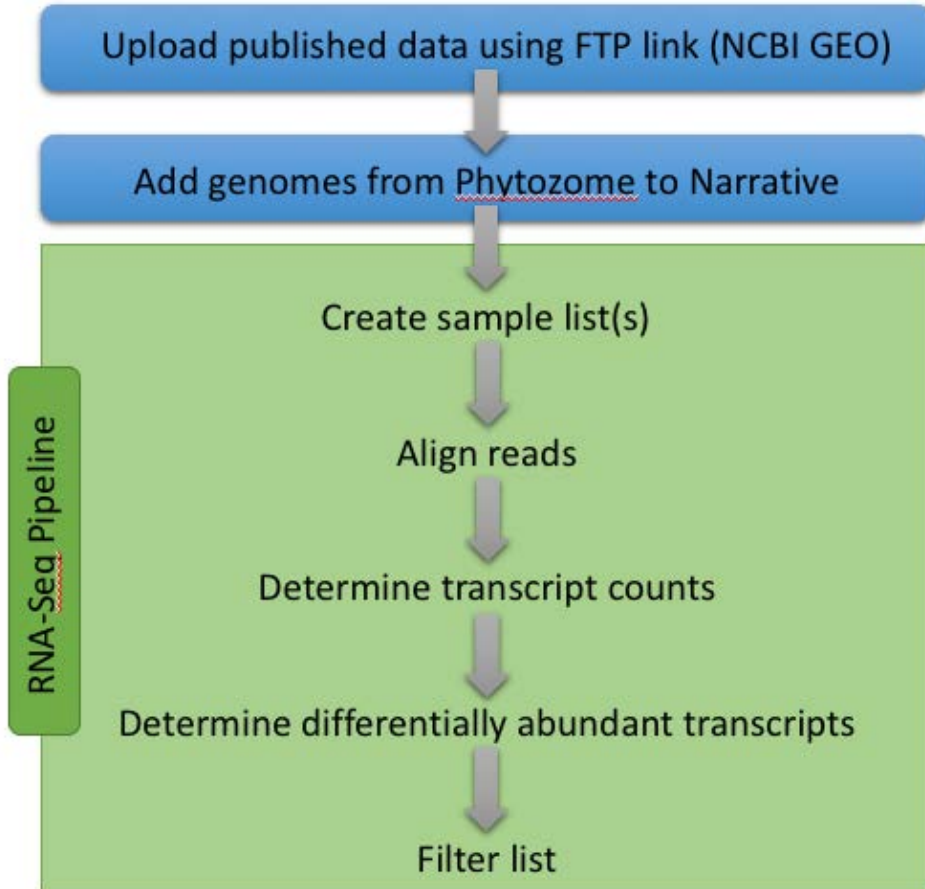
The ModelSEED algorithm



# Uncovering differential response to Zn levels in plants

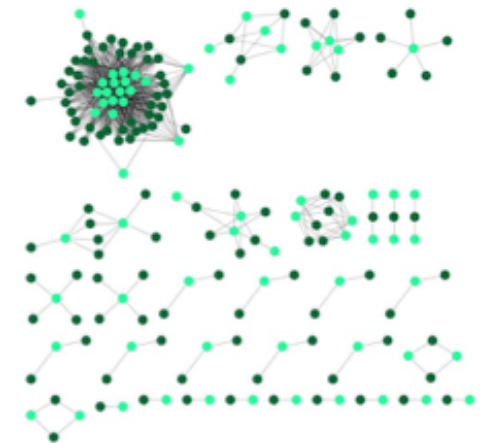
Crysten E. Blaby-Haas, BNL

- Identifying differentially expressed genes in *Chlamydomonas* in presence of perturbed Zn levels
- Comparing with differential expression in similar conditions in other plants
- Identified clusters of genes with similar expression profiles and mapped clusters to functions
- Identified clusters of genes with evolutionarily conserved responses



➤ Evolutionarily conserved responses

Clusters of rice and *Chlamydomonas* homologs



Examples of conserved responses:

Transport:  
ZIP-type Zn transporters  
ABC transporters  
P-type ATPases  
Amino acid transporter  
Potassium transporter  
Sulfate transporter

Zn-dependent enzymes:  
Dehydrogenases  
Carbonic anhydrases  
Hydrolases  
chlorophyll degradation

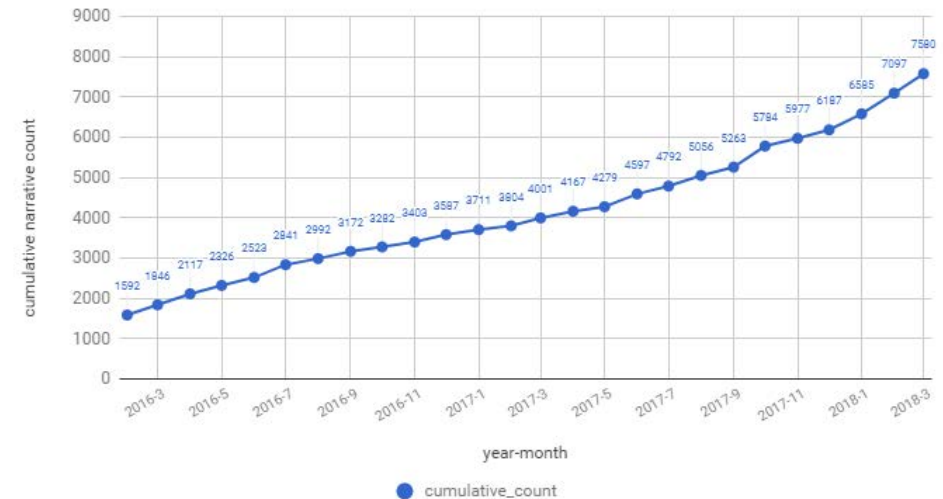
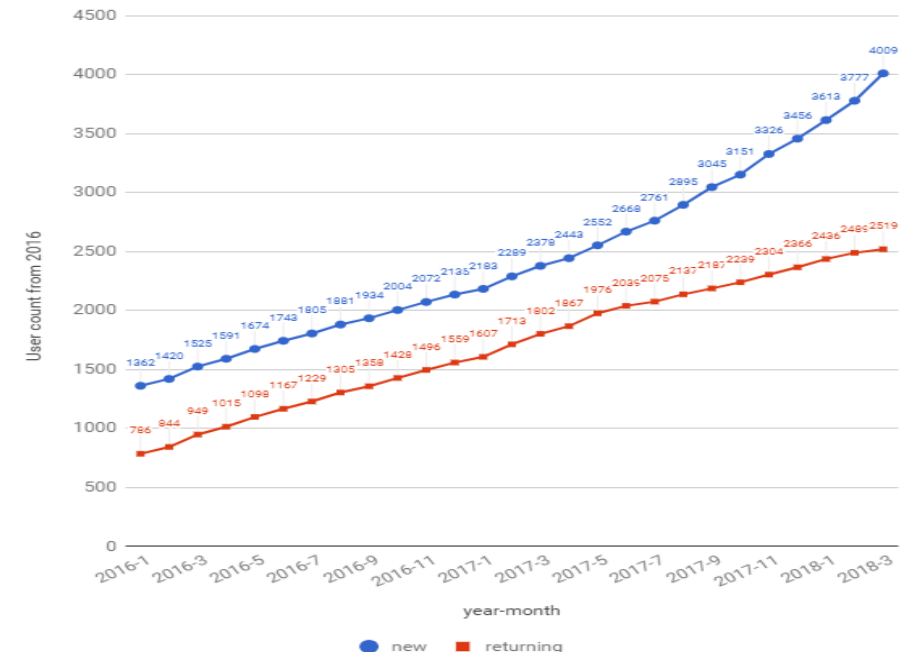
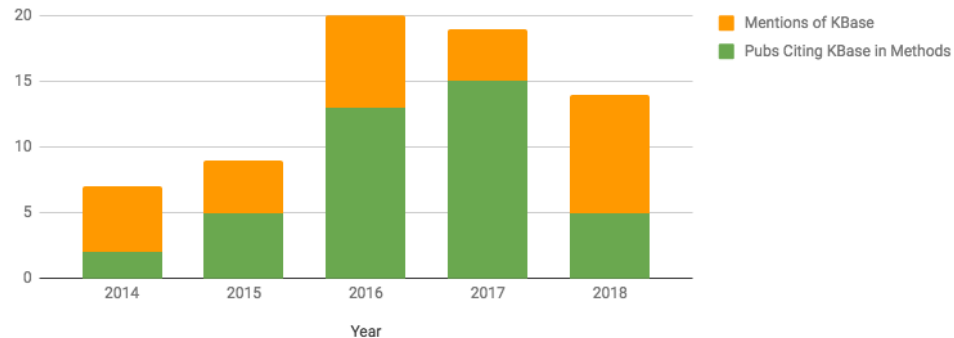
Signaling:  
AP2-TFs  
MYB-TFs  
B-BOX-TFs  
kinases

# Growth in KBase use

More users, more Narratives, more publications

- Total number of users > 4000 (active usage >2500)
- Over 7000 Narratives created by external users
- Monthly unique logins > 150
- 40 Publications citing KBase (2015-present) in their methods

KBase Publications and Mentions per Year (April 6, 2018)



# User Engagement Efforts



## 2017 and 2018 to Date:

- User Engagement Events in 26 locations
- Total attendance of more than 960



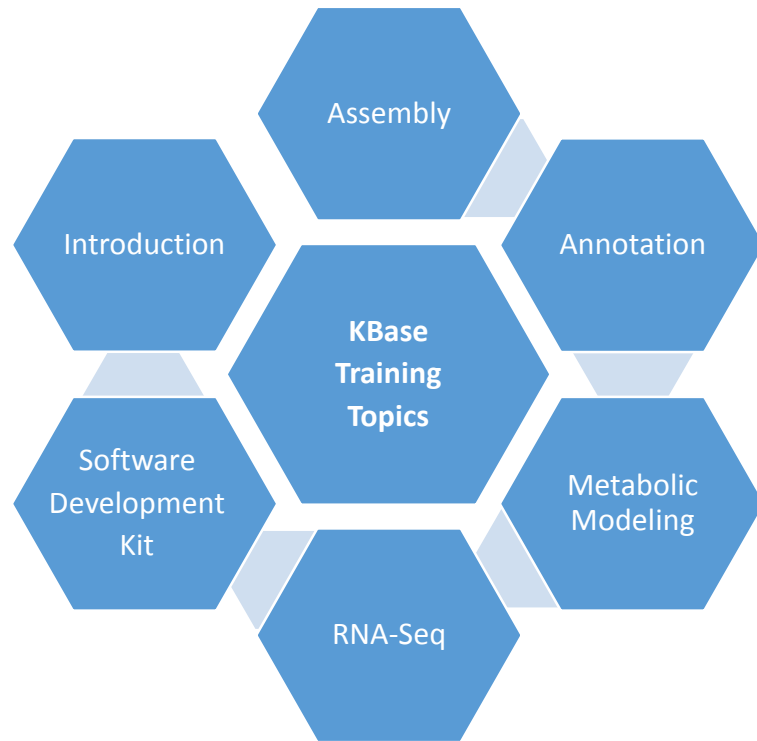
## Future:

- Increased focus on recurrent scheduled webinars and remote events
- Targeted events with SFAs and other key stakeholders



# We can work with you to help your work and teams

- We can provide workshops and webinars at your site
- Learn how to use KBase to help organize your projects data and analyses
- We have support for scientist/developer interchange
- We are starting a **User Working Group** where we help organize strategically important science and development by engaged groups using KBase infrastructure.



Join us: [engage@lists.kbase.us](mailto:engage@lists.kbase.us)

# Transition to New Stage

Base functionality has been introduced into the system to support the following areas:

- Assembly, Annotation, Comparative Genomics, Metabolic Modeling ,Microbiome Analysis, Gene Expression Analysis
- Bulk data management including improved upload from NCBI, FTP, local machines
- Robust search of KBase (and public data at JGI)

System is stable enough that 3rd parties can start bringing in their own data and tools with some assistance from KBase staff

- Ready for co-development with other DOE Facilities
- Future expansion in scientific capabilities will be community-powered
- User Working Groups (UWGs) will be formed to target specific scientific areas of interest
- Goals are to promote cultivation of power users, incorporation of high value data sets, and sharing/access to tools of interest to the scientific community

Project focus will be towards user growth and knowledge engine development



# Collaboration with JGI



## Strategic achievements:

- Established joint JGI/KBase leadership group to coordinate efforts
- Generated a 4-year roadmap with milestones to mark collaboration progress
- Working closely with Dr. Igor Grigoriev on metabolic models of select fungal microbes
- Incorporating reference data from Phytozome
- Working on Metagenome Assembled Genomes with Dr. Emiley Eloë-Fadrosh and IMG
- Initiated co-development activities starting with shared service to compute genome homologies across JGI and KBase public data



## Tactical progress:

Developed a prototype data conduit between JGI and KBase (Pull to KBase)

- Enables users to search JGI databases captured in JAMO via search interface in KBase
- Supported data types can be transferred directly into KBase Narratives

Wrapping JGI-developed analysis tools and pipelines as KBase apps

- BBtools (JGI reads data preprocessing pipeline)
- MetaBat (JGI Metagenome binning tool)
- HipMer (high-performance assembler for paired-end reads)
- Significant progress on metagenome annotation pipelines

The screenshot shows the KBase JGI Search interface. At the top, there's a search bar with '1068248 fasta' entered. Below the search bar is a table with columns: #, Project ID, Title, PI, Date, Type, Scientific name, and Metadata. The table contains two rows, with the second row highlighted in orange. Below the table, there's a 'Sequencing Project' details panel with fields for Title, Name, ID, Status, PI, Year, and Comments.

#	Project ID	Title	PI	Date	Type	Scientific name	Metadata
1	1068248	Bifidobacterium longum subsp. suis DSM 2021	Klenk	8/9/2015	fasta	Bifidobacterium longum	lib:Ar7WVZ
2	1068248	Bifidobacterium longum subsp. suis DSM 2021	Klenk	8/9/2015	fasta	Bifidobacterium longum	lib:Ar7WVZ

**Sequencing Project**

Title	Exploiting the genomes of the Actinobacteria: plant growth promoters and producers of natural products and energy relevant enzymes united in a taxonomically unresolved phylum
Name	Bifidobacterium longum subsp. suis DSM 2021
ID	1068248
Status	In Progress
PI	Klenk, Hans-Peter
As of	2015-06-10
Year	2015
Comments	




# Collaboration with JGI - MycoCosm

The screenshot displays the MycoCosm website interface. On the left, there is a navigation menu with sections like '1000 Fungal Genomes project', 'Genomic Encyclopedia of Fungi', 'Plant Feedback Health', 'Microbial Bioreactors', 'Plant Pathogenesis', 'Genocidal', 'Microbiomes', 'Lignocellulose Degradation', 'Sweet Fermentation', 'Industrial Organisms', 'Fungal Diversity', 'Submit CBP proposal', 'Announcements', and 'Releases'. The main content area features a search bar, a tree diagram of fungal genomes, and a list of genomes under the 'DATA' section. A 'Build Fungal Model' window is open on the right, showing a 'Report' section with a pie chart titled 'Published Model Integration Statistics'. The pie chart shows the following data:

Species	Percentage
Aspergillus_terreus_NIH2624	17.2%
Saccharomyces_cerevisiae_5288c	5.7%
Mucor_circinelloides_CBS277	5.9%
Scheffersomyces_stipitidis_CBS	7%
Candida_glabrata_ASM254	7%
Komagataella_phaffii_GS115	6.8%
Yarrowia_lipolytica_CLIB122	5.7%
Candida_tropicalis_MYA-3404	8%
Kluyveromyces_lactis_NRR1	5.5%
Eremothecium_gossypii_ATCC_10895	8.1%
Aspergillus_oryzae_RIB40	11.5%
Penicillium_rubens_Wisconsin	5.5%

- KBase users now able to access published MycoCosm KBase reference data
- Over 130 MycoCosm genomes ported into KBase
- New app to construct genome-scale models of fungal genomes

# Joint JGI-KBase Roadmap

Roadmap		FY17	FY18	FY19
	<b>Cross-connectivity between JGI and KBase</b>			
	■ Search and import JGI data within KBase	◎	◎	
	■ JGI tools and pipelines available in KBase		◎	◎
	■ Cross-links and ID maps between JGI Portals and KBase			◎
	<b>Build a diverse, engaged user community</b>			
	■ Joint communication strategy		◎	
	■ Joint engagement strategy (Joint User Calls and User Working Groups)		◎	◎
	<b>Enable scientific discovery</b>			
	■ Co-design and development of compute infrastructure		◎	◎
	■ JGI-KBase metabolic model resource for metagenome, microbial, plant, and fungal communities	◎	◎	◎

# Collaboration with EMSL



## Four categories of activity:

### *Integrating tools with KBase*

- MFAPipeline tool for computing flux predictions from labeling patterns determined from NMR data
- NW-chem for prediction of chemical properties from structure

### *Supporting search and import of data from EMSL into KBase*

- Enabling import of metabolomics/NMR data initially as metabolite/value pairs and chemical formula/value pairs
- Enabling import of proteomics data as protein/value pairs and peptide/value pairs

### *Scientific collaborations*

- EMSL and PNNL are working on some fungal systems and will join the KBase/Mycocosm collaboration to annotate, model, and study their genomes
- Applying NWchem to predict thermodynamic properties for compounds in the KBase biochemistry and integrate this data in metabolic models

### *Codevelopment of infrastructure*

- Developing a service to capture, store, represent, search, and compare experimental conditions, which will be shared and linked to datasets at JGI, EMSL, and KBase

## 3 year roadmap:

### **Year 1:**

- MFAPipeline
- Fungal modeling collaboration

### **Year 2:**

- NWChem
- Experimental conditions service
- NMR data import
- Proteomics data import

### **Year 3:**

- Metabolomics data import
- NWchem prediction of thermodynamic properties

# Scaling Computation with DOE HPC Facilities

KBase will leverage ASCR Compute Resources at NERSC to:

- Quickly deal with high demand or to enable large-scale bulk analysis
- Enable large-scale precomputation and value added analysis of reference data and user data to enable Knowledge Propagation
- Support HPC-enabled applications such as HipMER in HipMCL
- **Support JGI co-development activities around homology, taxonomy, environmental “similarity”, ID-Mapping**



KBase’s Execution Framework leverages NERSC’s Shifter container technology to seamlessly run KBase applications.



# Establishing new User Working Groups (UWGs)

KBase will work with the broader community to expand scientific functionality via User Working Groups - thematic subgroups of BER researchers that will help organize and coordinate the development of data and analyses within and across their programs in defined areas of interest.

Goals of UWG is via KBase:

- Organize a concerted community effort to expand KBase functionality and data resources in a tightly focused research area
- Build a user community in each major research area to spread awareness of KBase capabilities and support new user training
- Obtain organized user feedback on KBase functionality and data in a variety of focus areas
- Work together to design new tools and workflows, and facilitate new scientific collaborations

Three initial kernel UWGs under development based on community input (likely to evolve over time):

- Metabolism: metabolic modeling, metabolomics, cheminformatics, integrating chemistry and biology
- Functional Genomics: RNA-seq, TN-seq, proteomics, GWAS, genome annotation, discovery of gene function
- Microbiome: amplicon analysis, metagenome analysis, predicting species interactions within a microbiome

Metrics of success:

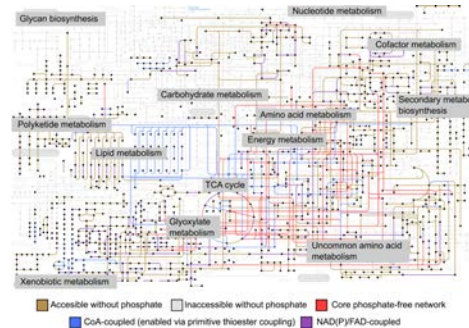
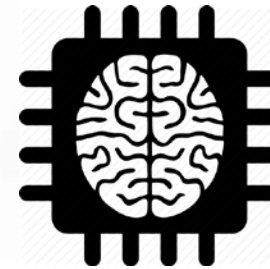
- Scientific publications emerging from interactions within UWG
- Demonstrated use of UWG tools, data, and workflows by the KBase user community
- Presentations at major conferences

# Future Work

# KBase Roadmap Synopsis 2018-2020 (v1)

## Key Areas of Focus


- **Collaboration with DOE Facilities**
  - Major Co-development Efforts with
    - JGI
    - EMSL
- **Platform Development**
  - Search
  - Knowledge Engine
  - Scalable Compute (HPC)
- **Scientific Method Expansion**
  - User Working Groups
    - Metabolism
    - Functional Genomics
    - Microbiome
- **Ongoing Support and Maintenance**
  - Development and Operations
  - General Outreach
- **Management and Publications**





# Projects, Groups and Labs

## Project Page


 **Lignin Characterization and Redesign in the *Miscanthus* Cell Wall**

**Project Description**


Our research focuses on characterizing *Miscanthus*, a biofuel feedstock grass with rapid growth, low mineral content, and high biomass yield. *Miscanthus* can be used as input for ethanol production, often outperforming other alternatives in terms of biomass and gallons of ethanol produced. Broad goals of our research include:

- Improve yield, sustainability, and water and nitrogen use efficiency of *Miscanthus*
- Understand its carbon partitioning and nutrient cycling
- Understand the structure, function, and organization of the *Miscanthus* genome
- Enable *Miscanthus* to be efficiently bred or manipulated for biofuels


**Project Principal Investigators**



Jill Smith



Carlos Rivera



Jane Peters

**Labs**

- [Smith Lab – \*Miscanthus\* genomics](#)
- [Rivera Lab – Cell wall recalcitrance](#)
- [Peters Lab – Breeding and genetic diversity](#)

**Our Narratives**


**Build Plant Metabolic Model**


- Build Plant Metabolic Model
- Compare Two Metabolic Models
- Run Flux Balance Analysis
- View Media

**Sorghum Transcriptome Models Based On**

- Build Plant Metabolic Model
- Compare Two Metabolic Models
- Run Flux Balance Analysis
- View Media
- 26 markdown cells

## Lab Page

 **The Smith Lab**




**Jill Smith, PI**


[smithj@university.edu](mailto:smithj@university.edu)

The Smith lab seeks to better understand and characterize the genome of the *Miscanthus x giganteus*. This perennial grass hybrid of *M. sinensis* and *M. sacchariflorus* has high photosynthetic efficiency and low water use relative to other biofuel feedstock crops. Our lab identifies genes involved in cell wall formation and structure that may be targeted for modification to improve yield while maintaining disease resistance.


**Personnel**




Jason Moore




Sam Anthony



Amy Todd



Allen Richards



Kay Woods

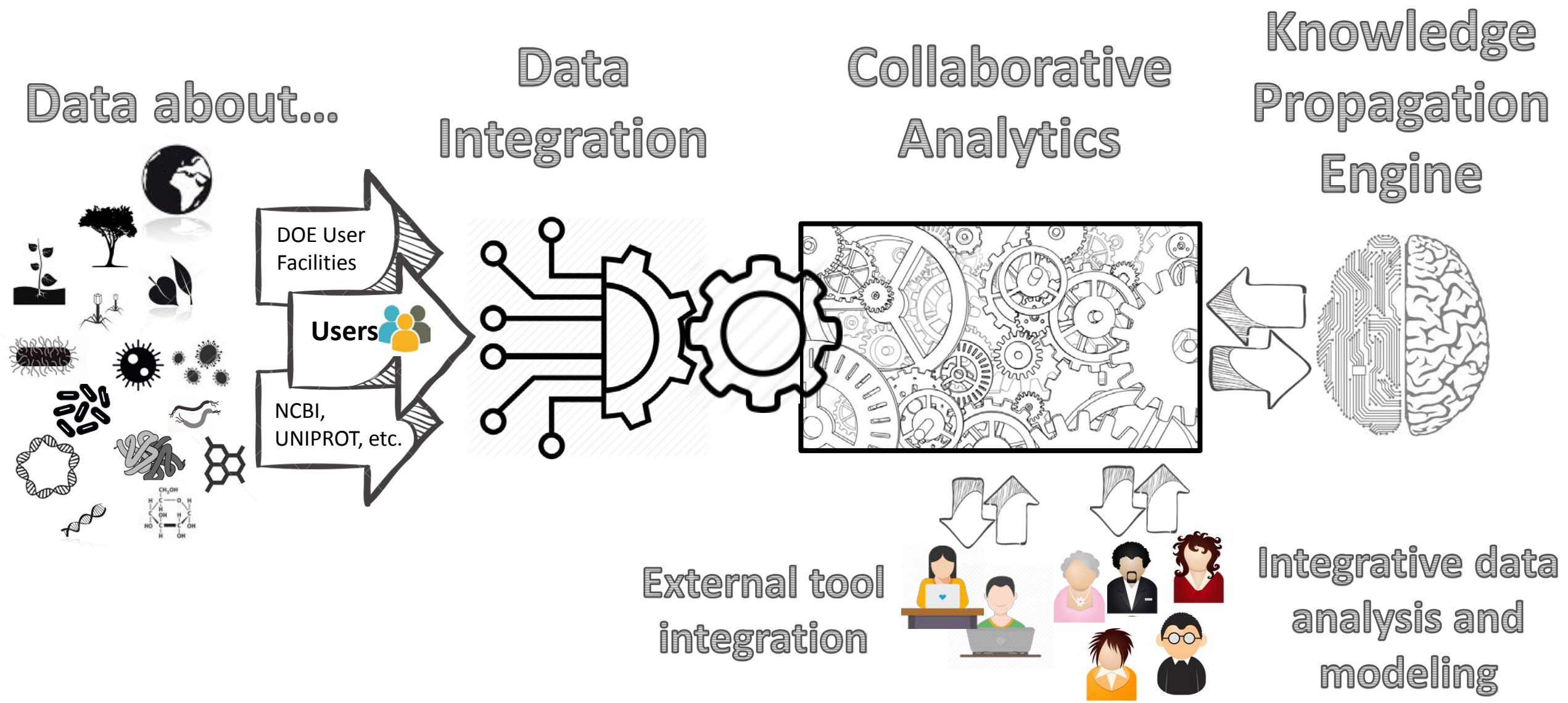
**Projects**

- ▾ **Genes involved in secondary cell wall formation x**  
Identification of genes essential for lignin biosynthesis, transport, and polymerization.
- **Regulation of secondary cell wall formation x**
- **Modification - suppression of gene expression x**
- **Genes supporting phenolics integration into lignin x**
- **Cell wall and disease resistance x**

- Allows projects to organize their people, data and analyses
- Allows them to track metrics of progress and use
- Allows them to publish and track external citation and utility

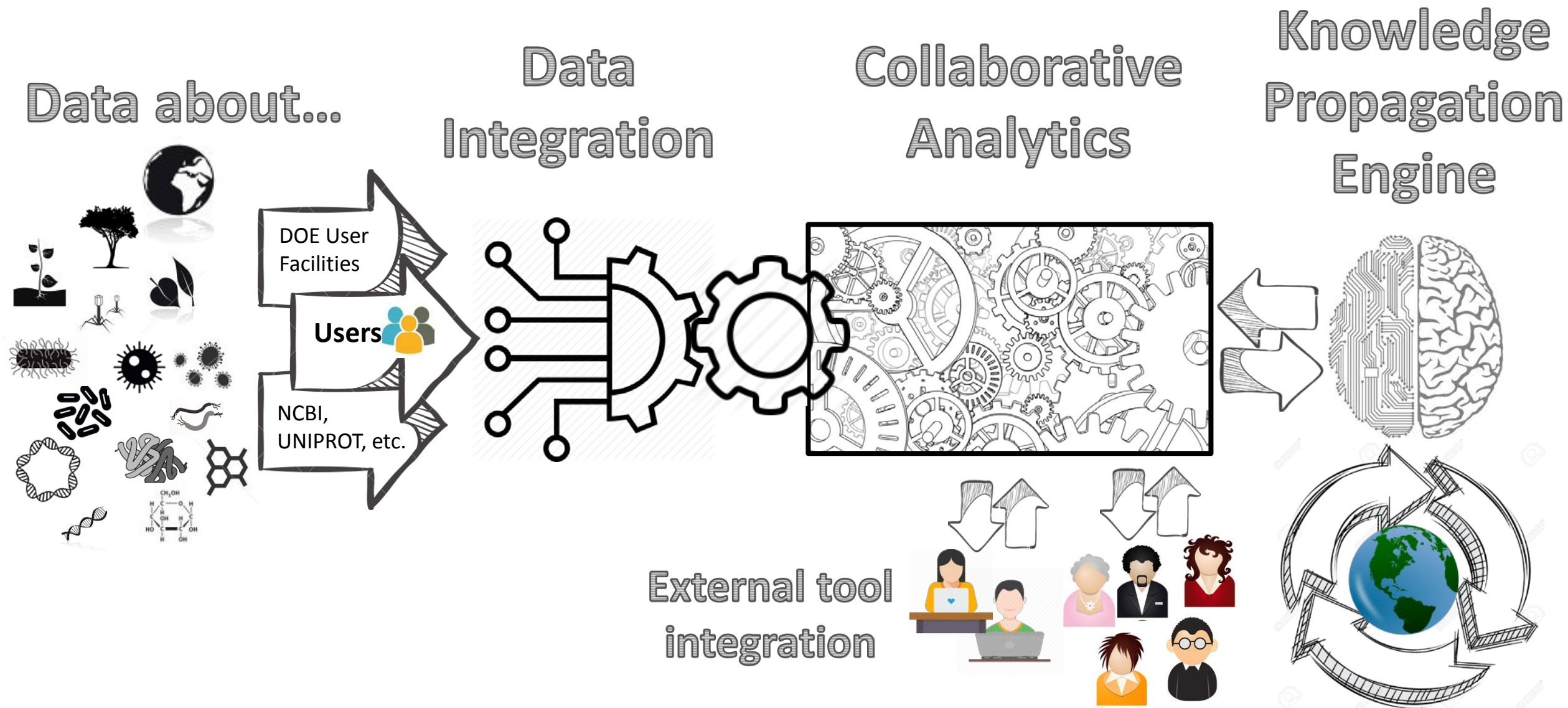
# KBase is a knowledge creation and discovery environment

- ❖ Integration of primary and derived products into a data model that supports human and machine learning analysis of all shared and published results across the system and automatic propagation of new results to biologically-related entities.

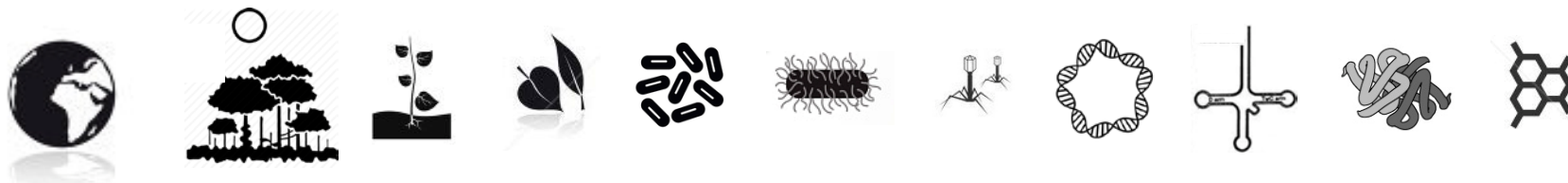


# KBase is a knowledge creation and discovery environment

- ❖ Supports sharing to individuals, groups and the world.
- ❖ Discovery of results and constant update of “knowledge” from integration and analysis.



Biological Inquiry



Scientific Community



KBase Users

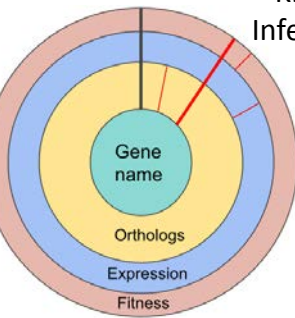


Reference Source Data

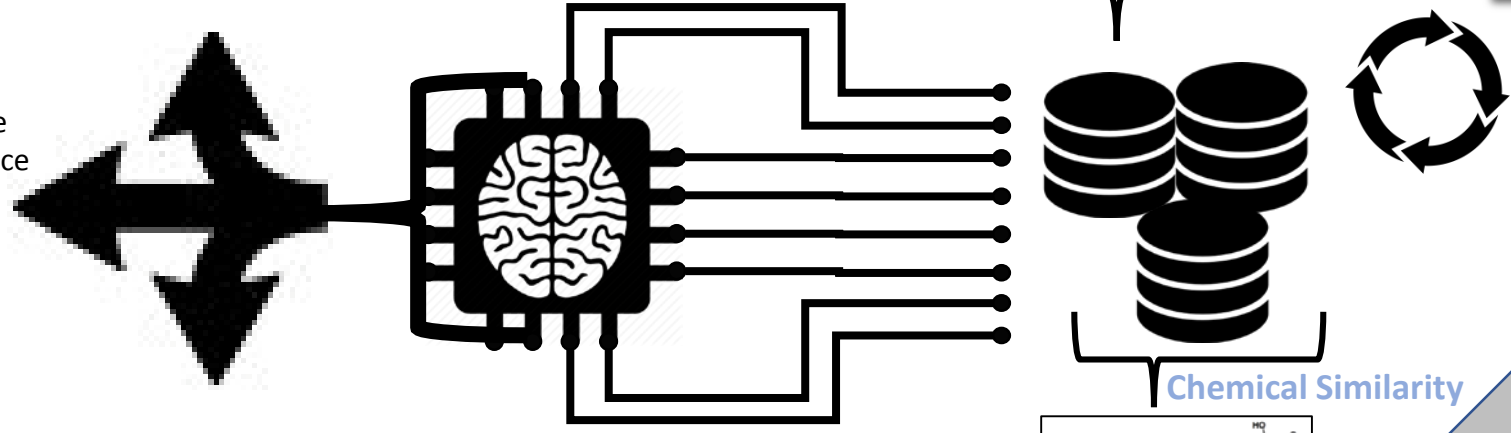


Narrative Development

Reference-Source Function



KBase Inference

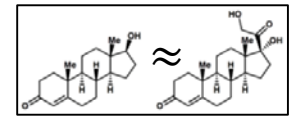
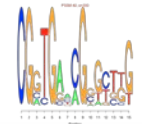
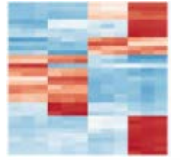


KBase Data and Apps

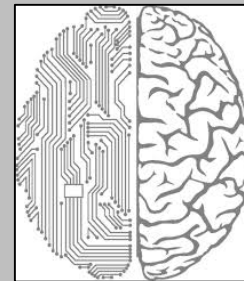
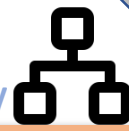
- Feeds
- Badges

- Evidence-based reannotation
- Propagation by orthology
- Metabolic model update

Biclustering Functional enrichment Motif finding



Taxonomy



RESKE

- ❖ Relation Engine
- ❖ Search
- ❖ Knowledge Engine

Informers

Propagators

Meta-analyzers

Homology

Connectors

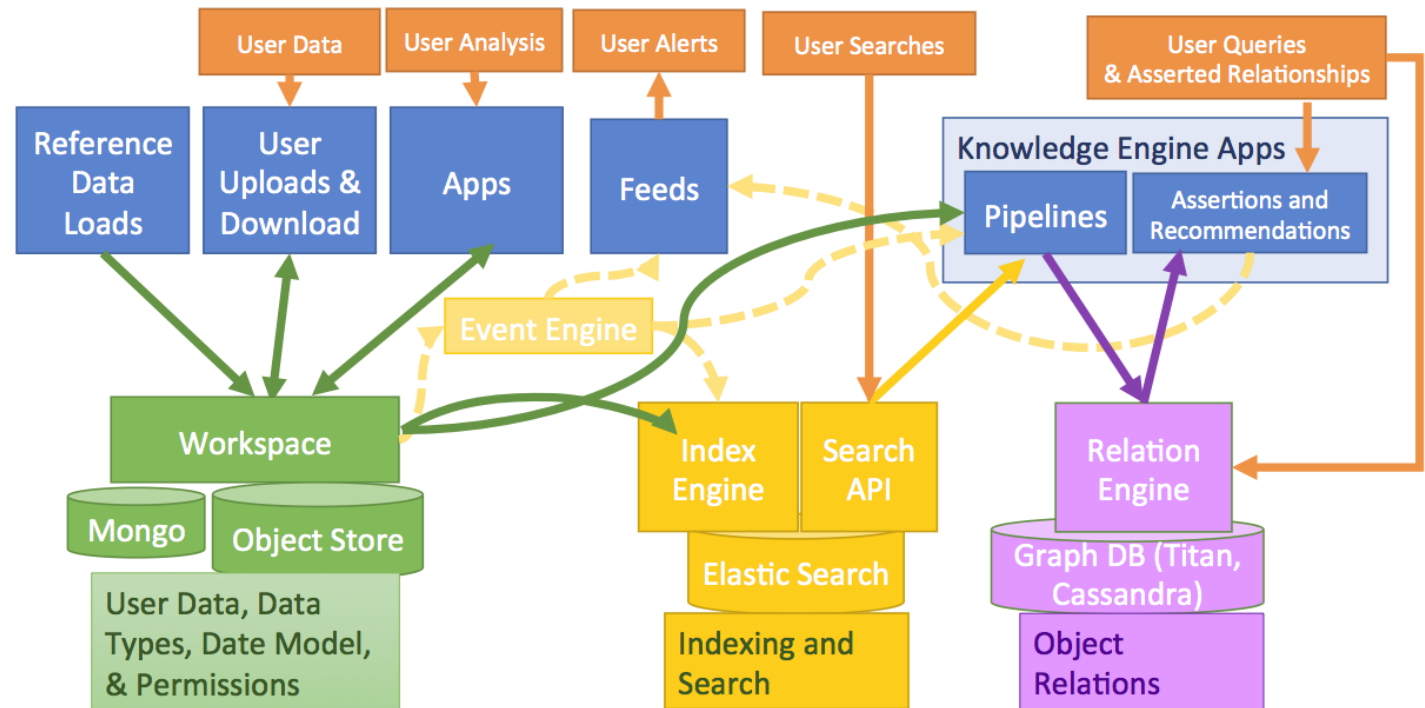
# Knowledge Engine

*Make new discoveries in biological function through integration of data and results shared by users across the system*

Incremental development plan:

1. Advanced search of all data and results across system
2. Relation Engine
3. Knowledge Engine

## Knowledge Engine Architecture





# PROTOTYPE



Dashboard



Catalog



Narratives



Search Data



Account



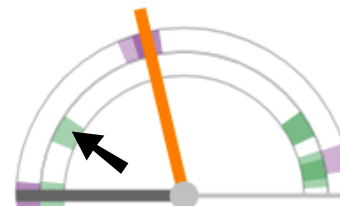
Feeds  
3 new

## Genes

Filter genes

Gene	Distance	User term	Ortholog term
PGA1_c13170	0.43	acetoin catabolic process GO:0045150	nitrogen fixation GO:0009399
WP_044041547.1	0.42	signal transduction by protein... GO:0023014	C4-dicarboxylate transport GO:0015740
cobA	0.42	oxidation-reduction process GO:0055114	glycerol transport GO:0015793
PGA1_RS02130	0.42	phosphorylation GO:0016310	cell wall organization GO:0071555
WP_014881553.1	0.42	phosphorylation GO:0016310	cell wall organization GO:0071555
PGA1_c19160	0.42	oxidation-reduction process GO:0055114	phenylacetate catabolic proc... GO:0010124
WP_014881857.1	0.42	alpha-amino acid metabolic p... GO:1901605	cardiolipin biosynthetic process GO:0032049
PGA1_RS09815	0.42	oxidation-reduction process GO:0055114	Gram-negative-bacterium-ty... GO:0043165
ftsW	0.42	cell wall organization GO:0071555	coenzyme biosynthetic process GO:0009108
PGA1_RS04735	0.42	cell wall organization GO:0071555	coenzyme biosynthetic process GO:0009108
PGA1_c35370	0.42	signal transduction by protein... GO:0023014	heme transport GO:0015886
PGA1_RS03090	0.42	DNA repair GO:0006281	antibiotic transport GO:0042891
WP_014879421.1	0.42	establishment of localization i... GO:0051649	magnesium ion transmembra... GO:1903830
PGA1_c09510	0.42	cell wall organization GO:0071555	coenzyme biosynthetic process GO:0009108

## Functional Profile



### Tooltip

Type	Fitness
Term	protein secretion by the type II secretion system
Id	GO:0015628
p-value	0.0361
distance	0.156

### Legend

GO Term	Id	Distance	p-value
<b>User</b>			
acetoin catabolic process	GO:0045150		
<b>Ortholog</b>			
nitrogen fixation	GO:0009399	0.43	0.004
<b>Fitness</b>			
cell motility	GO:0048870	0.83	0.007
acetoin catabolic process	GO:0045150	0.00	0.015
regulation of bacterial-type flagellum assembly	GO:1902208	0.93	0.015
protein secretion by the type II secretion system	GO:0015628	0.16	0.036
regulation of transcription, DNA-templated	GO:0006355	0.95	0.037
<b>Expression</b>			
nitrogen fixation	GO:0009399	0.43	0.004
acetoin catabolic process	GO:0045150	0.00	0.004
alpha-amino acid metabolic process	GO:1901605	0.40	0.017
organic substance transport	GO:0071702	0.93	0.021

# Thank you!



*KBase is a multi-institutional collaboration*



**DOE Systems Biology Knowledgebase**

**Thank you!**

**Questions?**

INTEGRATION and  
MODELING *for*  
PREDICTIVE BIOLOGY



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science

Office of Biological and Environmental Research