The DOE Systems Biology Knowledgebase

KBase Progress and Goals

Adam Arkin
Physical Biosciences Division
Lawrence Berkeley National Laboratory
DOE Systems Biology Knowledgebase

Components Enabling Its Development

Plants and Microbes for Energy and Environment

Biological Principles
- Metabolism Integration
- Interactions Visualization
- Proteins Algorithms
- Gene Expression Mathematics

Computing

2010 Knowledgebase R&D Project
Year-long effort funded by American Recovery and Reinvestment ACT (ARRA).
Results from this project, which was completed in September 2010, are underpinning Knowledgebase development:
- DOE Systems Biology Knowledgebase Implementation Plan
- ARRA pilot projects

User Community Data and Resources
The Knowledgebase will leverage and establish critical partnerships with other synergistic efforts:
- DOE Joint Genome Institute
- National Center for Biotechnology Information
- iPlant Collaborative
- DOE Office of Advanced Scientific Computing Research
- Others

University-Led Projects to Develop Computational Biology and Bioinformatic Methods Enabling the Knowledgebase
In 2010, the DOE Office of Biological and Environmental Research awarded funding to 11 projects for basic research that will support Knowledgebase development in four areas:
- Omic data integration
- Integrated pathway reconstruction
- Genomic annotation
- Whole cellular simulations
• **DOE mission**: *predict, control and design* the biological components of energetic processes and environmental balance.

• Complex missions with rapidly expanding, intricately related diverse data types require ways to augment scientists’ ability to:
  – Filter information
  – Focus attention
  – Ask the right questions
  – Leverage other minds
  – Make the impossible possible. 😊
Enable DOE Mission Science

Microbes

Communities

Plants

DOE Office of Science
Office of Biological and Environmental Research
Before I tell you how...

DID WE MEET SPECIFICATIONS?
Distributed Development

• Four major National Labs and more than 10 collaborating universities/research centers

• 50 people with diverse cultures and expertise and with different institutional and group alliances

• Enthusiasm to create a system to revolutionize understanding and application of biological systems.

• Deliver version 1 in 18 months and prove it can support DOE scientists!
Community Specifications for KBase: 2011

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

i. For microbial systems, from 100-1000 microbes: Reconstruct and Predict Metabolic and Gene Expression Regulatory Networks to Manipulate Microbial Function

ii. For plant systems, for 10 key plants related to DOE missions: Integrate Phenotypic and Experimental Data and Metadata to Predict Biomass Properties from Genotype and Assemble Regulatory Data to Enable Analysis, Cross-Comparisons, and Modeling

iii. For Microbial Communities: Model Metabolic Processes within Microbial Communities and Mine Metagenomic Data to Identify Unknown Genes
KBase 2013: 18 Months Deliverables Exceed Spec.

- A Federated infrastructure and 10GBit/s transfer capabilities.
  - Built for high speed data transfer over ESNET using 100 GBit/s rates.

- 2 petabytes data storage and 2000 cores for data processing including interface and between high performance computing and cloud computational resources
  - KBase Magellan has 12,000 cores for data processing via both Open Stack Cloud interfaces and Cluster Services
  - KBase has >3 Petabytes of storage capacity

- Development of core Knowledgebase integrated data and workflow analysis management tools including Application Programming Interfaces, semantic and user Interfaces.
  - Integrated KBase API specified and operational. Used by third parties to integrate function and build apps.
  - Integrated data model aware of 925 data types encompassing sequence reads, contigs, genomes, genome features, transcription data, fitness data and more.
  - 40 Interface description documents leading to 821 functions that can be compiled into use for PERL, Python, Java, and R.
  - Prototype Search, Workflow and Novel Narrative/Notebook interfaces for navigating, analyzing and building knowledge in KBase
For microbial systems, from 100-1000 microbes: Reconstruct and Predict Metabolic and Gene Expression Regulatory Networks to Manipulate Microbial Function

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

For plant systems, for 10 key plants related to DOE missions: Integrate Phenotypic and Experimental Data and Metadata to Predict Biomass Properties from Genotype and Assemble Regulatory Data to Enable Analysis, Cross-Comparisons, and Modeling

i. Over 175 eukaryotic genomes including many variants of Poplar, Arabidopsis, Sorghum, Chlamydomonas, Brachypodium, Miscanthus, Switchgrass as well as many other algae and fungi.
ii. Phenotypes for genome variants of plants and services for calling the genetic variation among individuals.
iii. Services for variation calling, mapping genotype-to-phenotype via GWAS style analysis and tools for candidate gene filtering, trait modeling, and pathway enrichment,
iv. 731 gene expression experiments in Arabidopsis and Poplar; Plant co-expression network analysis for all.
v. Initial plant metabolic modeling

For Microbial Communities: Model Metabolic Processes within Microbial Communities and Mine Metagenomic Data to Identify Unknown Genes

i. Access to 11,000 metagenomes (>21 TBp)
ii. Integrated KBase access to QIIME functionality
iii. New tools for metagenome sequence quality assessment and experimental design
iv. Services for taxonomic and gene identification, abundance, and a host of other functions
Captain, I am detecting large quantities of win in this sector.
The community document is not a specification, though...

WHY DO WE NEED KBASE?
Our goals drive our specification

• We believe it is necessary to solve the grand challenges in biological control of environmental and energy process.

• We believe it is necessary to deal with deep issues in scientific communication and reproducibility.

• We need to lower the barriers to computationally efficient use of advanced algorithms and data from diverse producers.

• We believe we need to do this together--leveraging many minds.
The number of bioinformatics resources is huge

- An increasing number of data warehouses since biology is becoming a big data discipline (e.g., NCBI, Ensembl, etc.)

- Specialized applications and databases for relatively generic analyses (e.g., MG-RAST and MicrobesOnline)

- Evolving libraries of sophisticated computational biology algorithms for use in programming environments (e.g., Bioconductor)

- Workflow tools that allow the chaining of these algorithms together by non-programmers (e.g., Galaxy and Taverna)

- Workflow sharing tools to allow people to use each other’s work products

- Open-access publication of journal articles with increasing use of semantic tagging

- Scientific social networks (e.g., ResearchGate, Epernicus, etc.)

So what is missing?!
• KBase drives data through models to predictions and experimental design.

• KBase accelerates reproducible, reusable, and transparent science.

• KBase deeply enables scientists to work together to approach complex biological problems.

• KBase gives credit where it is due and privacy where it is needed.

• KBase is an open software and data environment to which others can contribute and with which others can build.
Empowering Users

User Access

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

Integrated Data

Predictive models
Hypothesis
Visualizations
Comparative analyses

Analysis

Simulation and Prediction

Contributions from Community Developers

Services and Tools

DOE Office of Science
Office of Biological and Environmental Research
What can we do that was hard before?

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

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

**Compute**
- Large-scale assembly computations
- Large-scale BLAST and sequence alignment
- Metabolic model gapfilling
- Metabolic model reconciliation

**Workspace**
- Genome sequence
- Annotation
- Metabolic model reconstruction
- Phenotype simulation
- Phenotype Reconciliation

**Central store**
- Reference sequences
- Reference annotations
- Reference models
- Reference phenotypes
- Comparison genomes
- Comparison phenotypes

**Validation experiments**
Biological trends across the tree of life

- **Genome sequence**
  - Large-scale query for genomes

- **Workspace**
  - Genome sequence
  - Reference sequences

- **Central store**
  - Reference annotations
  - Reference models
  - Reference phenotypes

- **Construction application of templates**
  - Metabolic model
  - Predicted phenotypes

- **Statistical analysis of data**
  - Uploaded exp. data
  - Reconciled model
  - Comparison genomes
  - Comparison phenotypes

- **Phylogenetic distance**
- **Metabolic distance**

Assessing “completeness” of our understanding
Driving discovery in experimental biology

Predict nutrients required for biomass biosynthesis

Predict potential nutrients that may inhibit growth

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

Explore microbe-microbe codependency or microbe-host dependency hypotheses

Validate experiments

---

**Genome sequence**
- Assembly
- Annotation
- Modeling
- Regulatory network reconstruction
- Prediction of culture conditions
- Compare close cultivable organisms

**Workspace**
- Genome sequence
- Annotation
- Metabolic model
- Regulatory network
- Predict media

**Central store**
- Reference sequences
- Reference annotations
- Reference models
- Reference networks
- Reference Media
- Comparison genomes
- Comparison phenotypes

KBase
DOE Systems Biology Knowledgebase

DOE Office of Science
Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY
KBase Communities Supports Massive Metagenomics

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

Who are they?

What are they doing?

PHENYLALANINE, TYROSINE AND TRYP托PHAN BIOSYNTHESIS
KBase-Plants: Targeted genomes

- Poplar
- Arabidopsis
- Sorghum
- Chlamydomonas
- Brachypodium
- Miscanthus
- Switchgrass
From reads through variants to phenotypes
Align and call SNPs from 35M 80bp (14Gbp) reads with maize genome (zmb73v2)
Identified 372k high confidence SNPs

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

*estimated time
## Maize Population Analysis

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

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

*estimated time
Networks: Cross-cut example

I. Networks-based knowledge discovery

- Networks-based ontologies
  - SFAs
  - BRCs
  - ...  
  - datasets
  - ontologies

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

II. Facilitate building predictive models

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

III. Infrastructure for scientific social networks

- Networks of scientific communities and projects
  - Support narrative interface. Social networks connecting users and joint projects

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

- Networks quality control
  - Assign quality measures to KBase networks to allow healthy competition between algorithms and teams
I. Datasets: collected, processed, classified

<table>
<thead>
<tr>
<th>KBase domain</th>
<th>Dataset sources</th>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Microbes</td>
<td>7</td>
<td>5099</td>
</tr>
<tr>
<td>Plants</td>
<td>6</td>
<td>46</td>
</tr>
<tr>
<td>Communities</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network types</th>
</tr>
</thead>
<tbody>
<tr>
<td>REGULATORY_NETWORK</td>
</tr>
<tr>
<td>CO_FITNESS</td>
</tr>
<tr>
<td>PROT_PROT_INTERACTION</td>
</tr>
<tr>
<td>CO_EXPRESSION</td>
</tr>
<tr>
<td>METABOLIC_SUBSYSTEM</td>
</tr>
<tr>
<td>FUNCTIONAL_ABUNDANCE</td>
</tr>
<tr>
<td>FUNCTIONAL_ASSOCIATION</td>
</tr>
<tr>
<td>PHYLOTYPE_ABUNDANCE</td>
</tr>
</tbody>
</table>

II. Networks API: provide heterogeneous networks in unified format

Networks build methods

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

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

Datasets management methods

- list<Dataset> allDatasets()
- list<DatasetSource> allDatasetSources()
- list<NetworkType> allNetworkTypes()
Iterative exploration and building network components

powered by the developed Networks API

Gene clusters associated with genes from dock panel

Putative members of the same network component (docked genes)
ENSURING IMPACT

Working with collaborators early
Using KBase framework to enable PROM

- Nathan Price’s probabilistic regulation of metabolism (PROM) integrates metabolic models and functional data to make better predictions of growth given genetic or transcription variation.

- Difficult to access and for people to use.

- Previously applied to just two organisms.

- Now a KBase service that can be applied to any genome for which there is expression or variation data.

- Testing on *Shewanella oneidensis* MR-1 with transcription, TF knockout and growth/fitness data available in KBase.
Kirsten Hofmockel, Iowa State, used KBase services to analyze and compare metagenomes from different sized soil aggregates across different crop treatments.
Variation in Lignin Composition and Content

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

- Assemble a population
- Phenotype the population
- GWAS Analysis
  - Create a SNP library
  - "28%"
HOW ARE WE BUILDING KBASE?
Define over research “paths”

KBase Workflow Model

- Functional Inference in Genomes and Metagenomes
- Metabolic, Regulatory, and Community Network Inference
- Inference of gene structure and annotation by homology
- Direct measurement and Guilt-by-Association Function Inference
- Inference of networks and suggestions for gap-filling
- Behavioral prediction and design

Driving data towards dynamic models of function

Model-based Analysis & Eng.
Comparative Analysis of Predictions
Link design to KBase Concepts

KBase Workflow Model

- Functional Inference in Genomes and Metagenomes
- Inference of gene structure and annotation by homology
- Direct measurement and Guilt-by-Association Function Inference
- Behavior prediction and design
- Metabolic, Regulatory, and Community Network Inference
- Inference of networks and suggestions for gap-filling
- Driving data towards dynamic models of function

Measures of confidence and quality

Clearinghouse of formal predictions/hypotheses

User communities

Comparative Analysis of Predictions

Model-based Analysis & Engineering
Inference of gene structure and annotation by homology

Direct measurement and Guilt-by-Association Function Inference

Inference of networks and suggestions for gap-filling

Measures of Confidence and Quality

Clearinghouse of Predictions/Hypotheses

User Communities
• Labs are prototypes and applications that demonstrate KBase present and possible future function.

• Early stage functionality both provides utility and a way for the community to provide feedback.
IRIS is a web-based terminal to give you access to the full KBase command line without the need to install anything!
### Command Line

- **Command-line environment for programmers**

- **Available via downloadable Macintosh DMG or Ubuntu (linux) image**

- **Installs clean environment for KBase application development**

### Populus reads to SNPs in Five Easy Steps

<table>
<thead>
<tr>
<th>Step</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Identify reference genome&lt;br&gt;$ all_entities_Genome -f scientific_name</td>
</tr>
<tr>
<td>2.</td>
<td>Upload Reads to KBase cloud&lt;br&gt;$ jk_fs_put_pe populus.1.fq.gz populus.2.fq.gz populus</td>
</tr>
<tr>
<td>3.</td>
<td>Align Reads with Bowtie2&lt;br&gt;$ jk_compute_bowtie -in=populus.pe -org=populus -out=populus_align</td>
</tr>
<tr>
<td>4.</td>
<td>Call SNPs with SAMTools&lt;br&gt;$ jk_compute_samtools_snp -in=populus_align -org=populus -out=populus_snps</td>
</tr>
<tr>
<td>5.</td>
<td>Merge and Download VCF files&lt;br&gt;$ jk_compute_vcf_merge -in=populus_snps --alignments=populus_align -out=populus.vcf&lt;br&gt;$ jk_fs_get populus.vcf</td>
</tr>
</tbody>
</table>
Prototype collaborative workspaces

View Objects Produced by Analysis in an Online Collaborative Workspace

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

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

Simple Tabular Views

Interactive metabolic maps and models

Comparative Model Viewer

Network Analysis Workbench

DOE Office of Science
Office of Biological and Environmental Research

U.S. DEPARTMENT OF ENERGY
Search Architecture

- KBase Search Web Application
  - KBase Search Developer API
    - External Application
      - External Application
    - KBase Solr Infrastructure
      - Central Store
      - Auxiliary Store
      - Workspace
Search with keyword: **thermocellum**

KBase Search

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

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

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

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

... 

Results organized using central data model

Link to external data in PUBMED
Gene page

KBase Search

thermocellum

Search Advanced

Carbon storage regulator

Genome: Clostridium thermocellum ATCC 27405
KBase Id: kbtg22.peg.1005
DNA Size: 229
Location: kbtg22.peg.2967730-2967730

Subsystem
1. Carbon storage regulator
Carbon storage regulator

Co-occurrences

# | Id 1 | Id 2 | Occurrences
---|------|------|-----------
1  | kbtg22.peg.1005 | kbtg22.peg.1459 | 26
2  | kbtg22.peg.1005 | kbtg22.peg.1214 | 172
3  | kbtg22.peg.1005 | kbtg22.peg.1848 | 49
4  | kbtg22.peg.1005 | kbtg22.peg.1957 | 164
5  | kbtg22.peg.1005 | kbtg22.peg.1919 | 164

DNA Sequence

Protein Sequence

Links to related content
Gene details
Sequences
Initial KBase Services

- IRIS
- KBase Notebook
- Auxiliary Store Service
- Communities API
- Metagenomics Analysis Tools R
- QC Service
- Annotation Service
- Central Store
- Core Model Viewer
- Experiment data
- FBA Modeling
- Metabolic Map Viewer
- Microbes Model Builder
- Phispy
- Probabilistic Annotation
- Protein Info service
- Regulation Service
- Similarity Service
- Translation Service
- Workspace Service
- KB Model Seed
- Tree Service
- Assembly Service
- Authorization Service
- Network Service
- Genotype Phenotype Service
- Genotyping Service
- Ontology Service Plant
- Expression Service
- Authentication and Authorization Client
- Cluster Service
- ERDB service
- File Type Service
- ID Service
- Registry
- Type Compiler
Data Stores: Storing a diverse range of biological data requires multiple approaches:
- Highly structured data in relational databases
- Frequently changing user data
- Large bulk data

The KBase Data Stores

- Petabytes of Raw Data
- Flexible Storage For Workspaces
- Structured Storage For Curated Data
Flux Balance Analysis Modeling Services

**FBA MODELING SERVICES**
- Model Reconstruction & Analysis Service
  - Command Line Client
  - Web Interface
  - KBASE Iris Interface
  - JSON RPC API

**Analysis Server**
- Model Reconstruction
  - Reconstruction of new models from genome annotations
- Model Curation
  - Functions permitting user manipulation of models
- Flux Balance Analysis
  - Analysis of pathway utilization, gene essentiality, and flux
- Model Gapfilling
  - Filling gaps in models to enable growth in specified conditions
- Phenotype Simulation
  - Load and simulation of growth phenotypes
- Phenotype Reconciliation
  - Optimization of model to fit phenotype data
- Model Data Retrieval
  - Enables the retrieval of model data, phenotypes, and FBA results
- Job Management
  - Enables management of long-running jobs

**Workers**
- Magellan Nodes
- Cluster Services

**Job Queues**
- Flux Balance Analysis
  - Phenotype simulation
  - Model gapfilling
  - Model gap generation
  - Model optimization

**Authentication**
- Globus Online user support

**Data Repository**
- WorkSpace: Collaborative store for models and analysis
- Central Store: Database of reference genomes, models, and data
Helping users and developers manage complexity

Website and Outreach
KBase Training Resources

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

- **User Manuals** are online procedural documents describing each of the services with detailed information about functionality and usage.

- **Tutorials** are both online training resources and in-person sessions* focused on using specific aspects of KBase. Online resources are in the form of written walk-throughs and video tutorials. In-person tutorials will typically be 2-3 days, and may be immediately followed by a 1-day workshop as part of the same agenda. In-person tutorials are bimonthly.

- **Workshops** are in-person sessions* focused on scientific problems. We will guide you through the use of KBase to accomplish tasks that are of interest to you. Currently, workshops will be single-day sessions that are paired with relevant tutorials (i.e., offered at the end of a tutorial session). Later in the project, workshops will be stand-alone, multiple-day events held annually.

- **Bootcamps** are in-person sessions focused on developing computational tools in KBase. Join our developers and help build KBase.

**plus ... Monthly webinars**
Scheduled KBase Events can be found at [http://kbase.us](http://kbase.us)

**Calendar**

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am Microbes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am Microbes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BERAC Presentations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am Microbes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DOE/NIFA Plant Feedstocks Genomics for Bioenergy</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Genomic Science Contractors-Grantees Meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am Microbes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10am Microbes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Events shown in time zone: Pacific Time

**Upcoming Training Sessions**

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

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

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

Contact us at [outreach@kbase.us](mailto:outreach@kbase.us) to register for a session.
Online Resources

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

help@kbase.us
Extensive Tutorials and Documentation

Tutorials

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

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

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

Getting Started

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

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

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

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

IRIS: a Browser-based Framework for Interacting with KBase

Introduction to the KBase IRIS Interface
Extensive Tutorials and Documentation

API and Command-Line Documentation

**Reference Documentation**

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

**CDM Command Line Scripts**

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

**CDM API**

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

**CDMI Entity-Relationship Command Line Scripts**

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

**CDMI Entity-Relationship API**

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

**ID Server Command Line Scripts**

The ID Server command line scripts allow ID allocation and lookup services. This is the core service for generating ID Base Identifier (ID Base I)

**CDM**

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

**CDMI Entity-Relationship Command Line Scripts**

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

**CDMI Entity-Relationship API**

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

**ID Server Command Line Scripts**

The ID Server command line scripts allow ID allocation and lookup services. This is the core service for generating ID Base Identifier (ID Base I)
MOAs and Collaborations

- Coordination and shared milestones with the BRCs
- Infrastructure collaborations with JGI, EMSL, NERC.
- Scientific collaborations with ENIGMA, PMI, and other SFAs
- MOAs with NSF NSF for iPLANT and pending for USDA
- Preparation for collaboration and support for future DOE calls
- Needs survey, bootcamps, and scientific discussions with >70 community scientists.
What next?
Building a starship in deep space?
• The data model must evolve to support the modeling mission

• A framework for turning bioinformatics algorithms’ output into models needs further development

• **Theory for integrative, cross-scale predictive biology under development**

• Much better ontologies for nearly everything

• Building the social tools.

• Need to improve data import, quality assessment and metadata

• Efficient incorporation of new third party algorithms and support with scalable computing.

• Growing a strong external development community while maintaining quality, stability and vision.

• Launching the KBase Foundation to ease licensing and growth of KBase user participation.

A more concrete view of the “Narrative” interface
Narrative Desiderata

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

People I’m following
- bobtheguy@kbase.us started a new narrative entitled "GWAS of Klebsiella spp."
- kristen@uchicago.edu commented on bobtheguy@kbase.us's narrative

Organisms I’m following

E Coli H25
- pipsqueak@concordia.gov started a new narrative entitled "GWAS Analysis of a New Strain of E.coli"
- qsquirrel@concordia.gov commented on a narrative entitled "E. coli H25 Regulatory Model"
- qsquirrel@concordia.gov cloned the narrative entitled "E. coli H25 Regulatory Model" as "E. coli H25 Regulatory Model with Gapfilling"
- qsquirrel@concordia.gov added a citation to the narrative entitled "E. coli H25 Metabolic Model"

Projects I’m following

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

Manage my subscriptions

You have:
- 3 notifications

Your narratives have been:
- commented on 14 times
- linked to 13 times
- cloned 3 times
- viewed 38 times

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

You have:
- 67 Contribution Points
- based on the following:
  - 1 published narrative
  - 24 comments
  - 13 evidence links
  - 2 public datasets
  - 0 user-defined functions
Untitled
Joe Scientist, Concordia National Laboratory
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory
I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

I'll start by getting the genome of E. coli F11 from the KBase Central Data Store.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

- **Upload a file**

- **Search the Central Data Store**
  
  ![](Escherichia coli F11)

  ![Escherichia coli F11](EscherichiacoliF11.png)
Search KBase

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

1. Escherichia coli TW10828
   - Bacteria
   - Contigs: 227
   - RNA Features: 122
   - Protein Encoding Genes: 5139
   - Genome Size: 6280267 bp
   - GC Content: 50.84 %
   - Details

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

3. Escherichia coli E482
   - Bacteria
   - Contigs: 94
   - RNA Features: 94
   - Protein Encoding Genes: 4752
   - Genome Size: 4633189 bp
   - GC Content: 50.70 %
   - Details
Searched for **Escherichia coli** found 11385 results.
Now showing 10 results.

- **4. Escherichia coli F11**
  - **Bacteria**
  - Contigs: 119
  - RNA Features: 96
  - Protein Encoding Genes: 5102

- **5. Escherichia coli 9.1649**
  - **Bacteria**
  - Contigs: 33
  - RNA Features: 112
  - Protein Encoding Genes: 5037

- **6. Escherichia coli W**
  - **Bacteria**
  - Contigs: 1
  - RNA Features: 0
  - Protein Encoding Genes: 6
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

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

**Escherichia coli F11 Genome, annotated**

- **Domain:** Bacteria
- **Taxon:** gamma-proteobacteria
- **KBase ID:** KBlg.691
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Contigs:** 2
- **Base pairs:** 5,280,267
- **PEGs:** 232
- **GC Content:** 50.64%

Functions

- annotation
- browse genome
- metabolic reconstruction
- build FBA model
- FBA
- gapfilling
- geno2pheno
- variation
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

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

<table>
<thead>
<tr>
<th><strong>Escherichia coli F11 Genome, annotated</strong></th>
<th><strong>KBase ID:</strong> KBr.g.691</th>
<th><strong>Contigs:</strong> 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain:</strong> Bacteria</td>
<td><strong>Source:</strong> Joe Scientist</td>
<td><strong>Base pairs:</strong> 5,280,267</td>
</tr>
<tr>
<td><strong>Taxon:</strong> gamma-proteobacteria</td>
<td><strong>Version:</strong> Mar 27, 2013 08:13:00 PST</td>
<td><strong>PEGs:</strong> 232</td>
</tr>
<tr>
<td><strong>Complete genome</strong></td>
<td></td>
<td><strong>GC Content:</strong> 50.64%</td>
</tr>
</tbody>
</table>
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

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

**Escherichia coli F11 Genome, annotated**  

- **Domain:** Bacteria  
- **Taxon:** gamma-proteobacteria  
- **KBase ID:** KBlg.691  
- **Source:** Joe Scientist  
- **Version:** Mar 27, 2013 08:13:00 PST  
- **Contigs:** 2  
- **Base pairs:** 5,280,267  
- **PEGs:** 232  
- **GC Content:** 50.64%
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

**Escherichia coli F11 Genome, annotated**

- **Domain:** Bacteria
- **Taxon:** gamma-proteobacteria
- **KBase ID:** KBlg.691
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Contigs:** 2
- **Base pairs:** 5,280,267
- **PEGs:** 232
- **GC Content:** 50.64%

Let's take a look at this genome.
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

**Escherichia coli F11 Genome, annotated**

- **Domain:** Bacteria
- **Taxon:** gamma-proteobacteria
- **Kbase ID:** Kbp.l.691
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Complete genome**
- **Contigs:** 2
- **Base pairs:** 5,280,267
- **PEGS:** 232
- **GC Content:** 50.64%

Let’s take a look at this genome.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

**Escherichia coli F11 Genome, annotated**
- **Domain**: Bacteria
- **Taxon**: gamma-proteobacteria
- **Source**: Joe Scientist
- **Version**: Mar 27, 2013 08:13:00 PST
- **Complete genome**

**KBase ID**: KBlg.691
**Contigs**: 2
**Base pairs**: 5,280,267
**PEGs**: 232
**GC Content**: 50.64%

Let's take a look at this genome.

**Genome Information Browser**
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

**Escherichia coli F11 Genome, annotated**
- **Domain:** Bacteria
- **Taxon:** gamma-proteobacteria
- **KBase ID:** K8lfg.691
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Contigs:** 2
- **Base pairs:** 5,280,267
- **PEGs:** 232
- **Complete genome**
- **GC Content:** 50.64%

Let's take a look at this genome.

Now I'm going to run a metabolic reconstruction on this genome. Let's see how the initial one comes out.

This is the first time I've done this for this particular strain, so it will be interesting to see which roles are found.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

Run metabolic reconstruction

Dataset:
E. coli F11 Genome

Other Options:

Now let's run the metabolic reconstruction. Once the initial one comes out, we'll look to see which roles are found.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

<table>
<thead>
<tr>
<th>Escherichia coli F11 Genome, annotated</th>
<th>KBase ID: KB[3.691]</th>
<th>Contigs: 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain: Bacteria</td>
<td>Source: Joe Scientist</td>
<td>Base pairs: 5,280,267</td>
</tr>
<tr>
<td>Taxon: gamma-proteobacteria</td>
<td>Version: Mar 27, 2013 08:13:00 PST</td>
<td>PEGs: 252</td>
</tr>
<tr>
<td></td>
<td>Complete genome</td>
<td>GC Content: 50.64%</td>
</tr>
</tbody>
</table>

Let's take a look at this genome.

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

Run Metabolic Reconstruction on E. coli F11 Genome

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

<table>
<thead>
<tr>
<th>Escherichia coli F11 Metabolic Reconstruction</th>
<th>Source: calculated</th>
<th>Rules: 29,294</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version: Mar 27, 2013 08:13:00 PST</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

---

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

- **KBBase ID:** KB[3.691]
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Complete genome**

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

---

Let’s take a look at this genome.

**Genome Information Browser**

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

**Run Metabolic Reconstruction on E. coli F11 Genome**

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

---

**Escherichia coli F11 Metabolic Reconstruction**

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

- **Roles:** 29,294

---

**Refining the Metabolic Model for *Escherichia coli* F11**
I've grown this without serine, but the genes for serine are missing in this automated reconstruction. I think you'll need to add kb|g.362.peg.287, kb|g.382.peg.123, and kb|g.382.peg.898.
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

**Escherichia coli F11 Genome, annotated**

- **Domain:** Bacteria
- **Taxon:** gamma-proteobacteria
- **KBase ID:** KB|5.691
- **Source:** Joe Scientist
- **Version:** Mar 27, 2013 08:13:00 PST
- **Complete genome**

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

Let's take a look at this genome.

**Genome Information Browser**

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

**Run Metabolic Reconstruction on E. coli F11 Genome**

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

**Escherichia coli F11 Metabolic Reconstruction**

- **Source:** calculated
- **Version:** Mar 27, 2013 08:13:00 PST
- **Roles:** 29,294

1 Comment
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

Run build FBA model

Dataset:
- E. coli F11 Genome

Other Options:
- [run] [cancel]

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

Joe Scientist, Concordia National Laboratory

Run Metabolic Reconstruction on E. coli F11 Genome
Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4:15:33pm

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

Run Build FBA Model on E. coli F11 Genome
Started: 9/27/2012 4:15:32pm
Finished: 9/27/2012 4:15:33pm

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

Joe Scientist, Concordia National Laboratory

Run Metabolic Reconstruction on *E. coli* F11 Genome

**Run FBA**

**Dataset:**
- *E. coli* F11 Genome FBA Model

**Other Options:**
- run
- cancel

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

Joe Scientist, Concordia National Laboratory

**Run Build FBA Model on *E. coli* F11 Genome**

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

**Escherichia coli F11 FBA Model**

Source: calculated  
Version: Mar 27, 2013 08:18:00 PST  
Reactions: 29,294  
Compounds: 2038  
Annotations: 234

**Run FBA on *E. coli* F11 Metabolic Model**

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

**Escherichia coli F11 FBA Result**

Source: calculated  
Version: Mar 27, 2013 08:18:00 PST  
P(Growth): 0.98
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

Run Build FBA Model on E. coli F11 Genome

- Upload a file
- Search the Central Data Store

Escherichia coli F11 FBA Result
Source: calculated
Version: Mar 27, 2013 08:13:00 PST

P(Growth): 0.98
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

Upload a File to your Narrative

**File:**

- Choose File
- ecoli_F11_growth02.1213.txt

**Dataset Type:**

- select...

**Dataset Name:**

- upload file

*Escherichia coli* F11 FBA Result

- Source: calculated
- Version: Mar 27, 2013 08:13:00 PST
- P(Growth); 0.98

Refining the Metabolic Model for *Escherichia coli* F11
Refining the Metabolic Model for *Escherichia coli* F11

Joe Scientist, Concordia National Laboratory

**Escherichia coli F11 FBA Model**

- Source: calculated
- Version: Mar 27, 2013 08:13:00 PST
- Reactions: 29,294
- Compounds: 2098
- Annotations: 234

**Run FBA on E. coli F11 Metabolic Model**

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

**Escherichia coli F11 FBA Result**

- Source: calculated
- Version: Mar 27, 2013 08:13:00 PST
- F1 Growth: 0.98

**Escherichia coli F11 Growth**

- Source: user uploaded
- Version: Mar 27, 2013 08:13:00 PST
- growth x nickel abundance
Refining the Metabolic Model for Escherichia coli F11

Joe Scientist, Concordia National Laboratory

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

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

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

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

Joe Scientist, Concordia National Laboratory
Newsfeed

People I'm following

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

Organisms I'm following

E Coli F11

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

Projects I'm following

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

Manage my subscriptions
Concept: Interactive community knowledge

Narrative Graphs Provide Measures of Activity and Influence

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

Narrative Code Versioning and Scenario Branching

Narrative Query Update

Narrative Data Change

Project Linkage and Citation
Into the future

• The data model must evolve to support the modeling mission

• A framework for turning bioinformatics algorithms’ output into models needs further development

• **Theory for integrative, cross-scale predictive biology under development**

• Much better ontologies for nearly everything

• Building the social tools.

• Need to improve data import, quality assessment and metadata

• Efficient incorporation of new third party algorithms and support with scalable compute.

• Growing a strong external development community while maintaining quality, stability and vision.

• Launching the KBase Foundation to ease licensing and growth of KBase user participation.

A more concrete view of the “Narrative” interface
Thanks to our Team and DOE!

<table>
<thead>
<tr>
<th>Name</th>
<th>Institute</th>
<th>Department</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkin, Adam</td>
<td>LBNL</td>
<td>Management</td>
</tr>
<tr>
<td>Baumohl, Jason</td>
<td>LBNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Best, Aaron</td>
<td>BNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Bischof, Jared</td>
<td>ANL</td>
<td>Communities (Folker)</td>
</tr>
<tr>
<td>Bowen, Ben</td>
<td>LBNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Bronner, Steven</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Brettin, Tom</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Canon, Shane</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Chan, Stephen</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Chandonia, John-Marc</td>
<td>LBNL</td>
<td>Microbes, Networks</td>
</tr>
<tr>
<td>Chivian, Dylan</td>
<td>LBNL</td>
<td>Communities</td>
</tr>
<tr>
<td>Cholia, Shreyas</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Conrado, Neal</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Cottingham, Bob</td>
<td>ORNL</td>
<td>Management</td>
</tr>
<tr>
<td>Davidson, Brian</td>
<td>ORNL</td>
<td>Outreach</td>
</tr>
<tr>
<td>Dehal, Paramvir</td>
<td>LBNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>De Jongh, Matt</td>
<td>Hope</td>
<td>Microbes</td>
</tr>
<tr>
<td>Desai, Narayan</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Devold, Scott</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Diz, Terry</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Drake, Meghan</td>
<td>ORNL</td>
<td>Outreach</td>
</tr>
<tr>
<td>Fang, Gang</td>
<td>Yale</td>
<td>Plants</td>
</tr>
<tr>
<td>Fei He</td>
<td>BNL</td>
<td>Plants</td>
</tr>
<tr>
<td>Frybarger, Paul</td>
<td>ANL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Gerstein, Mark</td>
<td>Yale</td>
<td>Plants</td>
</tr>
<tr>
<td>Glass, Elizabeth</td>
<td>ANL</td>
<td>Outreach</td>
</tr>
<tr>
<td>Gordon, India</td>
<td>ANL</td>
<td>Administration</td>
</tr>
<tr>
<td>Greiner, Arnette</td>
<td>LBL/NERSC</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Gurtowski, James</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Harris, Nomi</td>
<td>LBNL</td>
<td>Management</td>
</tr>
<tr>
<td>Harrison, Travis</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Henderson, Matt</td>
<td>LBNL</td>
<td>Communities</td>
</tr>
<tr>
<td>Henry, Chris</td>
<td>ANL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Howe, Adria</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Jochumik, Marcin</td>
<td>LBNL</td>
<td>Microbes/Communities</td>
</tr>
<tr>
<td>Keegan, Kevin</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Keller, Keith</td>
<td>LBNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Kora, Guru</td>
<td>ORNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Kumari, Sunita</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Land, Miriam</td>
<td>ORNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Lauer, Michael</td>
<td>ORNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Lochovsky, Lucas</td>
<td>Yale</td>
<td>Plants</td>
</tr>
<tr>
<td>Maslov, Sergey</td>
<td>BNL</td>
<td>Management</td>
</tr>
<tr>
<td>Meyer, Folker</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Mouton, Steve</td>
<td>ORNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Mulligan, Silvia</td>
<td>ANL</td>
<td>Administration</td>
</tr>
<tr>
<td>Novickchov, Pavel</td>
<td>LBNL</td>
<td>Infrastructure/Microbes</td>
</tr>
<tr>
<td>Oh, Taeyun</td>
<td>UC Davis</td>
<td>Plants</td>
</tr>
<tr>
<td>Olsen, Gary</td>
<td>UIUC</td>
<td>Microbes</td>
</tr>
<tr>
<td>Olson, Bob</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Olson, Dan</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Overbeek, Ross</td>
<td>ANL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Paczian, Tobias</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Palumbo, Tony</td>
<td>ORNL</td>
<td>Outreach</td>
</tr>
<tr>
<td>Pernello, Bruce</td>
<td>ANL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Pashkamak, Shiran</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Poon, Sarah</td>
<td>LBL/NERSC</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Price, Gavin</td>
<td>LBL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Ranjan, Priya</td>
<td>ORNL</td>
<td>Plants</td>
</tr>
<tr>
<td>Riehl, Bill</td>
<td>LBNL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Ronald, Pamela</td>
<td>UC Davis</td>
<td>Plants</td>
</tr>
<tr>
<td>Salazar, Jennifer</td>
<td>ORNL</td>
<td>Outreach</td>
</tr>
<tr>
<td>Schatz, Michael</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Seaver, Sam</td>
<td>ANL</td>
<td>Plants</td>
</tr>
<tr>
<td>Sharma, Manoj</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Sharma, Rita</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Sneddon, Michael</td>
<td>LBNL</td>
<td>Communities</td>
</tr>
<tr>
<td>Stevens, Rick</td>
<td>ANL</td>
<td>Management</td>
</tr>
<tr>
<td>Syed, Mustafa</td>
<td>ORNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Terry, Owenth</td>
<td>LBNL</td>
<td>Administration</td>
</tr>
<tr>
<td>Thomas, James</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Tingle, Nathan</td>
<td>Hope</td>
<td>Microbes</td>
</tr>
<tr>
<td>Trimbble, Will</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Tuskan, Jerry</td>
<td>ORNL</td>
<td>Plants</td>
</tr>
<tr>
<td>Wang, Daifeng</td>
<td>Yale</td>
<td>Plants</td>
</tr>
<tr>
<td>Ware, Doreen</td>
<td>CSHL</td>
<td>Plants</td>
</tr>
<tr>
<td>Weston, David</td>
<td>ORNL</td>
<td>Plants</td>
</tr>
<tr>
<td>Whitney, Cary</td>
<td>LBNL</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>Wilke, Andreas</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Wilkening, Jared</td>
<td>ANL</td>
<td>Communities</td>
</tr>
<tr>
<td>Xia, Fangfang</td>
<td>ANL</td>
<td>Microbes</td>
</tr>
<tr>
<td>Yoo, Shinjae</td>
<td>BNL</td>
<td>Plants</td>
</tr>
<tr>
<td>Yu, Dantong</td>
<td>BNL</td>
<td>Plants</td>
</tr>
</tbody>
</table>
Thank you!

Visit us at kbase.science.energy.gov
Contact us at outreach@kbase.us