



# ILCI Breeding Analytics Hub

**Workshop at the 2024 Annual Meeting  
San José, Costa Rica**

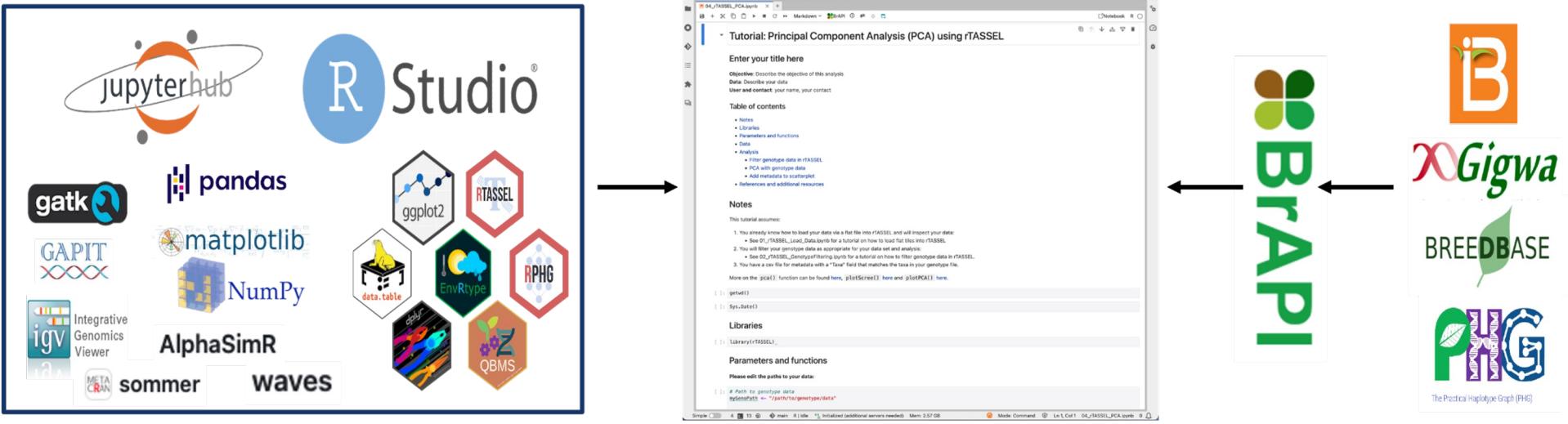


Cornell University

# ILCI Breeding Analytics Hub

## What is it?

The Breeding Analytics Hub is a cloud-based environment where software tools, code, documentation, and data are available in one place.



# What is it?

The hub is a managed environment:

- Software, tools, and packages are curated, installed and kept up to date.
- This includes custom R packages and web services developed specifically for ILCI researchers.

# What is it?

Computation is done on the cloud:

- Users aren't limited by the capacity of their computer or local server.
- Directories, data, and code are easily shared with other Hub users.

# What is it?

The Hub makes use of computational notebooks:

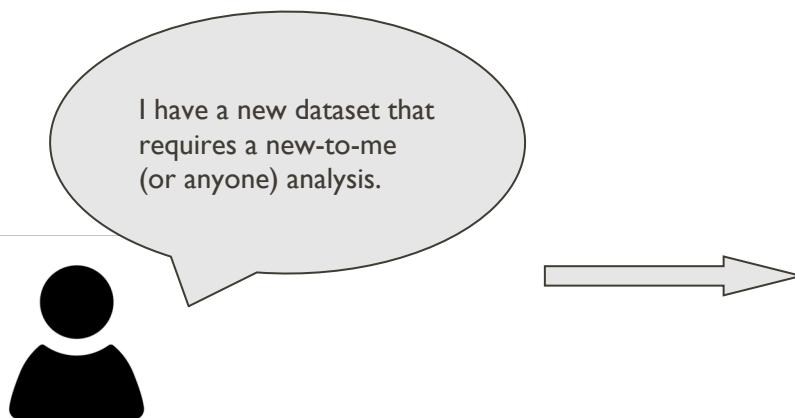
- Code and documentation are recorded in one place.
- Create reproducible data analysis projects.

# What is it?

The Hub is a shared space for collaboration, teaching and learning:

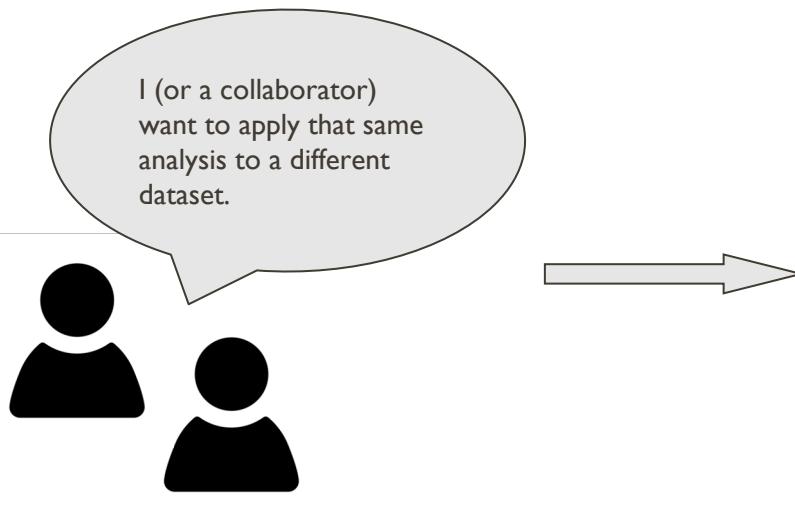
- Everyone is on the same server and in the same environment.
- Pass and share notebooks and data across users and institutions.
- Use templates, pre-populated notebooks that serve as tutorials and writeable notebooks.

# ILCI Breeding Analytics Hub Process



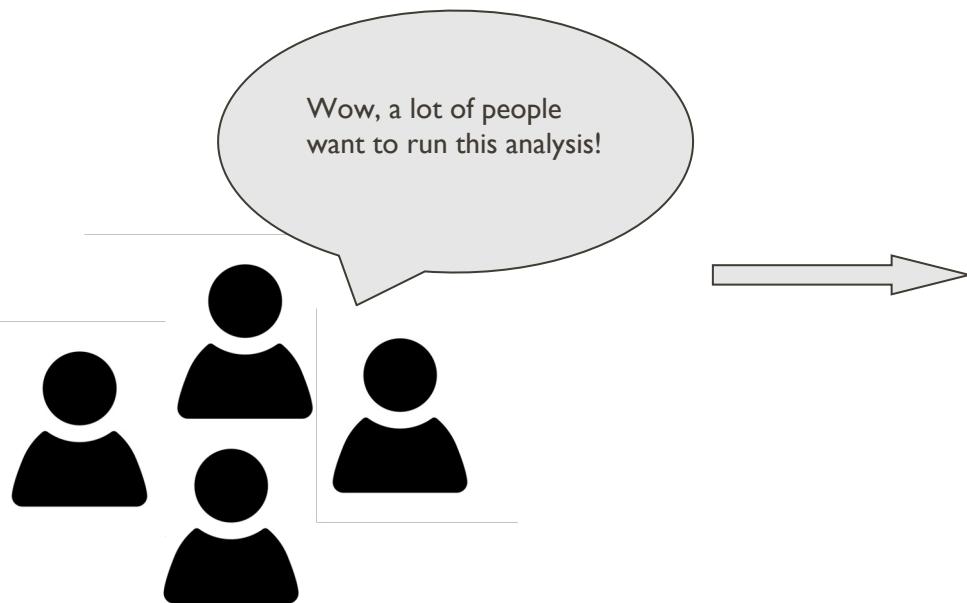
1. Upload data to the hub,
2. Create a documented notebook (independently or with a collaborator) with code,
3. Execute analysis.

# ILCI Breeding Analytics Hub Process



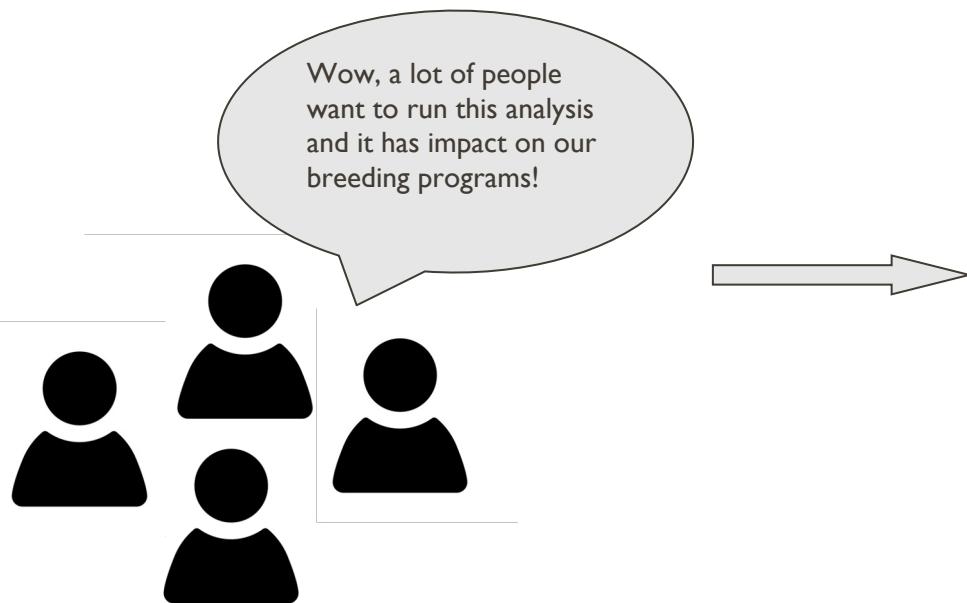
1. Upload data to the hub,
2. Re-use or share pre-written notebook, make changes to code as needed.
3. Execute analysis.
4. Evaluate and improve notebook.

# ILCI Breeding Analytics Hub Process



1. Upload data to a database,
2. Create a reproducible template that works with data retrieval,
3. Execute analysis.
4. Evaluate and improve template.

# ILCI Breeding Analytics Hub Process

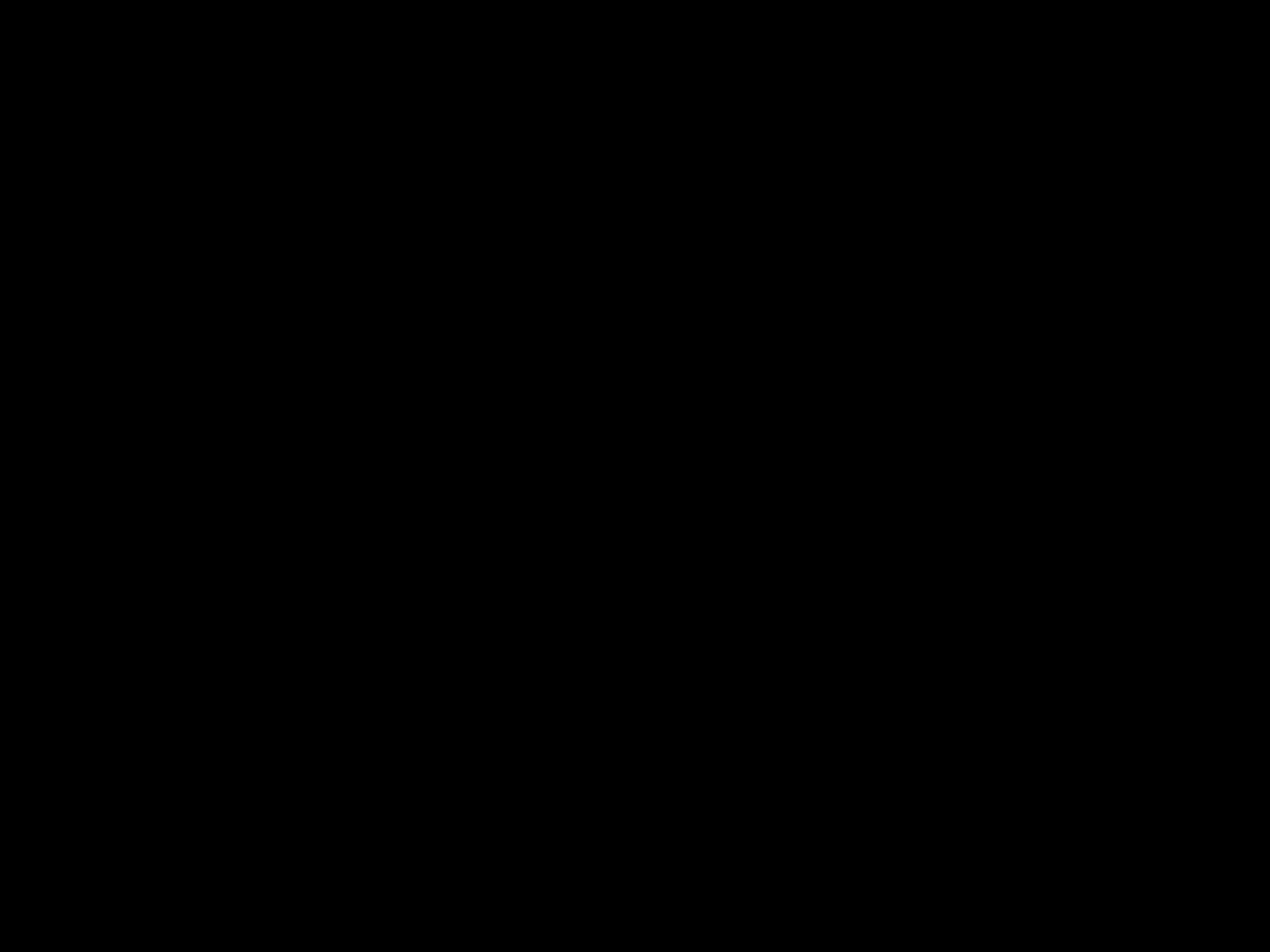


1. Create a new R package that simplifies and optimizes code or a webapp that automates the analysis.
2. Upload data to a database,
3. Execute analysis.
4. Evaluate and improve.

# ILCI Breeding Analytics Hub

## Featured tools

### Basic infrastructure



# ILCI Breeding Analytics Hub

## Featured tools

### Pre-populated computational notebooks

- Template tool

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
CCL_ESA	last year
eCornell	16 days ago
jobs	7 months ago
Markdown	17 days ago
PastAnalyses	last year
R	7 months ago
R.disabled	last year
shared	last year
shared_notebooks	2 years ago
sorghumDartag_CACCIA	3 months ago
templates_git	2 days ago
test_data	2 months ago
OrderAppendix_2_DS21-6352.zip	last year

Launcher

Notebook

- Python 3 (pykernel)
- Template
- Kotlin
- R
- RStudio [x]
- Shiny [x]

Console

- Python 3 (pykernel)
- Kotlin
- R

Other

- Terminal
- Text File
- Markdown File
- Python File
- R File
- Notebook Jobs

Show Contextual Help

Breeder Tools

Open the BrAPI Helper Panel

Jupyter

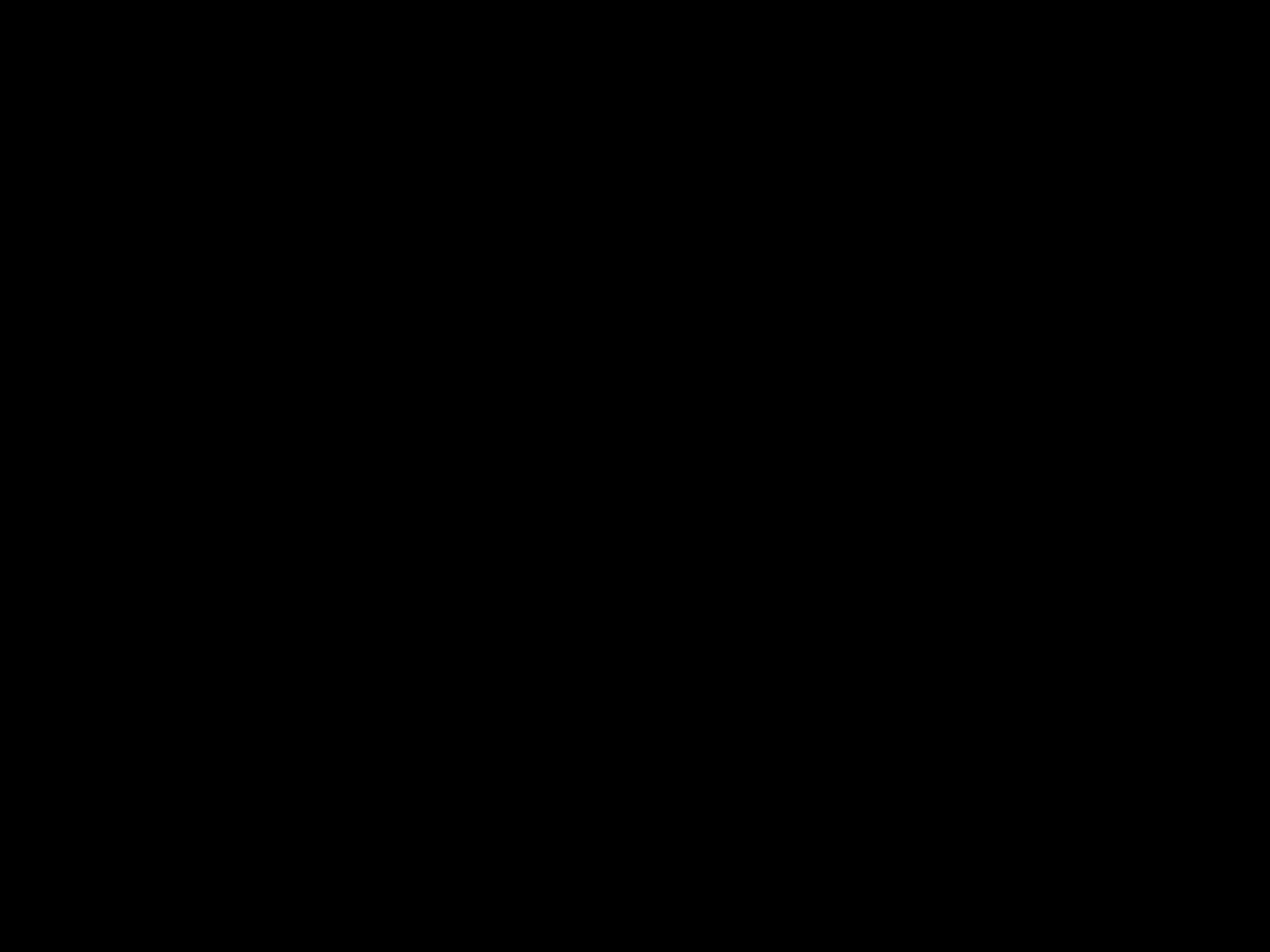
- Kotlin - Percent Format
- Python - Percent Format
- R - Percent Format
- MyST Markdown

# ILCI Breeding Analytics Hub

## Featured tools

### Background computation

- Job submitter tool



# ILCI Breeding Analytics Hub

## Featured tools

### Retrieving data from common databases - BrAPI Helper

The word 'data' is the central focus, rendered in a large, bold, purple font. Surrounding it are numerous other words, each representing a different aspect or tool related to breeding analytics. These words are arranged in a circular pattern around the center, with some being larger than others to indicate their relative importance or frequency. The colors of the words vary, creating a vibrant and diverse palette.

brapi  
analytics  
digitization  
access  
reduction  
error  
analysis  
routine  
mobility  
uptake  
no-files  
formatting  
provenance  
standardized  
consistency

# ILCI Breeding Analytics Hub

## Featured tools

### Retrieving data from common databases

- Why use databases?
  - consistency
  - standardized access
  - error reduction
  - integration opportunities

The diagram features the word 'data' in large, bold, purple font at the center. Surrounding it are various other words in different colors and sizes, representing different concepts related to data management and analysis:

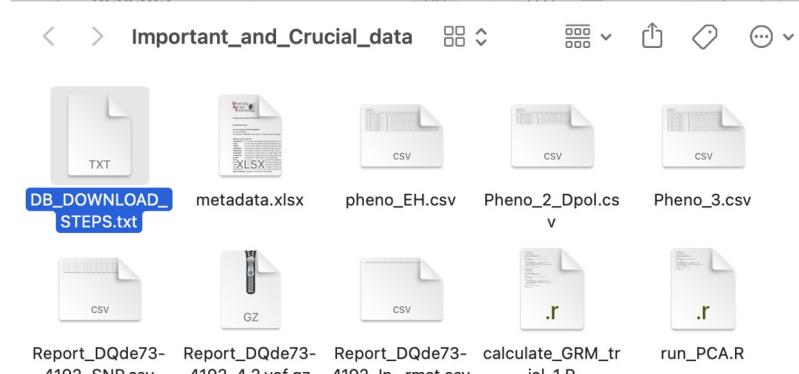
- Top left: brapi (orange)
- Top right: analytics (orange)
- Middle right: digitization (blue), access (blue)
- Bottom right: reduction (purple), error (purple), no-files (yellow)
- Bottom right (further down): analysis (red), routine (red)
- Middle left: consistency (purple)
- Middle left: standardization (yellow)
- Middle left: provenance (blue)
- Middle left: formatting (red)
- Middle left: uptake (yellow)
- Bottom left: mobility (orange)

# ILCI Breeding Analytics Hub

## Featured tools

### Retrieving data from common databases

- Challenges
  - effective and efficient data utilization
  - workflow fragmentation
  - extra steps (e.g. export, cleaning)



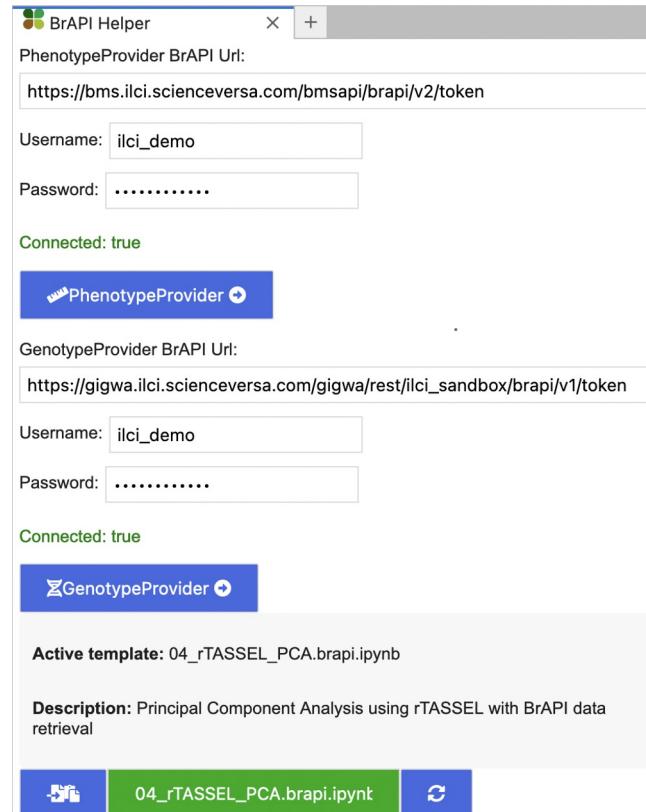
\* Fictional scenario. Any similarity to reality is *purely coincidental...*

# ILCI Breeding Analytics Hub

## Featured tools

### Retrieving data from common databases

- Closing the gap with the BrAPI Helper
  - JupyterLab extension (plugin)
  - streamline data access
  - enable data integration
  - enhance productivity (templates)



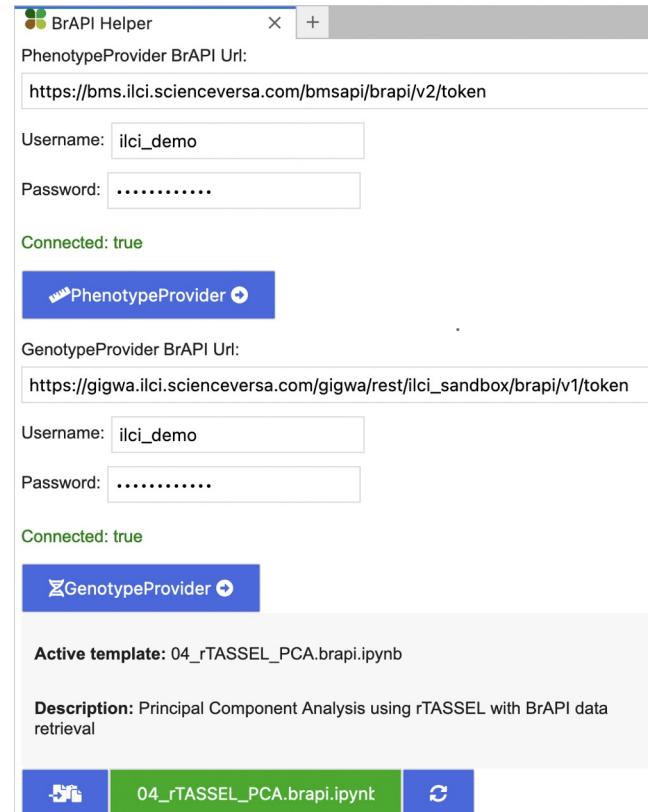
# ILCI Breeding Analytics Hub

## Featured tools

### Retrieving data from common databases: BrAPI Helper Demo

- Connect to phenotype and genotype DBs.
- Bring data into a Jupyter notebook.
- Produce visualizations.

[Start Demo](#)



# ILCI Breeding Analytics Hub

## Featured tools

### rTASSEL

- R library that interfaces with the **TASSEL 5 Java Library**

# ILCI Breeding Analytics Hub Featured tools

## rTASSEL - motivation

The diagram illustrates the rTASSEL tool, highlighting its three main interfaces: Graphical interface, Command line, and TASSEL API (Java). These interfaces are interconnected, represented by a central question mark icon.

**Graphical interface:** A screenshot of a software window showing a grid of data points, likely representing genotype or alignment information.

**Command line:** A terminal window displaying a command-line script:

```
$ ./run_pipeline.pl \
$TASSEL -fork1 -h \
allizea_gbs.t5.h5 \
-filterAlign \
-filterAlignMinFreq \
...
```

**TASSEL API (Java):** A code editor window showing Java code:

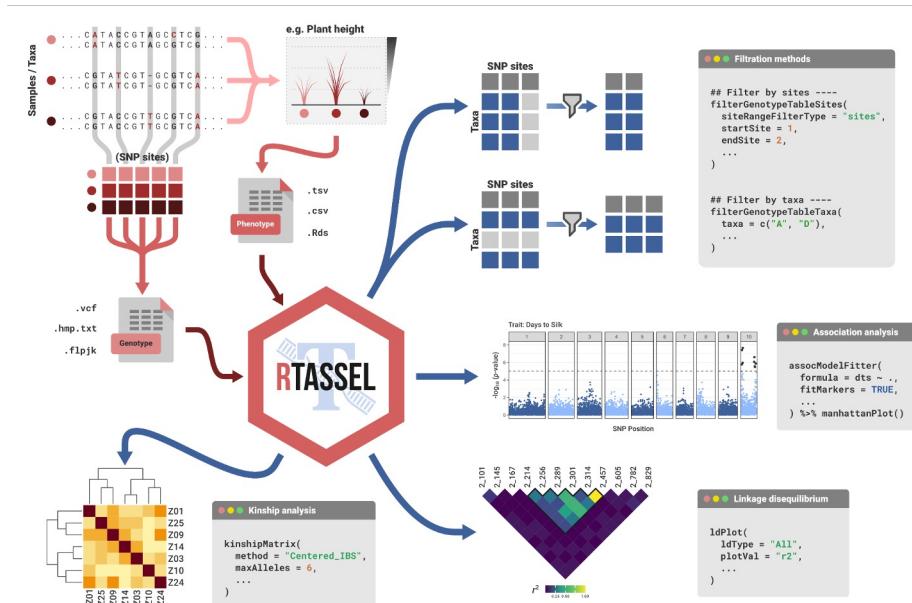
```
import net.maizegenetics.dna.*
Public class Filter {
    GenotypeTable myGeno;
    PositionList posList;
    String[] taxa;
    ...
}
```

**Intuitiveness**

**Extensibility**

# ILCI Breeding Analytics Hub Featured tools

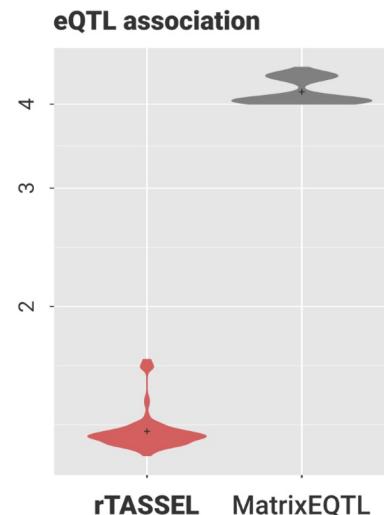
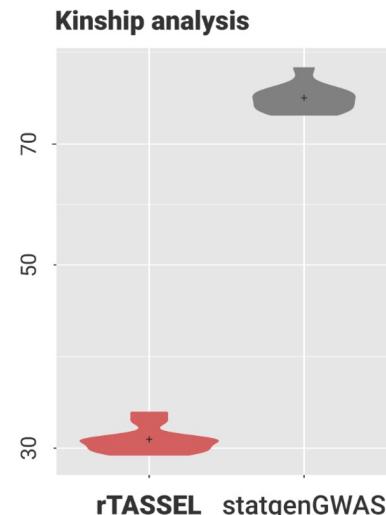
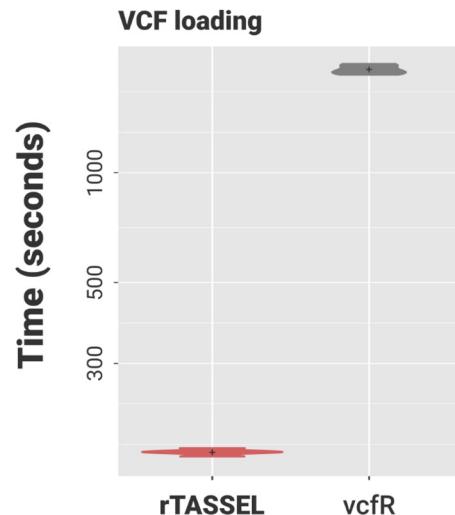
## rTASSEL - interface and interoperability



# ILCI Breeding Analytics Hub

## Featured tools

**rTASSEL is efficient**



# ILCI Breeding Analytics Hub

## Featured tools

**rTASSEL** is **collaborative** (available on the Hub!)

The screenshot shows the RStudio interface with several windows open:

- Code**: A code editor window containing R code related to Tassel analysis.
- Console**: The console output window showing the command `rTASSEL::manhattanPlot(tasGLM$GLM_Stats, "YLDPLOT", 5)`.
- Plots**: A Manhattan plot titled "Trait: YLDPLOT" showing -log10(p-value) on the y-axis (ranging from 0 to 6) against SNP Position on the x-axis. A horizontal dashed line is drawn at approximately 4.3, and a vertical dashed line is at position 809132.
- Environment**: The environment browser showing variables like `tasGLM`, `YLDPLOT`, and `GLM_Stats`.
- Files**: A file browser window showing files like `00_ciclesia_rTassel_main.ipynb`.
- Help**: A help browser window.
- BrAPI**: A sidebar for the BrAPI provider, showing status as "Connected: true" and a URL: [https://igv4.ici.ciceseversa.com/igv4/test/ki/\\_sandbox/brapi](https://igv4.ici.ciceseversa.com/igv4/test/ki/_sandbox/brapi).
- Genotype**: A VCF viewer window showing a chromosome view with controls for "Controls", "Color schemes", "Sorting", and "Export". It displays a grid of genotype data for chromosomes 1-10.
- GenotypeProvider**: A sidebar for the GenotypeProvider, showing status as "Connected: true" and a URL: [https://igv4.ici.ciceseversa.com/igv4/test/ki/\\_sandbox/genotypeprovider](https://igv4.ici.ciceseversa.com/igv4/test/ki/_sandbox/genotypeprovider).

```
[6]: ldRes %>% as.data.frame() %>% head()

A C
Locus1 Position1 Site1 NumberOfStates1 States1 Frequency1 Locus2 Position2 Site2
<chr> <int> <int> <int> <chr> <chr> <chr> <int> <int>
1 1 1947984 1 2 G:C NotImplemented 1 157104
2 1 2914066 2 2 T:G NotImplemented 1 157104
3 1 2914066 2 2 T:G NotImplemented 1 1947984
4 1 2914171 3 2 T:A NotImplemented 1 157104
5 1 2914171 3 2 T:A NotImplemented 1 1947984
6 1 2914171 3 2 T:A NotImplemented 1 2914066

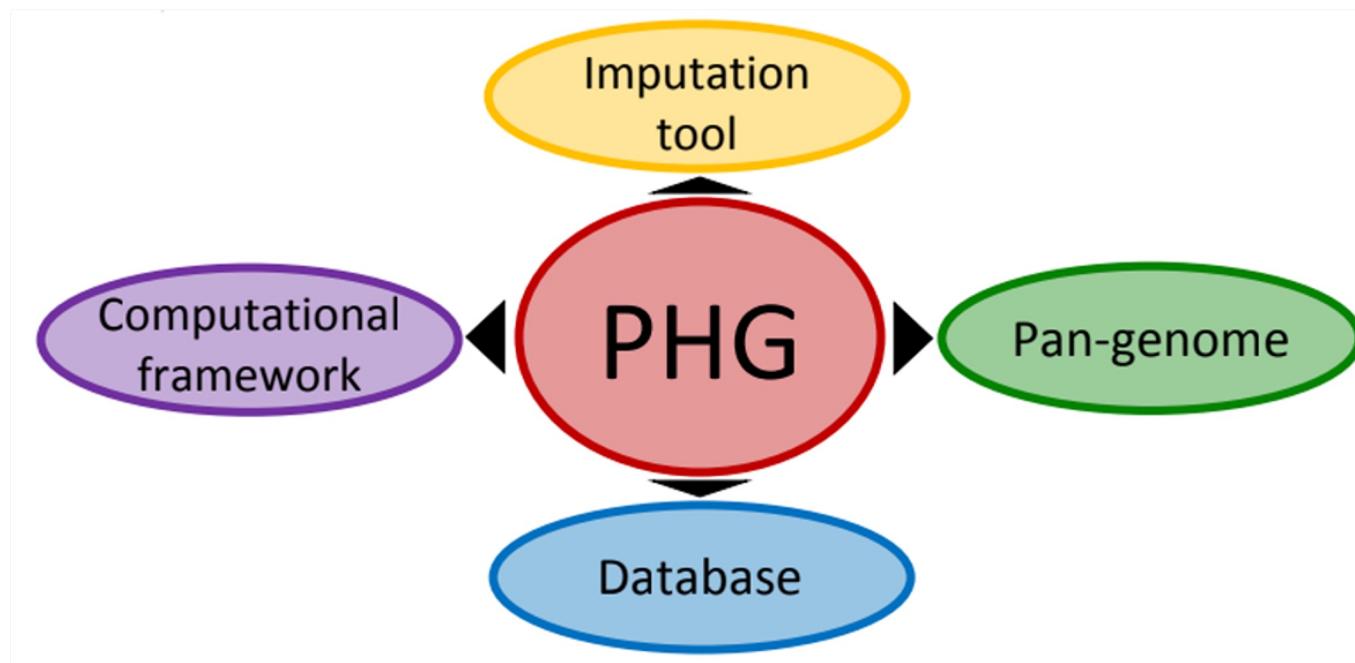
[7]: ## Visualize and pipe functions together -----
tagGeno %>%
  filterGenotypeTableSites(
    siteRangeFilterType = "position",
    startPos = 228e6,
    endPos = 308e6,
    startChr = 2,
    endChr = 2
  ) %>%
  ldPlot(
    ldType = "All",
    plotVal = "DPrime",
    verbose = FALSE
  )

```

# ILCI Breeding Analytics Hub

## Featured tools

### PHG/rPHG

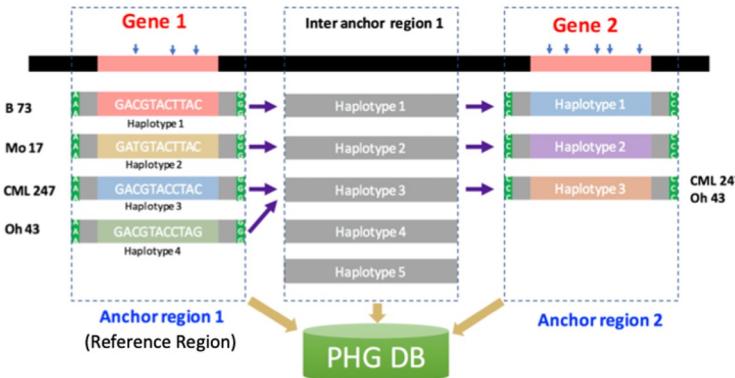


# ILCI Breeding Analytics Hub

## Featured tools

### The PHG is a pan-genome

- The PHG is a graph-based representation of genomic diversity for a breeding population or crop.
- Genomes are divided into genic and intergenic regions and represented as sequences of haplotypes.

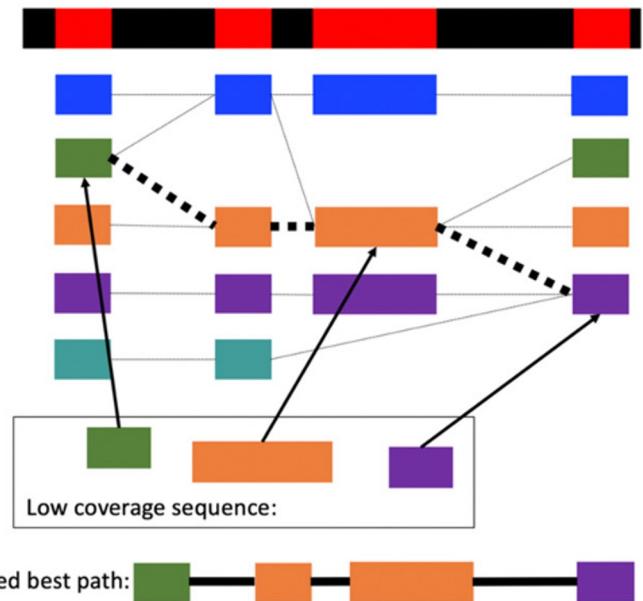


# ILCI Breeding Analytics Hub

## Featured tools

### The PHG is an imputation tool

- Align *low coverage* reads to PHG pangenome
- Find the most likely “path” through the graph (Viterbi/HMM)

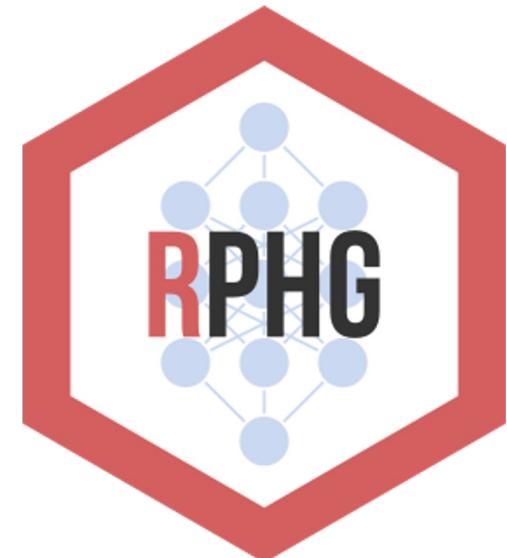


# ILCI Breeding Analytics Hub

## Featured tools

### The PHG is a database of haplotypes

- Associate haplotypes rather than SNPs with traits
  - Reduce computational and statistical dimensionality
- Use R tools to retrieve information
  - **rPHG**
  - Integrate with **rTASSEL**



# ILCI Breeding Analytics Hub

## Coming attractions

### Demo

- rPHG data retrieval
- PHG imputation

# G-Crunch

## What is it?

G-Crunch is a graphical tool for crunching genomics and phenomics data, enabling users to launch repeatable, pre-built analyses across various BrAPI-compliant data sources.

**G - Crunch - Work In Progress**

BrAPI Server Address: <https://bms.ilci.scienceversa.org>

BrAPI crop: sorghum

Username: [ilci\\_demo](#)

Password: [\\*\\*\\*\\*\\*](#)

Study ID: Load

5: Adugna\_trial\_1 Environment Number 2 | 7: JDLS\_TEST\_STUDY Environment Number 1

Unit Type: Plot ▾

Group By Accession

**BrAPI Phenotype Data Preview**

Load

WHERE

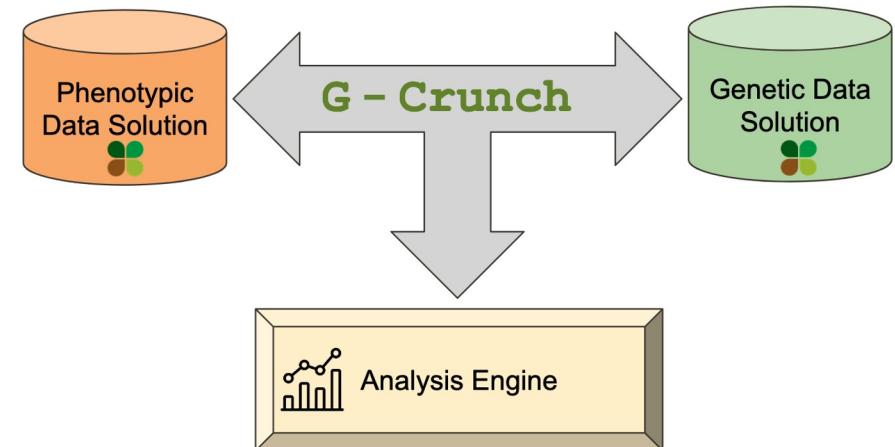
GStdYld\_C\_kgha

3.7e+3 from... Nothing selected. ...to 4.8e+3

GStdYld\_C\_kgha

3.7e+3 from... Nothing selected. ...to 4.8e+3

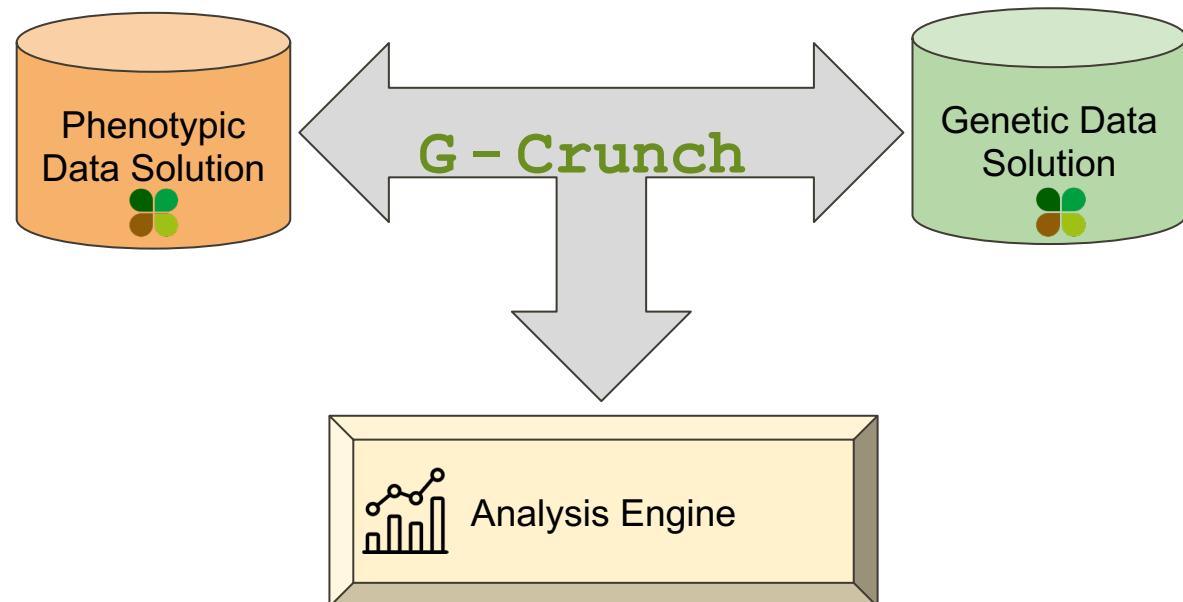
[+] AND [New Range] [New Group]



# G-Crunch

## What does it do?

G-Crunch is a tool to allow users to specify what analyses they want, using BrAPI



# ILCI Breeding Analytics Hub

# Coming attractions

The screenshot shows the G-Crunch web application interface. At the top, there are two tabs: "G-Crunch" and "G-Crunch Analysis Status". The main content area has a title "G - Crunch - Work In Progress".

**Phenotype Database:** A dropdown menu set to "Custom".

**BrAPI Server Address:** An input field.

**BrAPI crop:** An input field.

**Username:** An input field.

**Password:** An input field with masked text.

**Study ID(mult):** A dropdown menu set to "Load". Below it is a large input field containing a list of study IDs.

**Unit Type:** A dropdown menu set to "Plot".

**Group By Accession:** A checkbox.

**Generate Preview:** A button.

**Genotype Server Details [remove]**

**Genotype Database:** A dropdown menu set to "Custom".

**Genotype Server Address:** An input field.

**Genotype crop:** An input field.

**Genotype Server Username:** An input field.

# ILCI Breeding Analytics Hub

## Coming attractions

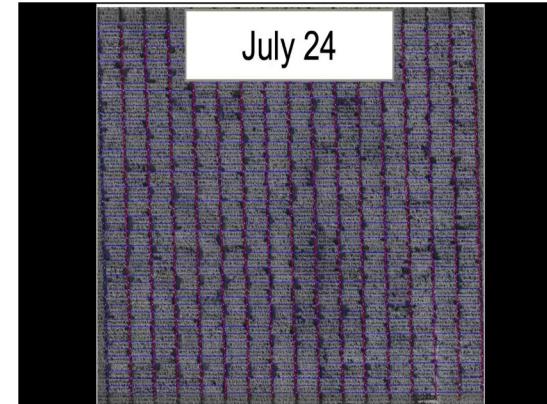
**Remote sensing and image analysis, ImageBreed**

- High-throughput phenotyping

# High-Throughput Phenotyping

## Remote sensing and image analysis (UAV)

- simplifies data collection
- several challenges remain:
  - features extraction challenging
  - integrating traditional phenotypic data
  - data management



# ILCI Breeding Analytics Hub - ImageBreed

## What is it?

ImageBreed is a tool for managing drone data, providing GUI workflows for tasks and metadata input, and phenotypic information extraction of data collection events.

Search Manage Analyze About

Notice: Undergoing System Maintenance Please Do Not Login!

Welcome to ImageBreed!

This system helps you manage field trials, process imaging events, and perform genetically informed predictions.

ImageBreed Tutorials

Manage Ground Rovers Manage Aerial Imagery Manage Field Experiments Genotypic Data

A Docker deployment is available.

Please cite ImageBreed if you find this software useful. A detailed description of the genotyping storage system is published. Please follow my research.

ImageBreed is now integrated with OpenDroneMap for creating orthophotomosaics and digital surface maps (DSM) from color images or MicaSense 5-channel images.

Manage Aerial Imagery: Run A Standard Process On Orthophotomosaics

Please use the Brave or Chrome web-browsers to view these image tools. In the future, other browsers may be supported.

Watch tutorial videos from processing aerial imaging events.

Intro 1 Imaging Event B... 2 Use GCPs 3 Rotate 4 Cropping 5 Thresholding 6 Plot Polygons 7 Apply 8 Indices 9 Phenotypes 10

This workflow will guide you through applying a standard process to your aerial imaging bands

Here you can take one of the imaging event bands you uploaded all the way through the process to plot image saving. This will require manual steps such as image rotation, cropping, and plot polygon templating. After you have completed this process for one imaging event band, you can apply it to all other imaging event bands and other calculated vegetative indices.

Go to Next Step

# ILCI Breeding Analytics Hub - ImageBreed

## What does it do?

ImageBreed uses computer vision to:

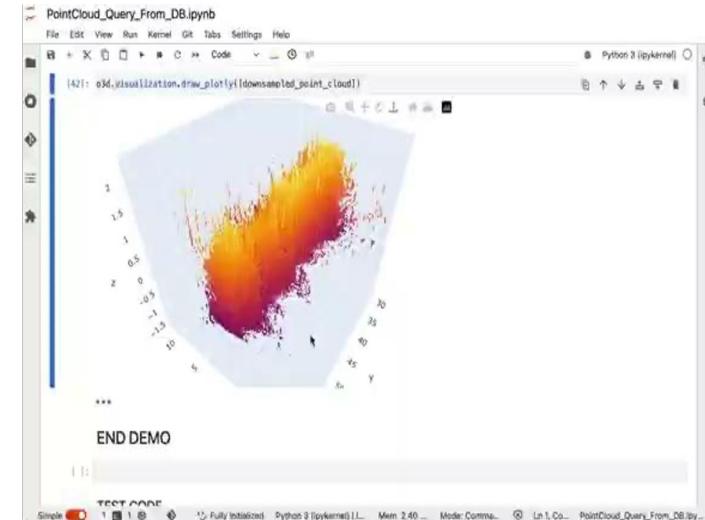
- enable plot assignment for phenotype extraction from UAV data
- vegetative index calculations
- metadata management for multiple field trials and image events

The screenshot displays the ImageBreed software interface, which includes several key components:

- Top Navigation Bar:** Shows "ILCI Breeding Analytics Hub" and "ImageBreed".
- Field Trial Management:** A table lists field trials with columns for "Field Trial", "Planting Date", "Imaging Event Date", "Number of Events", and "Number of Sites". One trial is highlighted: "FT147/150222 Normal" (Planting Date: 2022-01-01, Imaging Event Date: 2022-01-01, Number of Events: 1, Number of Sites: 296).
- Plot Assignment:** Two large images show agricultural plots with green crop rows. Red boxes highlight specific areas for analysis.
- Vegetative Index Calculations:** A progress bar at the bottom indicates the status of "Calculating phenotypes for all plot phenogen designs".
- Metadata Management:** A section titled "Manage Animal Imagery - Run A Standard Process On Orthophotosense" shows a progress bar for "Dilate Imagery to Calculate and Save in Database".
- Image Preview:** Below the progress bars, a row of small thumbnail images shows various agricultural scenes.

## What will it do?

- Expand support for BrAPI-enabled phenotype databases.
- Better support for LIDAR data “point clouds.”
- Codebase modernization to allow interaction with JupyterLab: ImageBreedPy.



# New tool concepts

## Development of concepts for the Breeding Analytics Hub

- What are your current challenges using the Hub?
- Or, if you haven't tried it, what is preventing you from using it?

# New tool concepts

## Development of concepts for the Breeding Analytics Hub

- What type of content would help you with your analyses needs?  
(i.e templates, tutorial notebooks, training)

# New tool concepts

## Development of concepts for the Breeding Analytics Hub

- What obstacles do you face in your crop improvement program that could be addressed by a new feature or tool in the Hub? (Do you want to be involved?)

ILCI Breeding Analytics Hub

# Test drive

Option to stay to test the Hub



The U.S. Government's Global Hunger & Food Security Initiative



Cornell University