

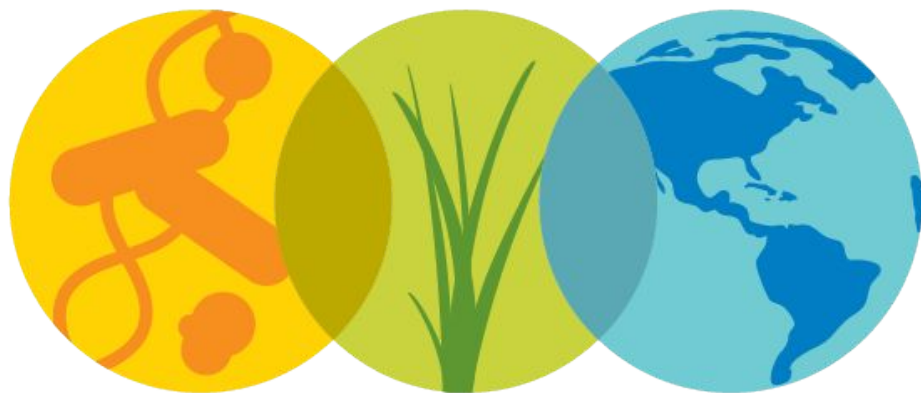


# **KBase Automation**

## **Group 26 - Senior Design**

Client/Advisor: Dr. Myra Cohen

Jake Veatch, Daulton Leach, Daniel Way, Sergey Gernega, Hunter Hall, Caleb Meyer



# KBase

PREDICTIVE BIOLOGY

## DOE Systems Biology Knowledgebase

<https://www.kbase.us/acknowledgements/>

# KBase Ecosystem

- **Web Application**
  - Non-technical (BIOLOGISTS)
  
- **Software Development Kit (SDK)**
  - Technical
  - Not the focus



# The Flux Balance Analysis (FBA)

The screenshot shows the KBase interface for the 'Run Flux Balance Analysis' workflow. The top navigation bar includes the KBase logo, the workflow name 'Gap&FBA 101 - V2020 - Copy', and user information 'Created by: Mikaela Cashman (mikaetacashman)'. The main interface is divided into several sections:

- DATA:** A list of input objects including 'Btheta.FBA.v3', 'Bob.v1', 'Bacteroides\_thetaiotaomicron\_V...', 'm127.media.v1', and 'base.media.v1'.
- APPS:** A list of application categories such as 'Comparative Genomics', 'Expression', 'Genome Annotation', 'Genome Assembly', 'Metabolic Modeling', 'Microbial Communities', 'Read Processing', 'Sequence Analysis', 'Uncategorized', 'Upload', 'Utilities', and 'Virus'.
- Run Flux Balance Analysis:** The main configuration area with a 'Run' button and sub-sections for 'Input Objects', 'FBA model', 'Media', 'Expression data set', and 'Parameters'.

The screenshot shows the KBase interface for the 'Run Flux Balance Analysis' workflow results. The top navigation bar is identical to the configuration page. The main interface is divided into several sections:

- DATA:** A list of input objects, identical to the configuration page.
- APPS:** A list of application categories, identical to the configuration page.
- Expression condition:** A section with a search bar and a message 'no available values'.
- Selected Items:** A section with a search bar and a message 'no selected values'.
- Expression threshold:** A numeric input field with the value '0'.
- Expression uncertainty:** A numeric input field with the value '0'.
- Activation coefficient:** A numeric input field with the value '1'.
- Max carbon uptake:** A numeric input field with the value '50'.
- Max nitrogen uptake:** A numeric input field with the value '50'.
- Max phosphate uptake:** A numeric input field with the value '50'.
- Max sulfur uptake:** A numeric input field with the value '50'.
- Max oxygen uptake:** A numeric input field with the value '50'.



# Problem

- Manual configuration
  - Which of these 100 solutions does organism X grow best?
- Take minutes to run
- Can't run tests sequentially (Web App)
- Exponential combination of inputs to test
  - 10 booleans =  $2^{10} = 1,024$  combinations
  - 8 floats with a range of 0-100 is infinite
- Why not use the SDK?
  - We are trying to mimic the user when testing
  - We want to keep it simple for the non-technical users

Expression threshold	<input type="text" value="0.5"/>
Expression uncertainty	<input type="text" value="0.5"/>
Activation coefficient	<input type="text" value="0.9141714"/>
Max carbon uptake	<input type="text" value="0"/>
Max nitrogen uptake	<input type="text" value="0"/>
Max phosphate uptake	<input type="text" value="0"/>
Max sulfur uptake	<input type="text" value="0"/>
Max oxygen uptake	<input type="text" value="0"/>



# Project Motivation

1. Primary
  - a. Automate programming experiments
  - b. Run experiments sequentially
  - c. Result - increase efficiency of biologists
  
2. Side Effect (Secondary)
  - a. Tool can be used to test KBase as system
  - b. Certain experiment configurations affect / break Kbase?



# Our Solution

- Take inputs from user either through a GUI or a file
- Automate browser functionality.
- Use a Job/Runner style to allow jobs to run sequentially
- Collect output from job(s) and return the output to the user in a file
- Output is viewable to catch errors and view configuration results



# Selenium

- Browser automation software
- Primary use is automation of web applications, but can be extended to any task that needs to be repeated numerous times automatically.

Perfect fit for running multiple Flux Based Analysis experiments (FBAs)





# Intended Users

1. Biologists
  - a. Allows biologists to view results from running many simulations to determine the desired outcome. Allows the Biologist to play with and configure the parameters to their likings. Can run X simulations at a time and try to find the preferred results.
2. KBase Developers
  - a. Projects output collection will flag errors and allow the developers to easily find and replicate bugs to allow for simpler debugging

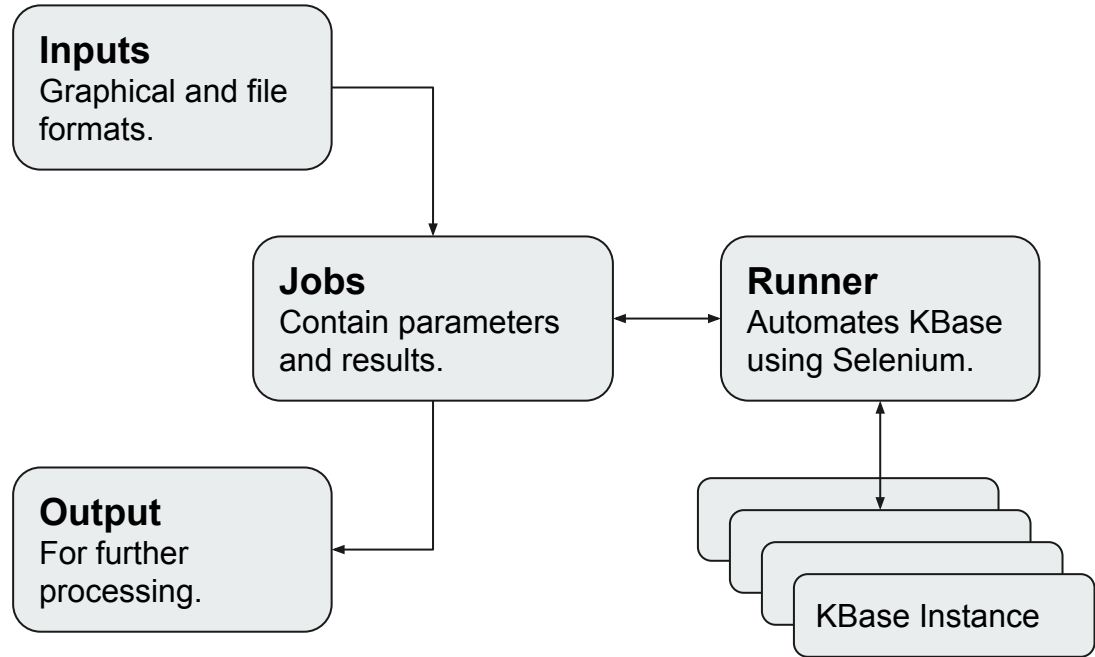
Realistically: Anyone with proper KBase credentials.



# Architecture

## Core objectives:

- Inputs, outputs, jobs, and runners should be modular
- System should be extensible for future formats and KBase application types





# Requirements

## Functional:

- Configure Selenium to interface with KBase apps
- User sets parameters to automate using Selenium
- App will sample the configuration space from set parameters
- Automate the testing of sample configurations
- Collect output data from KBase
- Users will be able to randomize inputs
- All inputs from the KBase GUI will be available

## Non-Functional:

- Will not noticeably disrupt KBase traffic flow
- App GUI will have similar layout as KBase GUI

## Environmental:

- Will require a valid and active KBase account
- Standalone application that can be executed from the users desktop
- Must be run on device which supports Selenium



# Standards Followed

IEEE 12207-2017 - ISO/IEC International Standard - Information Technology - Software Life Cycle Processes

This standard holds the purpose of defining a lifecycle standard for software projects from design, development, deployment, testing, and acquisition standpoints.

IEEE 1028-2008 - IEEE Standard for Software Reviews and Audits

This standard defines five different types of software reviews and audits that can be used in combination to increase the quality of software projects at various stages of their lifecycle.

IEEE 2430-2019 - IEEE Trial-Use Standard for Software Non-Functional Sizing Measurements

This standard defines how to size nonfunctional requirements in software.



# Engineering Constraints

- Have to use selenium to manipulate GUI, cannot use SDK.
- Unable to modify KBase in any way
- Future development teams must be able to modify project.



# Operational Environment

- Will require a valid and active KBase account
- Must be run on device which supports Selenium
- Standalone application that can be executed from the users desktop



# Security Concerns

- Less risk - the project is augmented current user capabilities
- User must provide authentication credentials

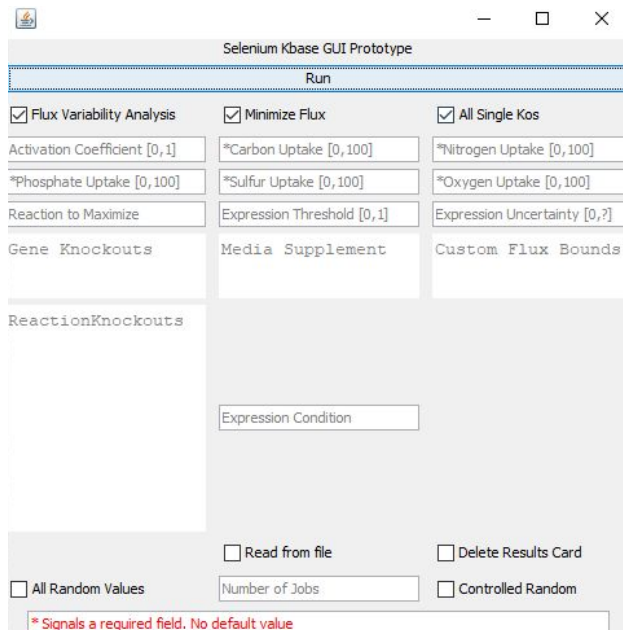


# Testing Plan

- Unit testing
- Continuous integration testing
- Acceptance testing
- Results



# Screenshots



Selenium Kbase GUI Prototype

Run

Flux Variability Analysis     Minimize Flux     All Single Kos

Activation Coefficient [0,1]    \*Carbon Uptake [0,100]    \*Nitrogen Uptake [0,100]

\*Phosphate Uptake [0,100]    \*Sulfur Uptake [0,100]    \*Oxygen Uptake [0,100]

Reaction to Maximize    Expression Threshold [0,1]    Expression Uncertainty [0,?]

Gene Knockouts    Media Supplement    Custom Flux Bounds

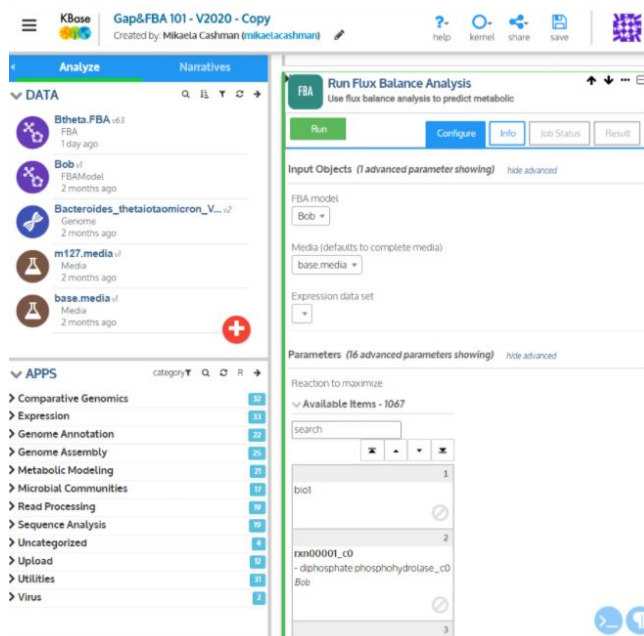
ReactionKnockouts

Expression Condition

Read from file     Delete Results Card

All Random Values    Number of Jobs     Controlled Random

\* Signals a required field. No default value



Kbase Gap&FBA 101 - V2020 - Copy  
Created by Mikaela Cashman (mikaelacashman)

Analyze Narratives

DATA

- Bheta.FBA v1.3 FBA 1 day ago
- Bob v1 FBAModel 2 months ago
- Bacteroides\_thetaiotaomicron\_V... v2 Genome 2 months ago
- m127.media v1 Media 2 months ago
- base.media v1 Media 2 months ago

APPS

- Comparative Genomics 30
- Expression 31
- Genome Annotation 23
- Genome Assembly 73
- Metabolic Modeling 21
- Microbial Communities 17
- Read Processing 39
- Sequence Analysis 39
- Uncategorized 4
- Upload 2
- Utilities 13
- Virus 2

Run Flux Balance Analysis  
Use flux balance analysis to predict metabolic

Run    Configure    Info    Job Status    Result

Input Objects (1 advanced parameter showing) hide advanced

FBA model  
Bob

Media (defaults to complete media)  
base.media

Expression data set

Parameters (16 advanced parameters showing) hide advanced

Reaction to maximize

Available Items - 1067

search

biol	1
rxn0001_c0 -diphosphate phosphohydrolyase_c0 Bob	2
	3



**DEMO**



# References

- *IEEE SA - The IEEE Standards Association - Home*, IEEE Standards Association, [standards.ieee.org/](https://standards.ieee.org/).
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*. 2018;36: 566. doi: 10.1038/nbt.4163



# Questions?

Thank you for your time.