THE AGAVE PLATFORM

SCIENCE AS A SERVICE FOR THE OPEN SCIENCE COMMUNITY

Rion Dooley @*deardooley*
deardooley@gmail.com

THE EVOLUTION OF A CYBERINFRASTRUCTURE

Once upon a time, most of us built garage-style clusters…

HPC systems have grown up since then and become much more powerful and sophisticated

Modern academic data centers were built upon a foundation of:

- *HPC*
- *Visualization*
- *Support*

To support cutting edge computational science.

But an unexpected thing happened...


BIG
DATA

WORKFLOWS NOW TECHNICALLY COMPLICATED

LANGUAGES

- Python 2 & 3
- R
- Julia
- Perl
- Matlab
- Java
- Scala, Clojure, etc
- .NET
- C++
- Swift
- Haskell
- Go
- Javascript

FRAMEWORKS

- MapReduce Hadoop, Storm, Pachyderm
- Event & Streaming: Kinesis, Azure Stream Analytics, Camel, Streambase
- Deep/Machine Learning: Watson, Azure BI, Tensorflow
- In-memory parsing: Kognito, Apache Spark
- New data warehouse: Snowflake
- Containers: Docker, Rocket, MESOS, Kubernetes
- Cloud: AWS, GCE, OpenStack, VMWare

HARDWARE

- Rise of many-core computing means 50-100 threads/node*
  - Xeon / Xeon Phi
  - GPU
  - OpenPower
  - ARM
- Multi-level memory architectures
- Hierarchical storage architectures
- FPGAs

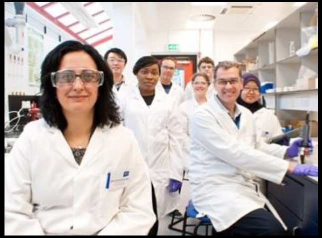
DIVERSE DISTRIBUTED RESEARCH TEAMS

Mike
- Computing novice
- Works remotely at partner site

Eliza
- Masters specific analysis skills
- Readily adopts new tech

Paulo
- Staff computational expert
- Supports multiple projects

Nikolaidas Group
- Mostly experimentalists
- Strict data sharing & access

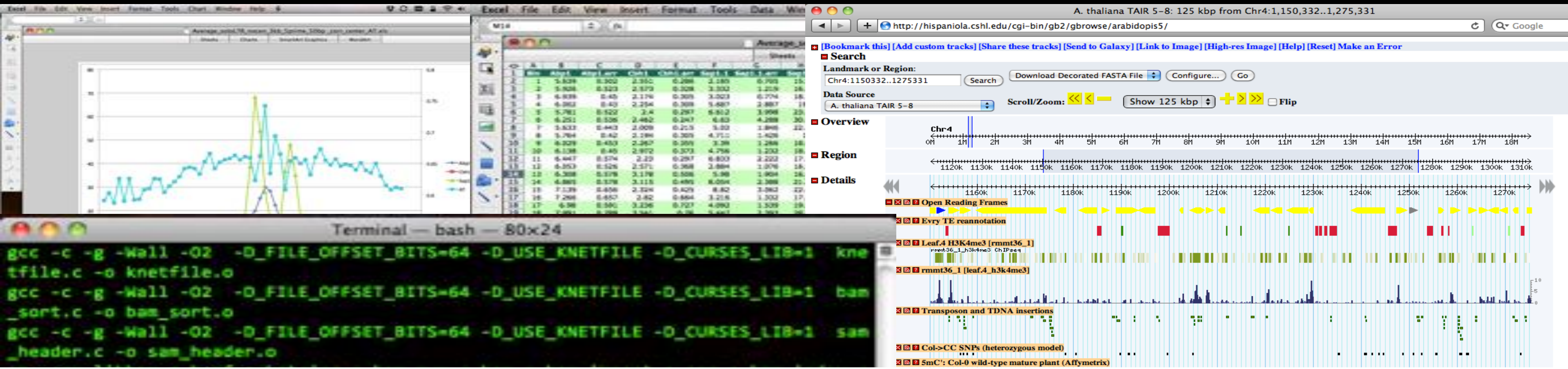


Roshan
- Computationally experienced
- Focused on interpretation

HOW DO WE HELP RESEARCHERS WITH SUCH DIVERSE NEEDS AND BACKGROUNDS?

Why Agave Was Built

- We used to build big HPC systems an pat ourselves on the back. But the world changed and we had to as well More iron wasn't the answer
- We started building software. We realized we couldn't build bespoke solutions for everyone
- We tried building portals. That would never scale.
- We tried enhancing an providing notebooks an "environments" online. Those were popular, but people outgrew those solutions.
- We tried giving them VM with entire stacks on them. That went well, but people needed to leverage things other than VM.
- What conversations always led to was that they wanted the kind of access our in-house devs had. They wanted cheat codes.
- Now we build APIs so they could level up their science. Those APIs are what we now call Agave



A. thaliana TAIR 5-8: 125 kbp from Chr4:1,150,332..1,275,331

http://hispaniola.cshl.edu/cgi-bin/gb2/gbrowse/arabidopis5/

Landmark or Region: Chr4:1150332..1275331

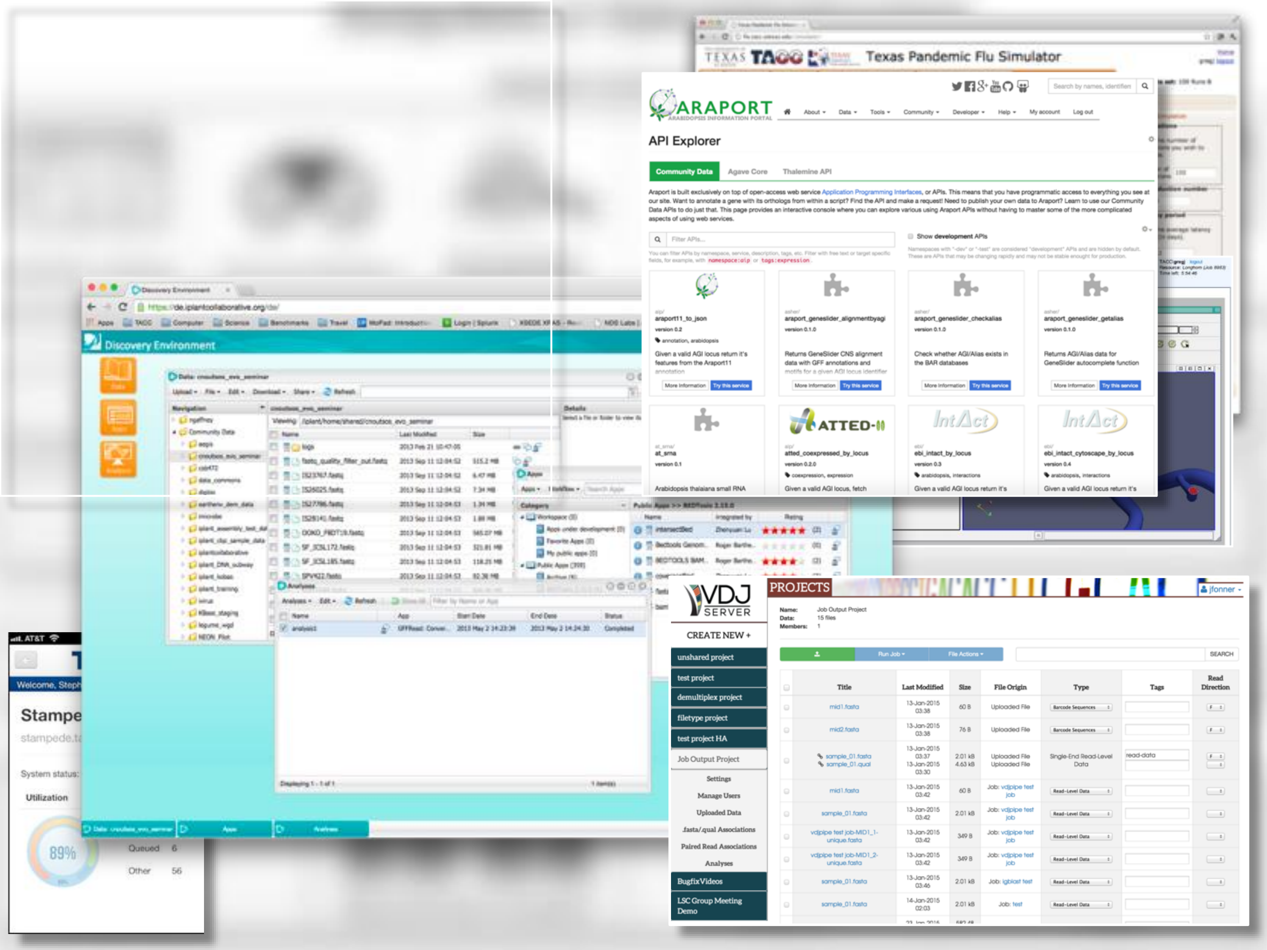
Data Source: A. thaliana TAIR 5-8

Scroll/Zoom: Show 125 kbp

```
gcc -c -g -Wall -O2  -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1  kne
tfile.c -o knetfile.o
gcc -c -g -Wall -O2  -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1  bam
_sort.c -o bam_sort.o
gcc -c -g -Wall -O2  -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1  sam
_header.c -o sam_header.o
```

MANY DOMAIN SCIENTISTS ARE NOT EXPERTS AT COMPUTING TECHNOLOGY. CREATE PURPOSE-BUILT, HIGHLY INTUITIVE INTERFACES

```
KNETFILE -D_CURSES_LIB=1  md5
```



Point-and-click interfaces

- Data management, sharing, and analysis
- Publishing reproducible analysis workflows
- Discovery of new or updated tools and data
- Interactive visualization of results

Backed by world-class computing and data capacity

GIVE EXPERTS ACCESS TO EVERY SINGLE ONE OF YOUR BUILDING BLOCKS. WEB SERVICE APIs EVERYWHERE. AUGMENT WITH PROFESSIONAL TOOLING.


Agave
Platform

AGAVE IS A MULTI-TENANT PAAS

DELIVERING **SCIENCE-AS-A-SERVICE** SOLUTIONS

IN *HYBRID* COMPUTATIONAL ENVIRONMENTS

-

BRIDGE BUILDING

OPEN SOURCE WITH FLEXIBLE DEPLOYMENT

Cloud Deployment

- Use your existing IdP
- White-label docs & tools
- Ansible for automation
- Pluggable logging
- Pluggable monitoring
- Managed or self-hosted
- Consulting available

On-Premise Deployment

- Use your existing IdP
- White-label docs & tools
- Ansible for automation
- Pluggable logging
- Pluggable monitoring
- Managed or self-hosted
- Consulting available

Hybrid Deployment

- Use your existing IdP
- White-label docs & tools
- Ansible for automation
- Pluggable logging
- Pluggable monitoring
- Managed or self-hosted
- Consulting available

WHAT DOES IT DO?

MANAGE DATA

RUN CODE

COLLABORATE ANYWHERE

CONNECT ANYTHING

AGAVE HELPS YOU MANAGE DATA

- Single, consistent interface to access distributed data
  - Agave "files" commands work on object store as well as linux clusters
- Managed, tenacious data movement
  - Will retry on failure
- Opinion-free metadata management
  - Name/value pairs
- Full provenance and searchable audit trail.
- Events, alerts, and notifications
- Horizontal scaling

SFTP

iRODS

globus online

Local

Amazon S3

SWIFT
an OpenStack Community Project

Microsoft Azure
Blob Storage

Google Drive

Dropbox

box

AGAVE HELPS YOU RUN CODE

- Bring your own code and/or leverage our catalog
- Run your apps as interactive, batch, or event driven processes
- Full lifecycle management
- Full provenance and searchable audit trail
- Reproducibility as a feature
- Publish entire experimental runs

slurm
workload manager

IBM

Adaptive
COMPUTING

PBS Works™

GRID ENGINE

HTCondor
High Throughput Computing

exec

docker

AGAVE HELPS YOU COLLABORATE MEANINFULLY

- Fine-grained access control
- Deep link to any resource in the API
- Pre-authenticated disposable links
- Alerts and notifications into existing apps
- Web standards come standard
- Integrations with popular frameworks and cloud services

slack
WebSockets
Email
Drupal
laravel
WordPress
express
django
Flask

AGAVE HELPS YOU INTEGRATE ANYWHERE

- Events, crons, monitors, and polling to stay informed
- Webhooks, web sockets, and custom vendor integrations to integrate with the world around you
- OAuth2, OIDC, and service accounts
- Self-service API Management
- Data mediation as a service
  - Can transform data formats
  - Access DBs, spreadsheets, etc.
- Support for enterprise IoT solutions

AWS Lambda

Pushbullet

zapier

IFTTT
ifthisthenthat

Google Home

nest

alexa

GROCERY STORE APPROACH

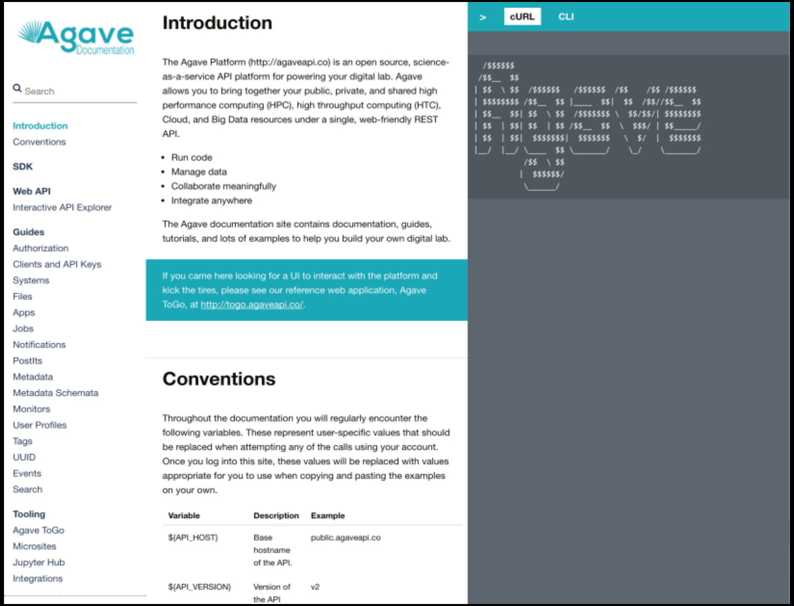
- Take what you want
- Leave the rest
- Self-paced
- No lock-in

Agave
Tooling

TOOLS THAT FIT YOUR TOOLBELT

- Client SDK: Python, JavaScript, Java, PHP, *Perl*, *R*
- Command Line Interface
- Plugins: AngularJS, Wordpress, Drupal, Tomcat
- Web applications (ToGo)
- Integrated environments (Jupyter Hub)
- Workflow management (End of Day)

DEVELOPER FRIENDLY DOCUMENTATION



Introduction

The Agave Platform (http://agaveapi.co) is an open source, science-as-a-service API platform for powering your digital lab. Agave allows you to bring together your public, private, and shared high performance computing (HPC), high throughput computing (HTC), Cloud, and Big Data resources under a single, web-friendly REST API.

- Run code
- Manage data
- Collaborate meaningfully
- Integrate anywhere

The Agave documentation site contains documentation, guides, tutorials, and lots of examples to help you build your own digital lab.

If you came here looking for a UI to interact with the platform and kick the tires, please see our reference web application, Agave ToGo, at http://togo.agaveapi.co/.

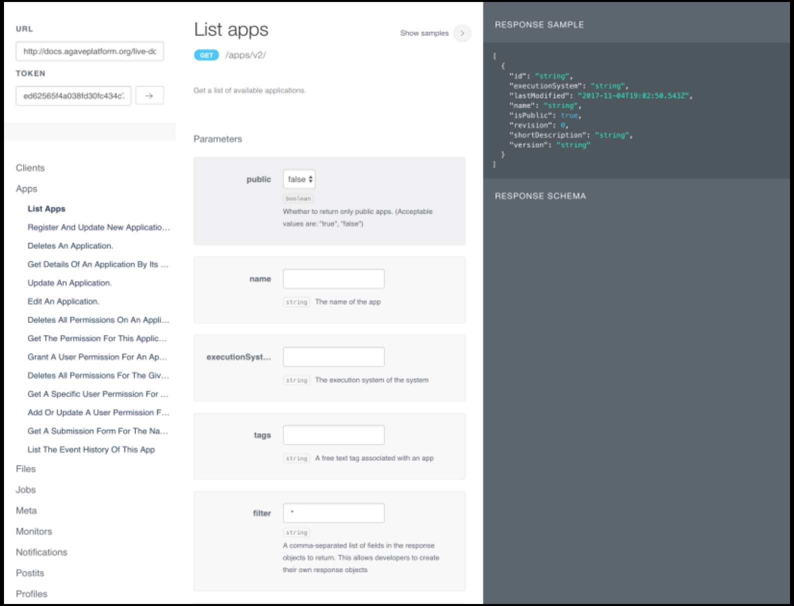
Conventions

Throughout the documentation you will regularly encounter the following variables. These represent user-specific values that should be replaced when attempting any of the calls using your account. Once you log into this site, these values will be replaced with values appropriate for you to use when copying and pasting the examples on your own.

Variable	Description	Example
${API_HOST}	Base hostname of the API.	public.agaveapi.co
${API_VERSION}	Version of the API	v2

```
  /$$$$$$
 /$$__  $$
| $$  \ $$  /$$$$$$   /$$$$$$  /$$    /$$ /$$$$$$
| $$$$$$$$ /$$__  $$ |____  $$|  $$  /$$//$$__  $$
| $$__  $$| $$  \ $$  /$$$$$$$ \  $$/$$/| $$$$$$$$
| $$  | $$| $$  | $$ /$$__  $$  \  $$$/ | $$_____/
| $$  | $$|  $$$$$$$|  $$$$$$$   \  $/  |  $$$$$$$
|__/  |__/ \____  $$ \_______/    \_/    \_______/
           /$$  \ $$
          |  $$$$$$/
           \______/
```

Developer guides



URL http://docs.agaveplatform.org/live-dc

TOKEN ed62565f4a038fd30fc434c

List apps

GET /apps/v2/

Get a list of available applications.

Parameters

public false
boolean
Whether to return only public apps. (Acceptable values are: "true", "false")

name
string The name of the app

executionSyst…
string The execution system of the system

tags
string A free text tag associated with an app

filter *
string
A comma-separated list of fields in the response objects to return. This allows developers to create their own response objects

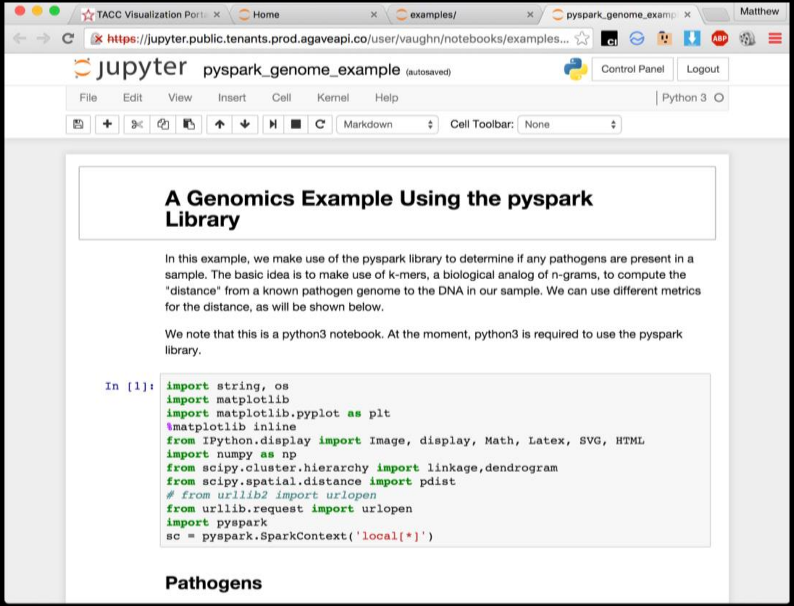
RESPONSE SAMPLE

```
[
  {
    "id": "string",
    "executionSystem": "string",
    "lastModified": "2017-11-04T19:02:50.543Z",
    "name": "string",
    "isPublic": true,
    "revision": 0,
    "shortDescription": "string",
    "version": "string"
  }
]
```

RESPONSE SCHEMA

Interactive API browser

COPY AND PASTE OR GIT CLONE



A Genomics Example Using the pyspark Library

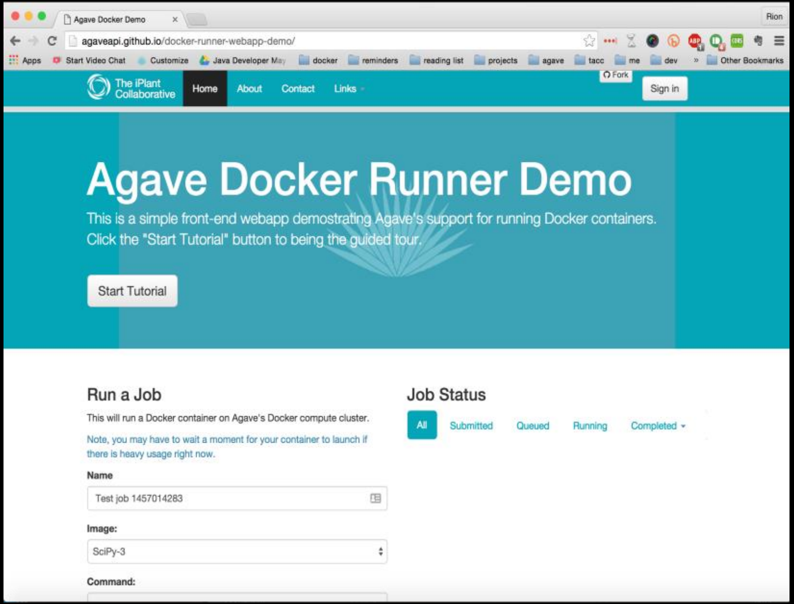
In this example, we make use of the pyspark library to determine if any pathogens are present in a sample. The basic idea is to make use of k-mers, a biological analog of n-grams, to compute the "distance" from a known pathogen genome to the DNA in our sample. We can use different metrics for the distance, as will be shown below.

We note that this is a python3 notebook. At the moment, python3 is required to use the pyspark library.

```
In [1]: import string, os
        import matplotlib
        import matplotlib.pyplot as plt
        %matplotlib inline
        from IPython.display import Image, display, Math, Latex, SVG, HTML
        import numpy as np
        from scipy.cluster.hierarchy import linkage,dendrogram
        from scipy.spatial.distance import pdist
        # from urllib2 import urlopen
        from urllib.request import urlopen
        import pyspark
        sc = pyspark.SparkContext('local[*]')
```

Pathogens

Jupyter integration



Agave Docker Runner Demo

This is a simple front-end webapp demostrating Agave's support for running Docker containers. Click the "Start Tutorial" button to being the guided tour.

Run a Job

This will run a Docker container on Agave's Docker compute cluster.

Note, you may have to wait a moment for your container to launch if there is heavy usage right now.

Name: Test job 1457014283

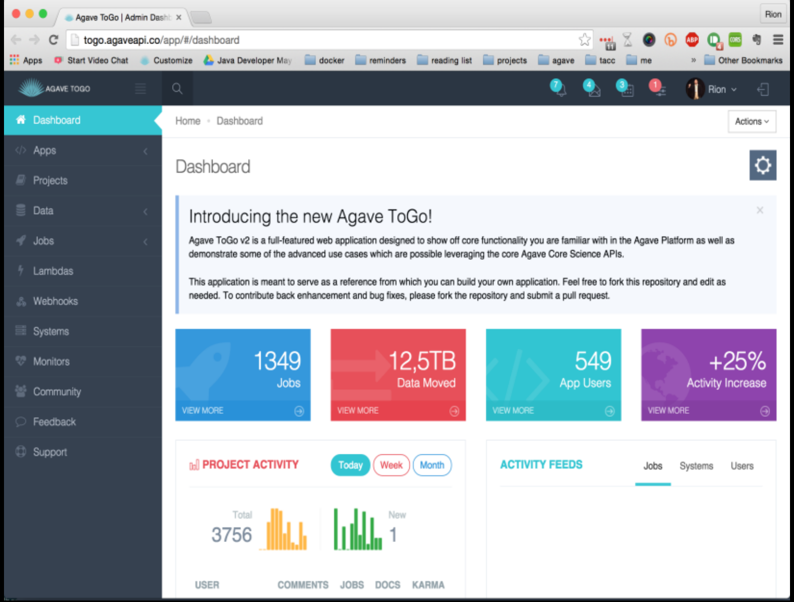
Image: SciPy-3

Command:

Job Status

Demos & samples

WEBAPPS *TOSTUDY* AND *TOGO*



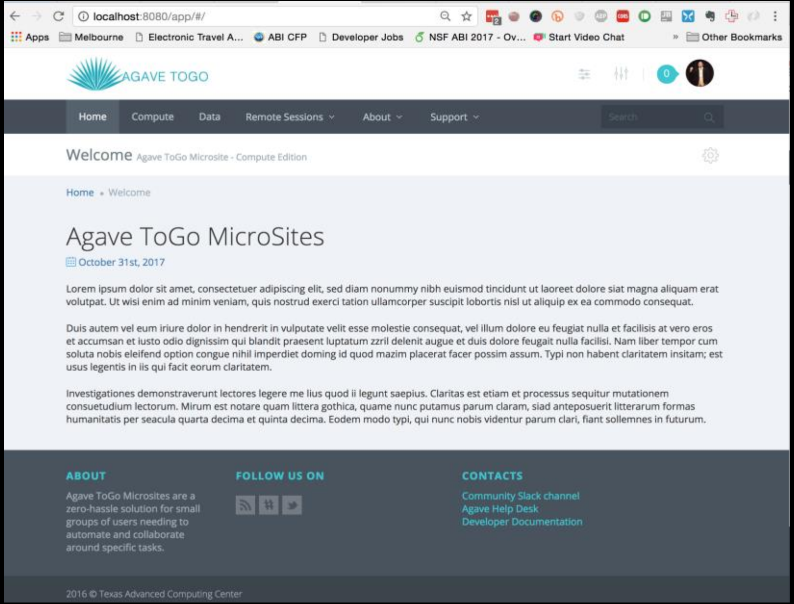
Dashboard

Introducing the new Agave ToGo!

Agave ToGo v2 is a full-featured web application designed to show off core functionality you are familiar with in the Agave Platform as well as demonstrate some of the advanced use cases which are possible leveraging the core Agave Core Science APIs.

This application is meant to serve as a reference from which you can build your own application. Feel free to fork this repository and edit as needed. To contribute back enhancement and bug fixes, please fork the repository and submit a pull request.

http://togo.agaveplatform.org/app



Welcome Agave ToGo Microsite - Compute Edition

Agave ToGo MicroSites

October 31st, 2017

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore siat magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla et facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue et duis dolore feugait nulla facilisi. Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Typi non habent claritatem insitam; est usus legentis in iis qui facit eorum claritatem.

Investigationes demonstraverunt lectores legere me lius quod ii legunt saepius. Claritas est etiam et processus sequitur mutationem consuetudium lectorum. Mirum est notare quam littera gothica, quame nunc putamus parum claram, siad anteposuerit litterarum formas humanitatis per seacula quarta decima et quinta decima. Eodem modo typi, qui nunc nobis videntur parum clari, fiant sollemnes in futurum.

ABOUT

Agave ToGo Microsites are a zero-hassle solution for small groups of users needing to automate and collaborate around specific tasks.

FOLLOW US ON

CONTACTS

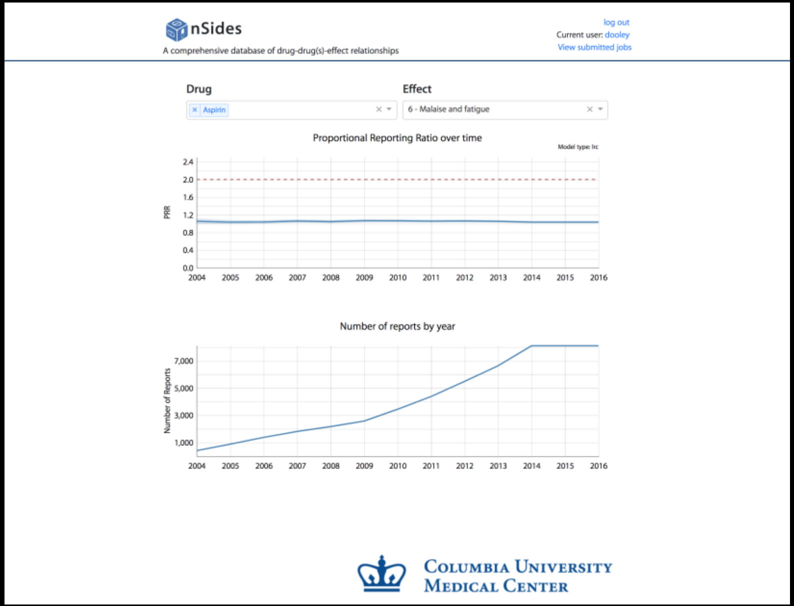
Community Slack channel
Agave Help Desk
Developer Documentation

2016 © Texas Advanced Computing Center

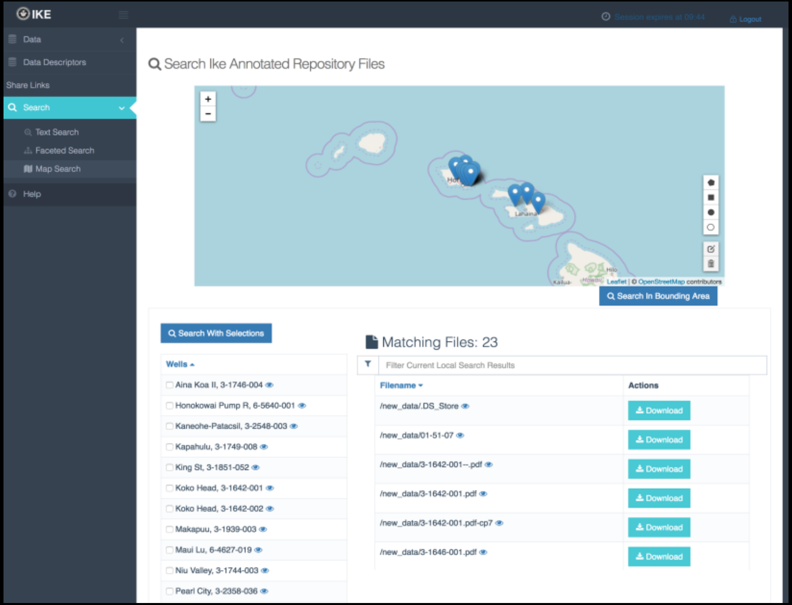
http://agaveplatform.github.io/microsites/app


WORLDWIDE PLATFORM USAGE
Google
Map data ©2016
Terms of Use

WHO'S USING THIS STUFF?



http://nsides.io



Search Ike Annotated Repository Files

Matching Files: 23

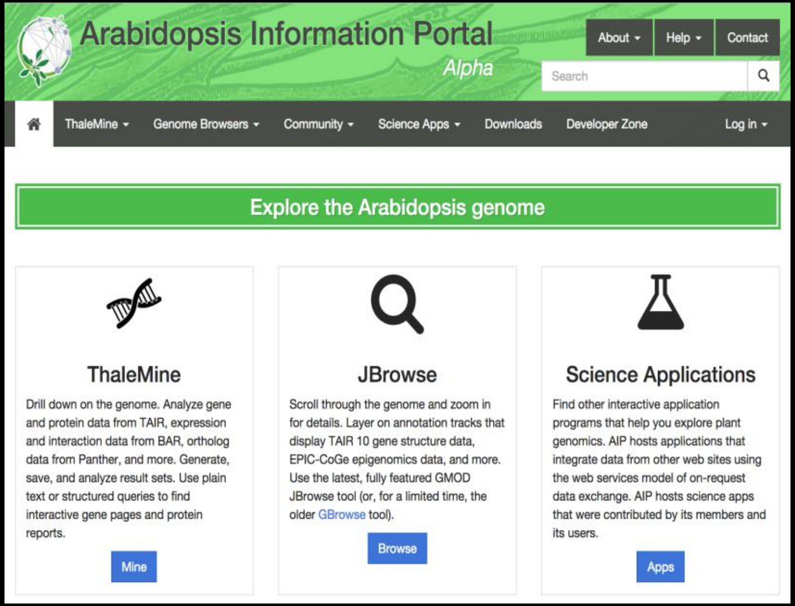
Wells ▴

- ☐ Aina Koa II, 3-1746-004
- ☐ Honokowai Pump R, 6-5640-001
- ☐ Kaneohe-Patacsil, 3-2548-003
- ☐ Kapahulu, 3-1749-008
- ☐ King St, 3-1851-052
- ☐ Koko Head, 3-1642-001
- ☐ Koko Head, 3-1642-002
- ☐ Makapuu, 3-1939-003
- ☐ Maui Lu, 6-4627-019
- ☐ Niu Valley, 3-1744-003
- ☐ Pearl City, 3-2358-036

Filename ▾	Actions
/new_data/.DS_Store	Download
/new_data/01-51-07	Download
/new_data/3-1642-001--.pdf	Download
/new_data/3-1642-001.pdf	Download
/new_data/3-1642-001.pdf-cp7	Download
/new_data/3-1646-001.pdf	Download

https://ikewai.its.uhawaii.edu

WHO'S USING THIS STUFF?



Explore the Arabidopsis genome

ThaleMine

Drill down on the genome. Analyze gene and protein data from TAIR, expression and interaction data from BAR, ortholog data from Panther, and more. Generate, save, and analyze result sets. Use plain text or structured queries to find interactive gene pages and protein reports.

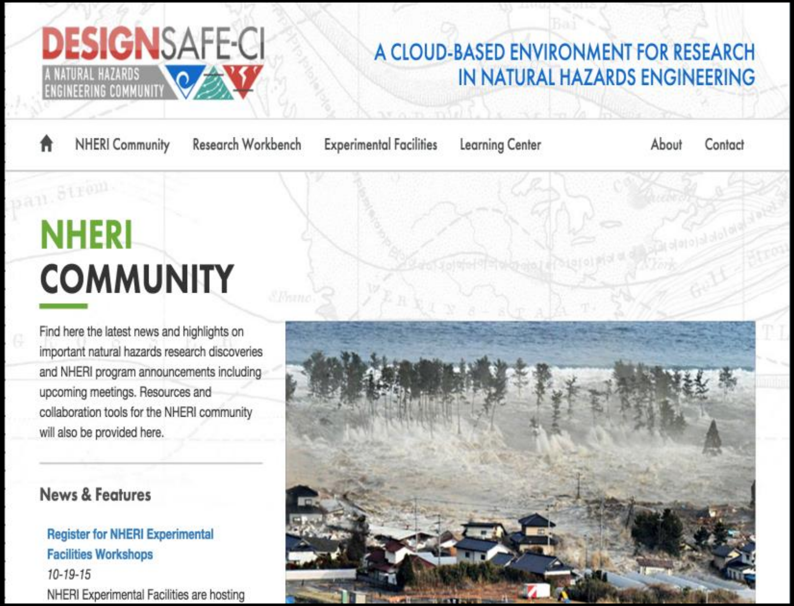
JBrowse

Scroll through the genome and zoom in for details. Layer on annotation tracks that display TAIR 10 gene structure data, EPIC-CoGe epigenomics data, and more. Use the latest, fully featured GMOD JBrowse tool (or, for a limited time, the older GBrowse tool).

Science Applications

Find other interactive application programs that help you explore plant genomics. AIP hosts applications that integrate data from other web sites using the web services model of on-request data exchange. AIP hosts science apps that were contributed by its members and its users.

http://araport.org



NHERI COMMUNITY

Find here the latest news and highlights on important natural hazards research discoveries and NHERI program announcements including upcoming meetings. Resources and collaboration tools for the NHERI community will also be provided here.

News & Features

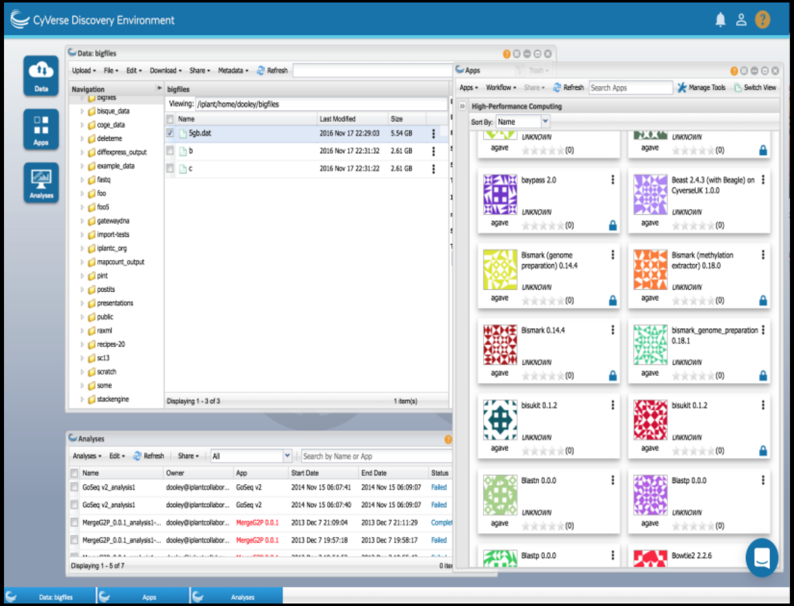
Register for NHERI Experimental Facilities Workshops

10-19-15

NHERI Experimental Facilities are hosting

https://www.designsafe-ci.org

WHO'S USING THIS STUFF?



bigfiles

Viewing: /iplant/home/dooley/bigfiles

	Name	Last Modified	Size
☑	5gb.dat	2016 Nov 17 22:29:03	5.54 GB
☐	b	2016 Nov 17 22:31:32	2.61 GB
☐	c	2016 Nov 17 22:31:22	2.61 GB

Displaying 1 - 3 of 3 1 item(s)

Analyses

Analyses Edit Refresh Share All Search by Name or App

	Name	Owner	App	Start Date	End Date	Status
☐	GoSeq v2_analysis1	dooley@iplantcollabor...	GoSeq v2	2014 Nov 15 06:07:41	2014 Nov 15 06:09:07	Failed
☐	GoSeq v2_analysis1	dooley@iplantcollabor...	GoSeq v2	2014 Nov 15 06:07:40	2014 Nov 15 06:09:07	Failed
☐	MergeG2P_0.0.1_analysis1-...	dooley@iplantcollabor...	MergeG2P 0.0.1	2013 Dec 7 21:09:04	2013 Dec 7 21:11:29	Complet
☐	MergeG2P_0.0.1_analysis1-...	dooley@iplantcollabor...	MergeG2P 0.0.1	2013 Dec 7 19:57:18	2013 Dec 7 19:58:17	Failed

Displaying 1 - 5 of 7

http://de.cyverse.org

WELCOME!

VDJServer is a free, scalable resource for performing immune repertoire analysis and sharing data. Manage, analyze, and archive your data through our web resource. You can also download our open source analysis software for local use.

UPLOAD

Manage and archive your data leveraging our mirrored, secure, high performance datastore.

ANALYZE

Use the latest software running on powerful supercomputers to crunch data at any scale.

PUBLISH

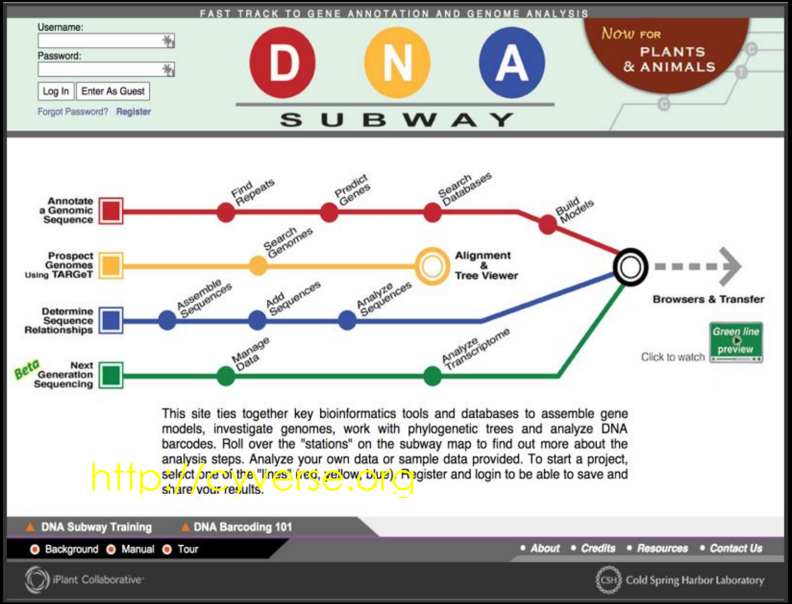
Create permanent URLs to your data, analysis pipelines, and results to support journal publications.

SHARE

Collaborate privately with individuals or contribute your data publicly to the scientific community.

https://vdjserver.org

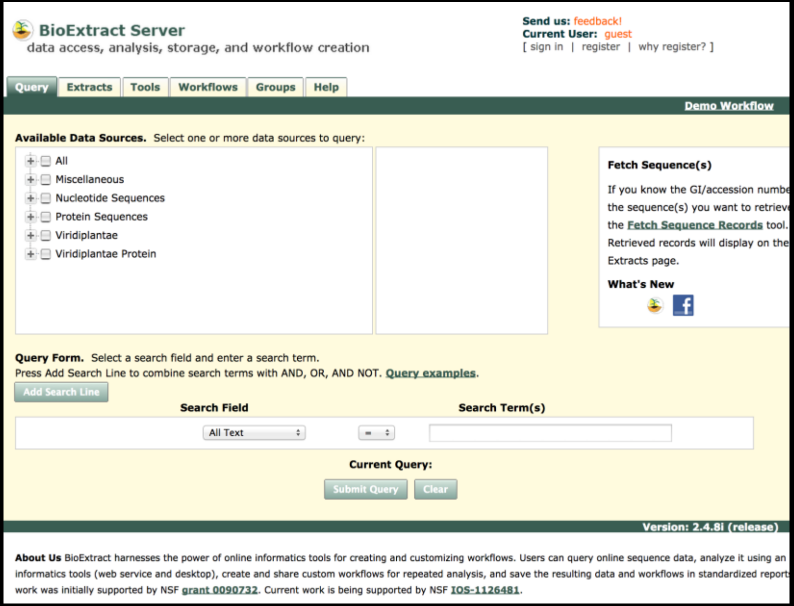
WHO'S USING THIS STUFF?



This site ties together key bioinformatics tools and databases to assemble gene models, investigate genomes, work with phylogenetic trees and analyze DNA barcodes. Roll over the "stations" on the subway map to find out more about the analysis steps. Analyze your own data or sample data provided. To start a project, select one of the "lines" (red, yellow, blue). Register and login to be able to save and share your results.

http://cyverse.org

http://dnasubway.org



Send us: feedback!
Current User: guest
[sign in | register | why register?]

Available Data Sources. Select one or more data sources to query:

- ☐ All
- ☐ Miscellaneous
- ☐ Nucleotide Sequences
- ☐ Protein Sequences
- ☐ Viridiplantae
- ☐ Viridiplantae Protein

Fetch Sequence(s)

If you know the GI/accession numbe
the sequence(s) you want to retrieve
the Fetch Sequence Records tool.
Retrieved records will display on the
Extracts page.

What's New

Query Form. Select a search field and enter a search term.
Press Add Search Line to combine search terms with AND, OR, AND NOT. Query examples.

Search Field | **Search Term(s)**

Current Query:

Version: 2.4.8i (release)

About Us BioExtract harnesses the power of online informatics tools for creating and customizing workflows. Users can query online sequence data, analyze it using an
informatics tools (web service and desktop), create and share custom workflows for repeated analysis, and save the resulting data and workflows in standardized report
work was initially supported by NSF grant 0090732. Current work is being supported by NSF IOS-1126481.

https://www.bioextract.org

WHO'S USING THIS STUFF?

What is iReceptor?

iReceptor is a distributed data management system and scientific gateway for mining "Next Generation" sequence data from immune responses. The goal of the project is to: *improve the design of vaccines and therapeutic antibodies by integrating Canadian and international data repositories of antibody and T-cell receptor gene sequences.*

iReceptor will provide a technology platform that will lower the barrier to immune genetics researchers who need to federate large, distributed, immune genetics databases in order to answer complex questions about the immune response. The focus of the iReceptor project is to leverage existing capabilities and technologies to build a new scientific platform for the immune genetics research community. In order to provide such a system, the iReceptor project will develop the following components:

- An iReceptor immune genetics data model and database design that enhances existing genomics data models,
- A set of data adaptor/import/conversion services for transforming immune genetics data sets into the iReceptor data model,
- An iReceptor data service that will expose access to a set of distributed immune genetics

LATEST NEWS

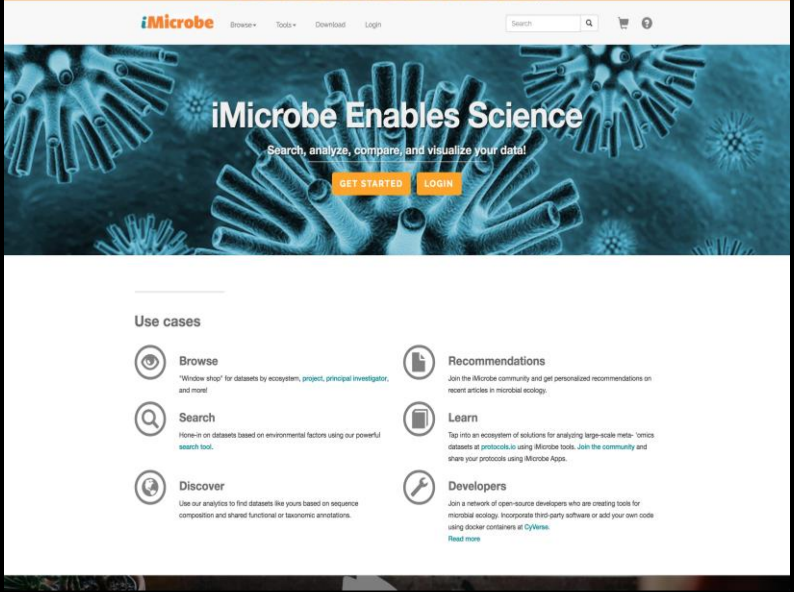
CANARIE Workshop Oct 20-21, 2014
10/20/2014 - 06:59

Antibody & T-Cell Receptor Data Integration Planning Meeting
09/10/2014 - 15:07

iReceptor CANARIE announcement
06/23/2014 - 11:31

iReceptor is live!!!
06/22/2014 - 00:44

https://ireceptorgw.irmacs.sfu.ca/



Note: This is a new beta version and still currently under development. Please let us know of any issues or suggestions!

iMicrobe Enables Science

Search, analyze, compare, and visualize your data!

Use cases

Browse
"Window shop" for datasets by ecosystem, project, principal investigator, and more!

Recommendations
Join the iMicrobe community and get personalized recommendations on recent articles in microbial ecology.

Search
Hone-in on datasets based on environmental factors using our powerful search tool.

Learn
Tap into an ecosystem of solutions for analyzing large-scale meta- 'omics datasets at protocols.io using iMicrobe tools. Join the community and share your protocols using iMicrobe Apps.

Discover
Use our analytics to find datasets like yours based on sequence composition and shared functional or taxonomic annotations.

Developers
Join a network of open-source developers who are creating tools for microbial ecology. Incorporate third-party software or add your own code using docker containers at CyVerse.
Read more

https://www.imicrobe.us


WHAT'S YOUR
POINT?


DID WE JUST BECOME BEST FRIENDS?

Agave Platform

https://agaveapi.co/

THANK YOU!

FOLLOW US

@agaveapi
https://agaveapi.co
https://slackin.agaveapi.co