

# How to load data to Galaxy and convert the Genotype Grid file to Flapjack project (not split)

**Milcah Kigoni**

m.kigoni@cgiar.org



# Step 1: Create a user account in Galaxy

One time only step

# Genotype-Grid file – Flapjack file Conversion

Step 1: open the Galaxy Server through this link

<http://13.250.212.83/>

or

<http://111.93.2.172:8088>

Create a user account here or log in



### Create account

Email address:

mkigoni@yahoo.com

Password:

.....

Strength

Confirm password:

.....

Public name:

wagio

Your public name is an identifier that will be used to generate addresses for information you share publicly. Public names must be at least three characters in length and contain only lower-case letters, numbers, dots, underscores, and dashes ('.', '\_', '-').

Submit



Workflow "Run Marker-assisted Backcrossing (MABC) analysis (Input: Intertek/LGC Genotype files)" has been imported.

You can [start using this workflow](#) or [return to the previous page](#).

Click here



## Step 2: Add HTPG Workflows

One time step

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. The 'Shared Data' menu is open, showing options: 'Access published resources', 'Histories', 'Workflows' (highlighted), 'Visualizations', and 'Pages'. A red box highlights the 'Shared Data' menu and the 'Workflows' option. A red arrow points from the 'Workflows' option to a text box at the bottom right that says 'Click on Shared data in the menu bar then workflows'.



Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Published Workflows

search name, annotation, owner, and tags

Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
Run Marker-assisted Backcrossing (MABC) analysis (Input: Intertek/LGC Genotype file)		wkigoni	★★★★★		~5 days ago
Run F1/Pedigree Verification (Input: HTPG Sample file and Intertek/LGC Genotype files)	Run this workflow to conduct an F1 or Pedigree verification	wkigoni	★★★★★		~5 days ago
Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)	Run this workflow to get a Flapjack project file from Intertek/LGC file	wkigoni	★★★★★		~5 days ago
Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype grid and HTPG Sample files)	Run this workflow to create split your genotype data into groups/populations in Flapjack	wkigoni	★★★★★		~5 days ago

History

search datasets

Unnamed history

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Click on import for all workflows

Click on Workflow to see imported workflows now in your user account

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Published Workflows

Chain tools into workflows

search name, annotation, owner, and tags

Advanced Search

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated↓
Run Marker-assisted Backcrossing (MABC) analysis (Input: Intertek/LGC Genotype files)		wkigoni	★★★★★		~5 days ago
Run F1/Pedigree Verification (Input: HTPG Sample file and Intertek/LGC Genotype files)	Run this workflow to conduct an F1 or Pedigree verification	wkigoni	★★★★★		~5 days ago
Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)	Run this workflow to get a Flapjack project file from Intertek/LGC file	wkigoni	★★★★★		~5 days ago
Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype grid and HTPG Sample files)	Run this workflow to create split your genotype data into groups/populations in Flapjack	wkigoni	★★★★★		~5 days ago

History

search datasets

Unnamed history

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Show all Workflows in the tool Panel

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Your workflows

search for workflow...

Name	Tags	Owner	# of Steps	Published	Show in tools panel
imported: Intertek (HTPG) - Flapjack Analysis		You	4	No	<input checked="" type="checkbox"/>
imported: Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype grid and HTPG Sample files)		You	5	No	<input checked="" type="checkbox"/>
imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)		You	3	No	<input checked="" type="checkbox"/>
imported: Run F1/Pedigree Verification (Input: HTPG Sample file and Intertek/LGC Genotype files)		You	3	No	<input checked="" type="checkbox"/>
imported: Run Marker-assisted Backcrossing (MABC) analysis (Input: Intertek/LGC Genotype files)		You	3	No	<input checked="" type="checkbox"/>

History

search datasets

Unnamed history

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Check this box for all workflows

Click on Analyse data to see all the workflows in the tool panels

**Galaxy / EiB-demo**

Analyze Data Workflow Shared Data Visualization Help User Using 0%

**Tools**

TOOLS FROM TOOLSHED

search tools

**GOBII TOOLS**

**Flapjack Tools**

**Workflows**

- All workflows
- imported: Intertek (HTPG) - Flapjack Analysis
- imported: Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype\_grid and HTPG Sample files).
- imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file).
- imported: Run F1/Pedigree Verification (Input: HTPG Sample file and Intertek/LGC Genotype files).
- imported: Run Marker-assisted Backcrossing (MABC) analysis (Input: Intertek/LGC Genotype files).

**Excellence in Breeding PLATFORM**

**Welcome to Galaxy!**

This instance is dedicated to host tools for the GOBII project, HTPG project and EiB platform.

Take an interactive tour: Galaxy UI History Scratchbook

**History**

search datasets

**Unnamed history**

(empty)

**i** This history is empty. You can [load your own data](#) or [get data from an external source](#)

NB: In future, all these workflows will be present on the tools pane and hence whenever you log into your account

## Step 1: Loading data into Galaxy

NB: This is a routine step whenever running an analysis in Galaxy that requires a file as the input

# Galaxy Interphase

The screenshot displays the Galaxy Interphase web interface in a browser window. The browser's address bar shows the URL 111.93.2.172:8083. The interface has a dark blue header with the 'Galaxy' logo and navigation links: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. A status bar on the right indicates 'Using 0 bytes'. The main content area is divided into three vertical panels, each outlined with a red border. The left panel, titled 'Tools', contains a search bar and a list of tool categories: 'Get Data', 'Send Data', 'Collection Operations', 'Text Manipulation', 'Filter and Sort', 'Join, Subtract and Group', 'Convert Formats', 'Extract Features', 'Fetch Sequences', 'Fetch Alignments', 'Statistics', 'Graph/Display Data', 'HTPG Tools', 'NGS', and 'Population Tools'. Under 'Workflows', there are links for 'All workflows' and 'HTPG Flapjack Analysis'. The middle panel, titled 'Hello, Galaxy is running!', provides instructions to customize the page by editing 'static/welcome.html' and offers buttons for 'Configuring Galaxy' and 'Installing Tools'. It also includes an interactive tour with links for 'Galaxy UI', 'History', and 'Scratchbook'. Below this, a paragraph describes Galaxy as an open platform for data-intensive research, developed by 'The Galaxy Team' with support from various institutions. The right panel, titled 'History', features a search bar and shows an 'Unnamed history' section that is currently empty. A blue information box states: 'This history is empty. You can load your own data or get data from an external source'.

## Tools panel:

This is where you find the tools and Workflows

## Analysis Panel:

This is where the analysis is done when a tool or workflow is run.

## Data panel:

This is where all Uploaded data and results files from analysis are found.

Click on Upload file

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Upload File from your computer

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

Excellence in Breeding PLATFORM

Welcome to Galaxy!

This instance is dedicated to host tools for the [GOBII](#) project, [HTPG](#) project and [EiB](#) platform.

Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

History

search datasets

Unnamed history

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

← → ↻

Not secure | 13.250.212.83

☆ 🔥 🔒 🟢 📁 🗨️ 💡 📧 📺 🆕 🏠 ⋮

Galaxy / EiB-demo

Analyze Data Workflow Shared Data Visualization Help User

Using 0%

Tools

search tools

Get Data

Upload File from your computer

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

Download from web or upload from disk

Regular

Composite

Collection

Drop files here

Type (set all):

Auto-detect

Q

Genome (set all):

----- Additional Species A...

Choose local file

Paste/Fetch data

Pause

Reset

Start

Close



Select your genotype Grid file in your computer

The image shows a web application interface with a Windows file explorer window open over it. The file explorer is displaying the contents of a folder named 'direct\_Intertek-Fj'. It lists two files: 'direct\_intertek-Fj\_conversion\_example\_output' and 'example\_grid\_file'. The 'example\_grid\_file' is selected and highlighted. A red rectangle is drawn around the 'File name' field, which contains 'example\_grid\_file', and the 'Open' button. The web application interface in the background includes a top navigation bar with various icons and a sidebar with links like 'Fetch Sequences', 'Fetch Alignments', 'Statistics', 'Graph/Display Data', 'VCF intersect', 'HTPG TOOLS', 'Pre-genotyping', and 'Post-genotyping'. The main content area has a large empty box with a dashed border and a 'Files here' label. At the bottom, there are controls for 'Type (set all):' (set to 'Auto-detect') and 'Genome (set all):' (set to '---- Additional Species A...'). A row of buttons at the very bottom includes 'Choose local file', 'Paste/Fetch data', 'Pause', 'Reset', 'Start', and 'Close'.

Open

HTPG\_workfl... > direct\_Intertek-Fj

Search direct\_Intertek-Fj

Organize New folder

This PC

3D Objects

Desktop

Documents

Downloads

Music

Pictures

Videos

Windows (C:)

LENOVO (D:)

DATA (E:)

Network

Name	Date modified
direct_intertek-Fj_conversion_example_output	8/15/2018 11:45 AM
<input checked="" type="checkbox"/> example_grid_file	8/15/2018 11:44 AM

File name: example\_grid\_file

All Files

Open

Cancel

Files here

Type (set all): Auto-detect

Genome (set all): ---- Additional Species A...

Choose local file

Paste/Fetch data

Pause

Reset

Start

Close

**Galaxy / EiB-demo**    Analyze Data   Workflow   Shared Data ▾   Visualization ▾   Help ▾   User ▾   Using 0%

**Tools**

search tools

**Get Data**  
Upload File from your computer

**Genomic Selection**

**Data Format Conversion**

**Marker Selection**

**Imputation**

**Cluster Analysis**

**GWAS tools**

**Text Manipulation**

**Collection Operations**

**Filter and Sort**

**Join, Subtract and Group**

**Extract Features**

**Fetch Sequences**

**Fetch Alignments**

**Statistics**

**Graph/Display Data**

**VCF intersect**



**HTPG TOOLS**

**Pre-genotyping**

**Post-genotyping**



### Download from web or upload from disk

Regular   Composite   Collection

Name	Size	Type	Genome	Settings	Status
 example_grid_file.xlsx	234.5 KB	Auto-dete... ▾	----- Additional Sp... ▾		100% ✓

Type (set all): Auto-detect ▾   Q

Genome (set all): ----- Additional Species A... ▾

 Choose local file    Paste/Fetch data   Pause   Reset   **Start**   Close

Click on Start to upload the file to Galaxy then Close

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Upload File from your computer

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

Excellence  
in Breeding  
PLATFORM

Welcome to Galaxy!

This instance is dedicated to host tools for the [GOBII](#) project, [HTPG](#) project and [EiB](#) platform.

Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

History

search datasets

Unnamed history

1 shown

234.53 KB

1: example\_grid\_file.xlsx

Your file will be added to the data panel

Click on the desired workflow for your data conversion

The screenshot displays the Galaxy / EiB-demo interface. On the left, a sidebar lists various tools and workflows. Two specific workflows are highlighted with red boxes and annotated with callouts:

- Workflow 1:** imported: Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype grid and HTPG Sample files). This workflow is associated with the callout: "To create a Flapjack file with split populations".
- Workflow 2:** imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file). This workflow is associated with the callout: "To convert a single Genotype file directly to Fj project file".

The main content area features the "Excellence in Breeding PLATFORM" logo and a "Welcome to Galaxy!" message. It states: "This instance is dedicated to host tools for the GOBII project, HTPG project and EiB platform." Below this, it offers an interactive tour with buttons for "Galaxy UI", "History", and "Scratchbook".

On the right side, the "History" panel shows a search bar and a list of datasets. The first entry is "1: example\_grid\_file.xlsx" with a size of 234.53 KB.

Step 2:

Running the “Directly convert a Genotype Grid File to a Flapjack Project”  
Workflow

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Workflow: imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)

Run workflow

History Options

Send results to a new history

YesNo

1: Intertek/LGC Genotype Grid file (Galaxy Version 1.0.0)

excel\_workbook

1: example\_grid\_file.xlsx

1: example\_grid\_file.xlsx

2: Intertek-Flapjack Converter (Galaxy Version 1.0.0)

3: Flapjack Project File (Galaxy Version 0.0.1)

Flapjack genotype file

Output dataset 'out\_file' from step 2

Flapjack map file

Nothing selected

Flapjack-formatted (tab-delimited text) input file of map data

Flapjack phenotype file

History

search datasets

Unnamed history

1 shown

234.53 KB

1: example\_grid\_file.xlsx

Information about what the workflow does and the required input is available here

Choose your input data here

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Workflow: imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)

History Options

Send results to a new history

YesNo

1: Intertek/LGC Genotype Grid file (Galaxy Version 1.0.0)

excel\_workbook

1: example\_grid\_file.xlsx

Select an excel file

Job Post Actions

Hide output 'sheet2'. Hide output 'sheet3'. Hide output 'sheet1'.

2: Intertek-Flapjack Converter (Galaxy Version 1.0.0)

3: Flapjack Project File (Galaxy Version 0.0.1)

Flapjack genotype file

Output dataset 'out\_file' from step 2

Flapjack map file

Nothing selected

Flapjack-formatted (tab-delimited text) input file of map data

Flapjack phenotype file

Run workflow

History

search datasets

Unnamed history

1 shown

234.53 KB

1: example\_grid\_file.xlsx

Click here to run analysis

Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

Successfully invoked workflow **imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file).**

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history

6 shown

234.53 KB

6: Flapjack CreateP  
roject on data 5

5: Flapjack\_file.gen  
otype

4: Worksheet3\_file.  
txt

3: Worksheet2\_file.  
txt

2: Worksheet1\_file.  
txt

1: example\_grid\_file.x  
lsx

Check here for the progress. All output files should turn green when the analysis is over



Galaxy / EiB-demo

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 0%

Tools

search tools

Get Data

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

✓

Successfully invoked workflow **imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file).**

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history

3 shown, 3 hidden

1.44 MB

6: Flapjack CreateProject on data 5

589.0 KB

format: **flapjack**, database: ?

Map/marker cache created in 3ms

Reading line 100 (7ms)

Reading line 200 (6ms)

Reading line 300 (4ms)

Reading line 400 (2ms)

Reading line 500 (3ms)

Reading line 600 (2ms)

Reading line 700 (2ms)

Reading line 800 (2ms)

Reading line 900 (2ms)

Reading line 10

Display in Flapjack: [View](#) ([Help](#))

binary data

5: Flapjack\_file.gen

Click on the Flapjack file to see details

## Tools

Get DataGenomic SelectionData Format ConversionMarker SelectionImputationCluster AnalysisGWAS toolsText ManipulationCollection OperationsFilter and SortJoin, Subtract and GroupExtract FeaturesFetch SequencesFetch AlignmentsStatisticsGraph/Display DataVCF intersect

## HTPG TOOLS

Pre-genotypingPost-genotyping

## TOOLS FROM TOOLSHED



Successfully invoked workflow **imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file).**

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

## History

**Unnamed history**3 shown, 3 [hidden](#)

1.44 MB

**6: Flapjack CreatePr  
object on data 5**

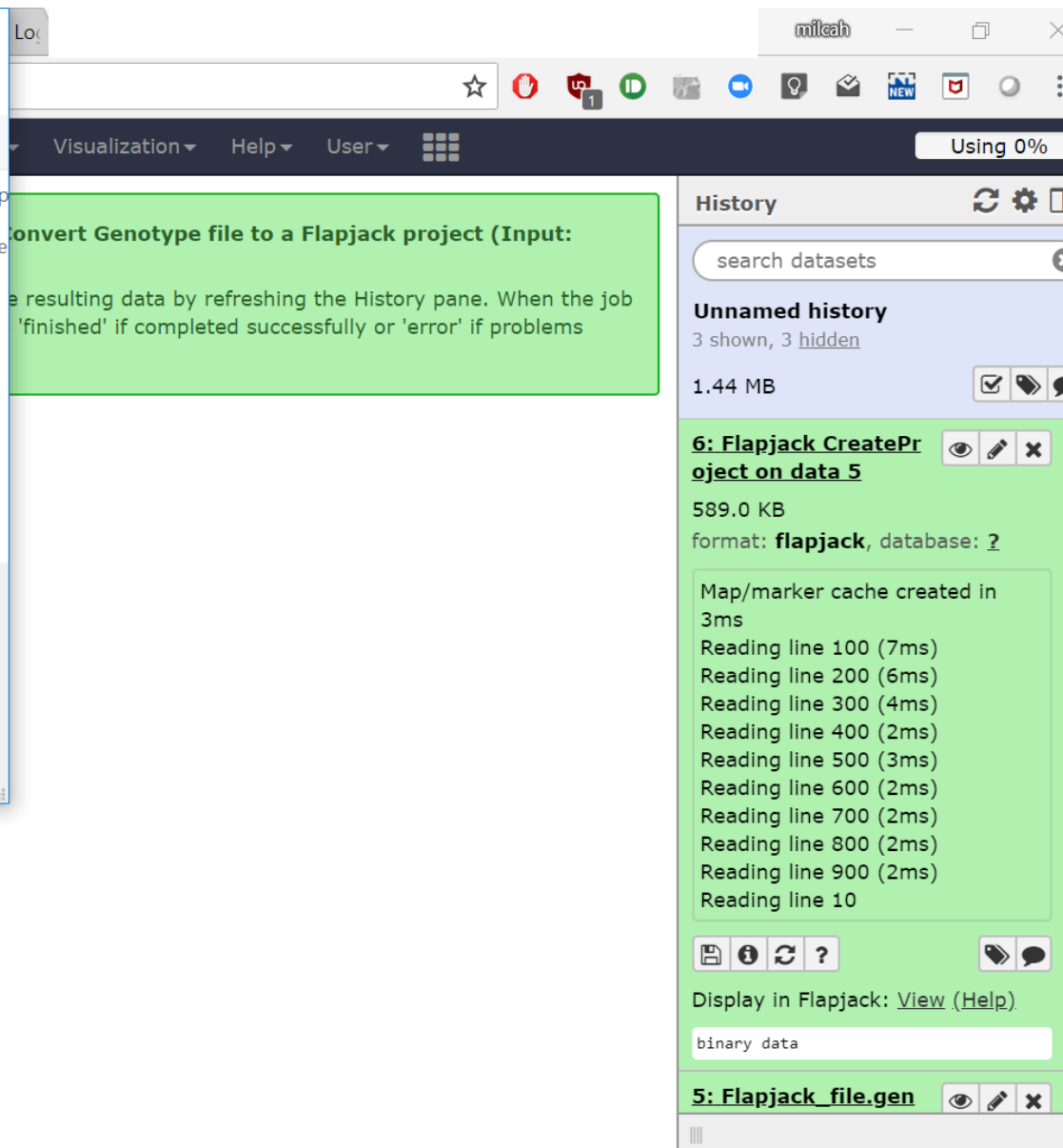
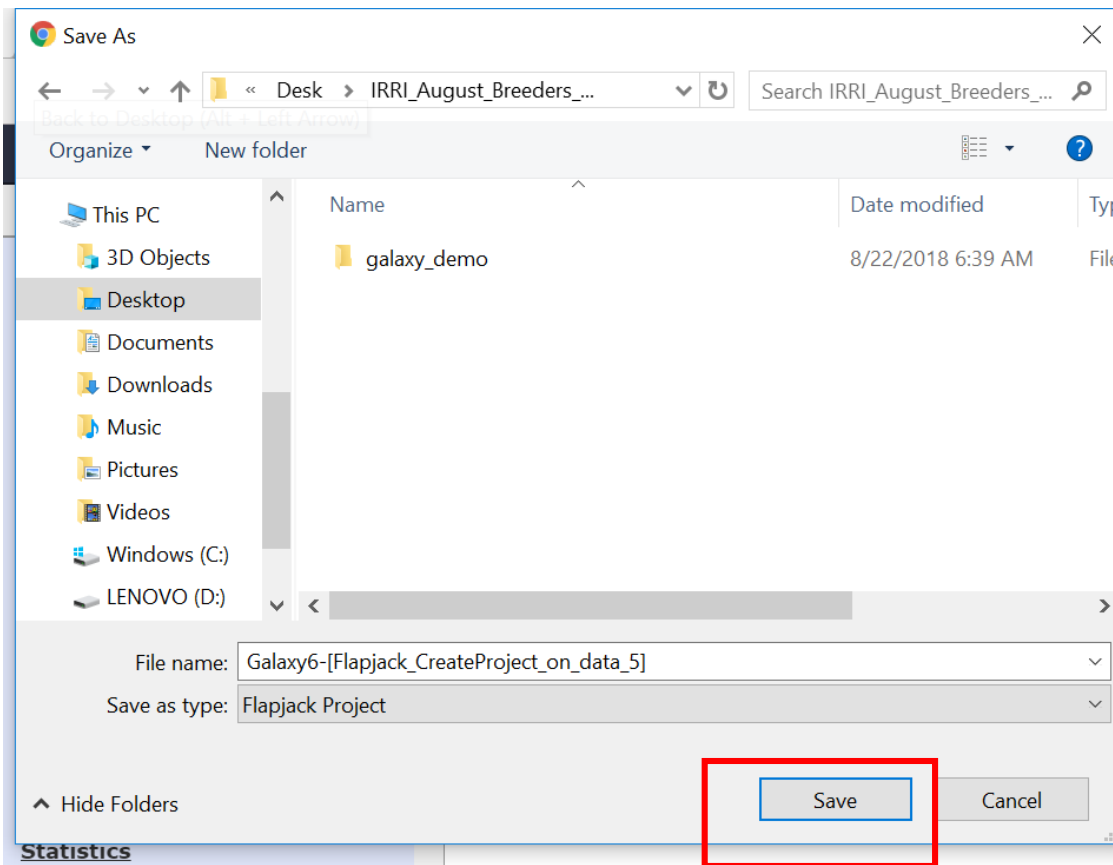
589.0 KB

format: **flapjack**, database: **2**

Map/marker cache created in  
3ms  
Reading line 100 (7ms)  
Reading line 200 (6ms)  
Reading line 300 (4ms)  
Reading line 400 (2ms)  
Reading line 500 (3ms)  
Reading line 600 (2ms)  
Reading line 700 (2ms)  
Reading line 800 (2ms)  
Reading line 900 (2ms)  
Reading line 10

**Download** in Flapjack: [View](#) ([Help](#)).

binary data



# Open the save file in Flapjack and run analysis



How to create a Split-populations Flapjack project file

## Step 1. Create the following analysis file

Cut

Copy

Format Painter

Paste

Clipboard

Calibri

11

A

A

B

I

U

Wrap Text

Alignment

General

Number

Conditional Formatting

Format as Table

Cell Styles

Styles

Insert

Delete

Format

Cells

AutoSum

Fill

Clear

Editing

Sort & Find & Filter

Select

J1

1

KBiosciences grid report

2

Grid version 1.03

3

More information is available in the Genotyping-.csv file. This file lists only the calls for each SNP on each well with a subject ID. When a subject ID is duplicated and the calls don't match

4

Project number

5

Order number

6

Plates

7

8

DNA \ Assay

snpOS0089

snpOS0096

snpOS0061

snpOS0054

snpOS0002

snpOS0007

snpOS0009

snpOS0024

snpOS0040

snpOS0295

snpOS0296

snpOS0297

snpOS0298

snpOS0299

9

2db17703b53a

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

10

494d342c5de2

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

11

43b60a8864c4

G:G

C:C

G:G

C:C

A:A

?

A:A

A:A

C:C

C:C

?

C:C

A:A

A:A

12

eb3c0afb963c

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

13

041e37041746

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

14

ed19fd708927

G:G

?

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

15

66324c99753e

G:G

?

C:C

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

16

33b8a0f69eea

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

17

45d4e08a7b57

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

18

edff59da7d58

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

19

f3d1fb26c36c

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

20

e55de89d3d4c

?

?

C:G

?

?

?

A:T

?

?

?

?

?

?

A:T

21

a0b87f5b99c5

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

22

9ca31f064b28

G:G

C:C

G:G

C:C

A:A

G:G

A:A

A:A

C:C

C:C

G:G

C:C

A:A

A:A

Genotype\_Grid\_file

Sample\_file

SNP\_info\_Summary\_file

## Worksheet 1: Genotype Grid file

FileHomeInsertPage LayoutFormulasDataReviewViewTell me what you want to do...Kigoni, Milcah (ICRISAT-Nairobi)

CutCopyFormat Painter

Clipboard

Calibri11A A

B I U

Font

Wrap TextMerge & Center

Alignment

General

\$ % , .0 .00

Number

Conditional FormattingTableCell Styles

Styles

InsertDeleteFormat

Cells

AutoSumFillClearSort & Find & FilterSelect

Editing

J5

xbl-2018-P3

	A	B	C	D	E	F	G	H	I	J	K	L
	dnarun_name	germplasm_name	dnasample_name	germplasm_pedigree	germplasm_type	germplasm_par1	germplasm_par1_type	germplasm_par2	germplasm_par2_type	dnasample_sample_group	dnasample_sample_group_cycle	dnasample_UUID
1	2db17703b53a	XBL-167-s1		inbred	parent					xbl-2018-P3		142cf886-a065-11e8-98d0-52
2	494d342c5de2	XBL-167-s2		inbred	parent					xbl-2018-P3		142cfc5a-a065-11e8-98d0-52
3	43b60a8864c4	XBL-167-s3		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142cfec6-a065-11e8-98d0-52
4	eb3c0afb963c	XBL-167-s4		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d0164-a065-11e8-98d0-52
5	041e37041746	XBL-167-s5		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d03d0-a065-11e8-98d0-52
6	ed19fd708927	XBL-167-s6		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d05f6-a065-11e8-98d0-52
7	66324c99753e	XBL-167-s7		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d0b82-a065-11e8-98d0-52
8	33b8a0f69eea	XBL-167-s8		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d0de4-a065-11e8-98d0-52
9	45d4e08a7b57	XBL-167-s9		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d1000-a065-11e8-98d0-52
10	edff59da7d58	XBL-167-s10		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d123a-a065-11e8-98d0-52
11	f3d1fb26c36c	XBL-167-s11		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d14ce-a065-11e8-98d0-52
12	e55de89d3d4c	XBL-167-s12		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d17d0-a065-11e8-98d0-52
13	a0b87f5b99c5	XBL-167-s13		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d2374-a065-11e8-98d0-52
14	9ca31f064b28	XBL-167-s14		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d2608-a065-11e8-98d0-52
15	f174d75ed456	XBL-167-s15		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d2856-a065-11e8-98d0-52
16	87d7e26e0a7f	XBL-167-s16		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d2aa4-a065-11e8-98d0-52
17	83aafcd2e60d	XBL-167-s17		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d2cca-a065-11e8-98d0-52
18	6769870c098a	XBL-167-s18		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d368e-a065-11e8-98d0-52
19	7d580709fcd7	XBL-167-s19		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d3bc0-a065-11e8-98d0-52
20	773ff9f46a82	XBL-167-s20		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d3eae-a065-11e8-98d0-52
21	b2c61fea5f3f	XBL-167-s21		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d414c-a065-11e8-98d0-52
22	79fe18e33844	XBL-167-s22		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d43ea-a065-11e8-98d0-52
23	132c1ded06a1	XBL-167-s23		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d4f3e-a065-11e8-98d0-52
24	21dddd729df9	XBL-167-s24		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d51a0-a065-11e8-98d0-52
25	c0d2d40b80a3	XBL-167-s25		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d53e4-a065-11e8-98d0-52
26	b4ff5601f33a	XBL-167-s26		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d55f6-a065-11e8-98d0-52
27	1b8badb4e7ac	XBL-167-s27		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d5858-a065-11e8-98d0-52
28	31250a1658c4	XBL-167-s28		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d5a74-a065-11e8-98d0-52
29	bf813c051e0e	XBL-167-s29		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d5ca4-a065-11e8-98d0-52
30	405d291372cb	XBL-167-s30		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d68de-a065-11e8-98d0-52
31	9e0dcf8473ee	XBL-167-s31		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d6b36-a065-11e8-98d0-52
32	22c6f248a26	XBL-167-s32		KR-104/Sika-105	F1	KR 104	DP	SIKA	RP	xbl-2018-P3		142d7a2c-a065-11e8-98d0-52

Genotype\_Grid\_file

Sample\_file

SNP\_info\_Summary\_file

## Worksheet 2: Your filled HTPG Sample file

NB: Splitting will be done according to how samples are group in column (dnasample\_sample\_group). See a completed example above

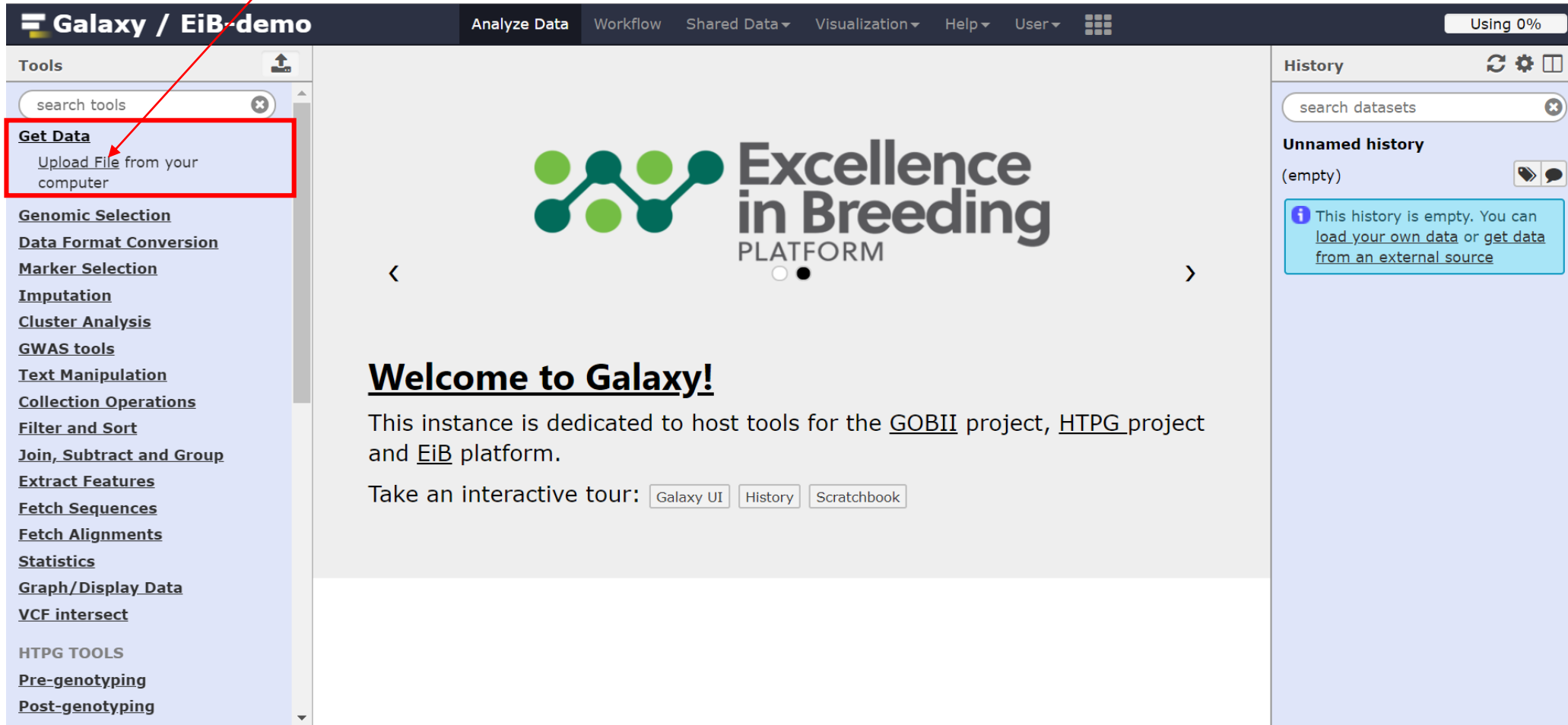
C18							
	A	B	C	D	E	F	G
1	<b>Intertek SNP ID</b>	<b>Customer_SNP_ID</b>	<b>Trait/Gene</b>	<b>Favorable_Allele</b>	<b>Unfavorable_Allele</b>		
2	snpOS0089	MSU7_3_31323659_[C/G]	Drought/DTY3.1	C	T		
3	snpOS0096	S12_17618268	Drought/DTY12.1	AG	TC		
4	snpOS0061	Xa21_SKEP	BLB/Xa21	A	C		
5	snpOS0054	xa5-S1_SKEP	BLB/xa5	G	A		
6	snpOS0002	S8_27520607	BLB/xa13	A	G		
7	snpOS0007a	Pi9-1a	Blast/Pi9	A	G		
8	snpOS0009	Pi2-1	Blast/Pi2	A	T		
9	snpOS0024	chalk5_576	Chalky/Chalk5	T	C		
10	snpOS0031	BadH2_3bp_E12	Fragrance/BADH2	T	C		
11	snpOS0040	s9_6774928	Submergence/Submergence	G	A		
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

**worksheet 3:** Your SNP/Marker summary information in (optional)



## Step 2. Upload your analysis file in Galaxy

Click on Upload file



The screenshot displays the Galaxy / EiB-demo web interface. The top navigation bar includes links for 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User', along with a 'Using 0%' status indicator. The left sidebar, titled 'Tools', contains a search bar and a list of tool categories. The 'Get Data' category is highlighted with a red box, and a red arrow points to the 'Upload File from your computer' option. The main content area features the 'Excellence in Breeding PLATFORM' logo and a 'Welcome to Galaxy!' message. Below the welcome message, it states: 'This instance is dedicated to host tools for the [GOBII](#) project, [HTPG](#) project and [EiB](#) platform.' and provides links for 'Galaxy UI', 'History', and 'Scratchbook'. The right sidebar, titled 'History', shows an 'Unnamed history (empty)' section with a message: 'This history is empty. You can [load your own data](#) or [get data from an external source](#)'.

**Galaxy / EiB-demo**

Analyze Data Workflow Shared Data Visualization Help User Using 0%

**Tools**

search tools

**Get Data**

Upload File from your computer

Genomic Selection

Data Format Conversion

Marker Selection

Imputation

Cluster Analysis

GWAS tools

Text Manipulation

Collection Operations

Filter and Sort

Join, Subtract and Group

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

VCF intersect

HTPG TOOLS

Pre-genotyping

Post-genotyping

**Excellence in Breeding PLATFORM**

**Welcome to Galaxy!**

This instance is dedicated to host tools for the [GOBII](#) project, [HTPG](#) project and [EiB](#) platform.

Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

**History**

search datasets

**Unnamed history**

(empty)

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Galaxy

111.93.2.172:8083

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Get Data

plink - Filter large VCF file

Upload File from your computer

UCSC Main table browser

UCSC Test table browser

UCSC Archaea table browser

EBI SRA ENA SRA

Get Microbial Data

BioMart Ensembl server

CBI Rice Mart rice mart

GrameneMart Central server

modENCODE fly server

Flymine server

Flymine test server

modENCODE modMine server

MouseMine server

Ratmine server

YeastMine server

Download from web or upload from disk

Regular Composite Collection

Name	Size	Type	Genome	Settings	Status
File3_HTPG_analysis_file(Galaxy_input).xlsx	697.6 KB	Auto-detect	----- Additional Species A...		100% ✓

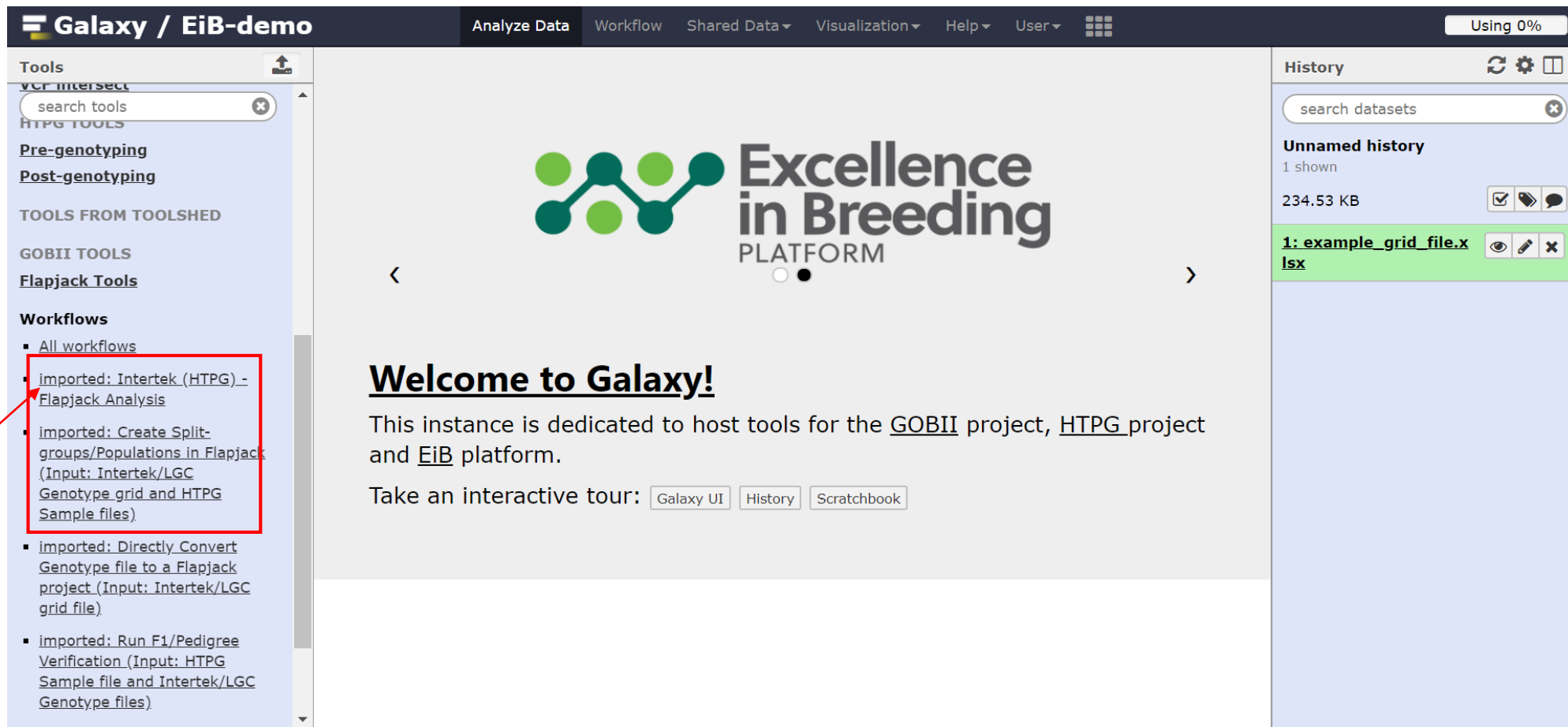
Type (set all): Auto-detect Genome (set all): ----- Additional Species A...

Choose local file Paste/Fetch data Pause Reset Start Close

1. Click here to browse for excel file in your computer

2. Click on start to load, then close

Click on the Create Split-groups/populations workflow or Intertek (HTPG) Flapjack Analysis



The screenshot shows the Galaxy / EiB-demo interface. The top navigation bar includes links for Analyze Data, Workflow, Shared Data, Visualization, Help, and User. The left sidebar contains a 'Tools' section with a search bar and a list of categories: VCF intersect, HTPG TOOLS, Pre-genotyping, Post-genotyping, TOOLS FROM TOOLSHED, GOBII TOOLS, Flapjack Tools, and Workflows. The 'Workflows' section is expanded, showing a list of workflows. A red box highlights the workflow 'imported: Intertek (HTPG) - Flapjack Analysis'. A red arrow points from a text box to this workflow. The main content area displays the 'Excellence in Breeding PLATFORM' logo and a 'Welcome to Galaxy!' message. The right sidebar shows a 'History' section with a search bar and a list of datasets, including '1: example\_grid\_file.xlsx'.

**Galaxy / EiB-demo**

Analyze Data Workflow Shared Data Visualization Help User Using 0%

**Tools**

search tools

HTPG TOOLS

Pre-genotyping

Post-genotyping

TOOLS FROM TOOLSHED

GOBII TOOLS

Flapjack Tools

**Workflows**

- All workflows
- imported: Intertek (HTPG) - Flapjack Analysis
- imported: Create Split-groups/Populations in Flapjack (Input: Intertek/LGC Genotype grid and HTPG Sample files)
- imported: Directly Convert Genotype file to a Flapjack project (Input: Intertek/LGC grid file)
- imported: Run F1/Pedigree Verification (Input: HTPG Sample file and Intertek/LGC Genotype files)

**Excellence in Breeding PLATFORM**

**Welcome to Galaxy!**

This instance is dedicated to host tools for the GOBII project, HTPG project and EiB platform.

Take an interactive tour: [Galaxy UI](#) [History](#) [Scratchbook](#)

**History**

search datasets

**Unnamed history**

1 shown

234.53 KB

**1: example\_grid\_file.xlsx**

To create a Flapjack file with split populations  
Click here

Click Run Workflow

Galaxy

111.93.2.172:8083/workflow/run?id=ebfb8f50c6abde6d

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

Get Data

Send Data

Collection Operations

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

HTPG Tools

NGS

Population Tools

Workflows

- All workflows
- HTPG/Intertek-Flapjack\_analysis

### Workflow: HTPG/Intertek-Flapjack\_analysis

Run workflow

History Options

Send results to a new history

Yes No

1: excel to txt extractor (Galaxy Version 1.0.0)

excel\_workbook

2: File3\_HTPG\_analysis\_file(Galaxy\_input).xlsx

2: File3\_HTPG\_analysis\_file(Galaxy\_input).xlsx

2: Intertek to Flapjack converter (Galaxy Version 1.0.0)

3: Intertek / Flapjack - SNP Sample File Processor (Galaxy Version 0.0.1)

4: Flapjack Split By Sample (Galaxy Version 0.0.1)

Flapjack genotype file

Output dataset 'output' from step 3

Flapjack phenotype file

Output dataset 'sampleFile\_sheet2' from step 1

Flapjack map file

Nothing selected

History

search datasets

Unnamed history

1 shown, 1 deleted

1.36 MB

2: File3\_HTPG\_analysis\_file(Galaxy\_input).xlsx

Galaxy

111.93.2.172:8083/workflow/run?id=ebfb8f50c6abde6d

☆🔥🛡️🟢📁💬🔦📧NEW📧

Galaxy

Analyze DataWorkflowShared DataVisualizationHelpUser

Using 5.9 MB

Tools

search tools

Get Data

Send Data

Collection Operations

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Fetch Alignments

Statistics

Graph/Display Data

HTPG Tools

NGS

Population Tools

Workflows

All workflows

HTPG/Intertek-Flapjack\_analysis

Successfully invoked workflow **HTPG/Intertek-Flapjack\_analysis**.

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Unnamed history

2 shown, 4 deleted, 2 hidden

2.63 MB

8: Flapjack Split By Sample on data 4 and data 7

2: File3\_HTPG\_analysis\_file(Galaxy\_input).xlsx

Click on  
the top file

Galaxy

111.93.2.172:8083/workflow/run?id=ebfb8f50c6abde6d

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Using 5.9 MB

Tools

search tools

[Get Data](#)

[Send Data](#)

[Collection Operations](#)

[Text Manipulation](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Convert Formats](#)

[Extract Features](#)

[Fetch Sequences](#)

[Fetch Alignments](#)

[Statistics](#)

[Graph/Display Data](#)

[HTPG Tools](#)

[NGS](#)

[Population Tools](#)

**Workflows**

- [All workflows](#)
- [HTPG/Intertek-Flapjack\\_analysis](#)

Successfully invoked workflow **HTPG/Intertek-Flapjack\_analysis**.

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

**Unnamed history**

2 shown, 4 [deleted](#), 2 [hidden](#)

2.63 MB

**8: Flapjack Split By Sample on data 4 and data 7**

1.0 MB

format: **flapjack**, database: ?

Map/marker cache created in 1ms

Map All Chromosomes has length 9.0

Importing traits from /data/galaxy-data/galaxy/database/tmp/tmpDG

No SQLite format 3 header

DB Serialization Cache: 0ms

Project saved in 736ms

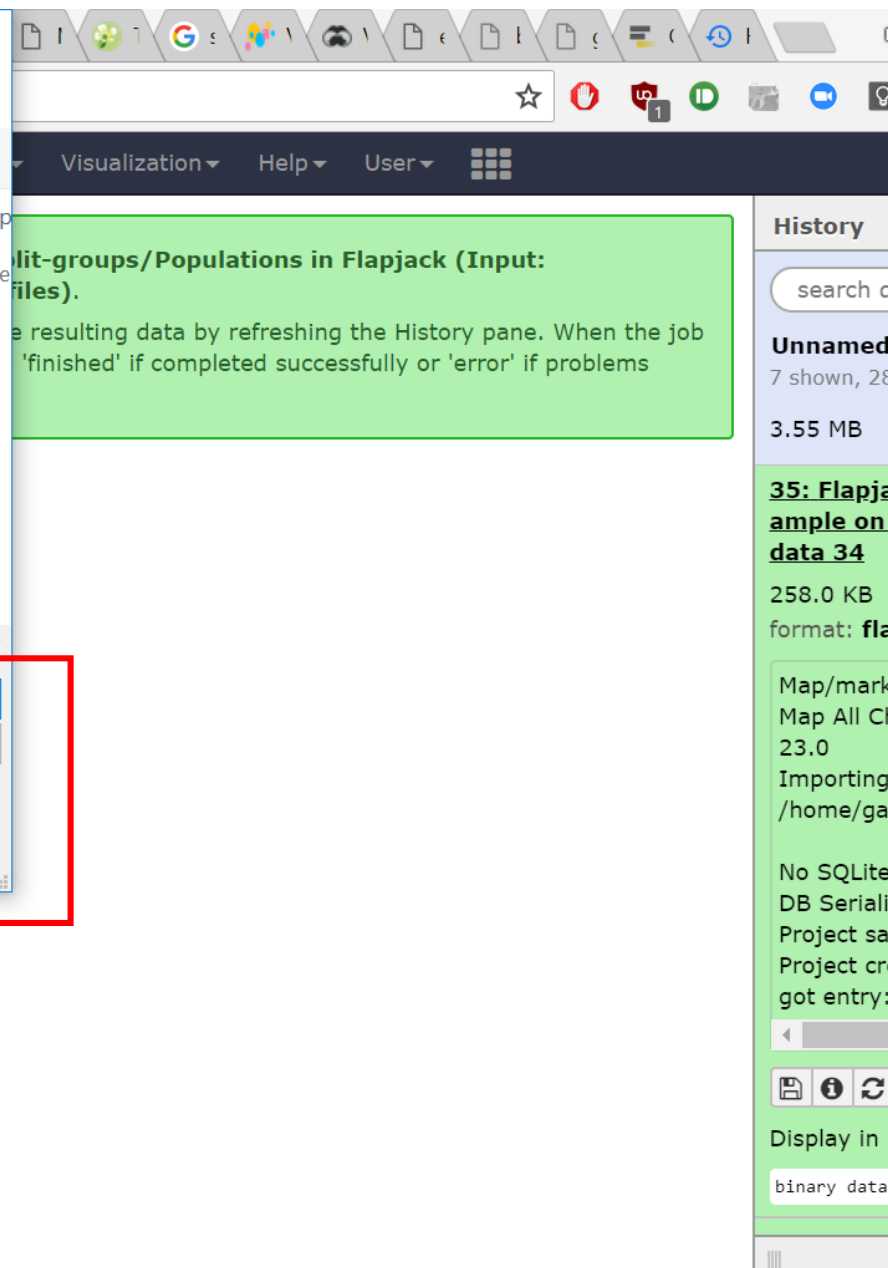
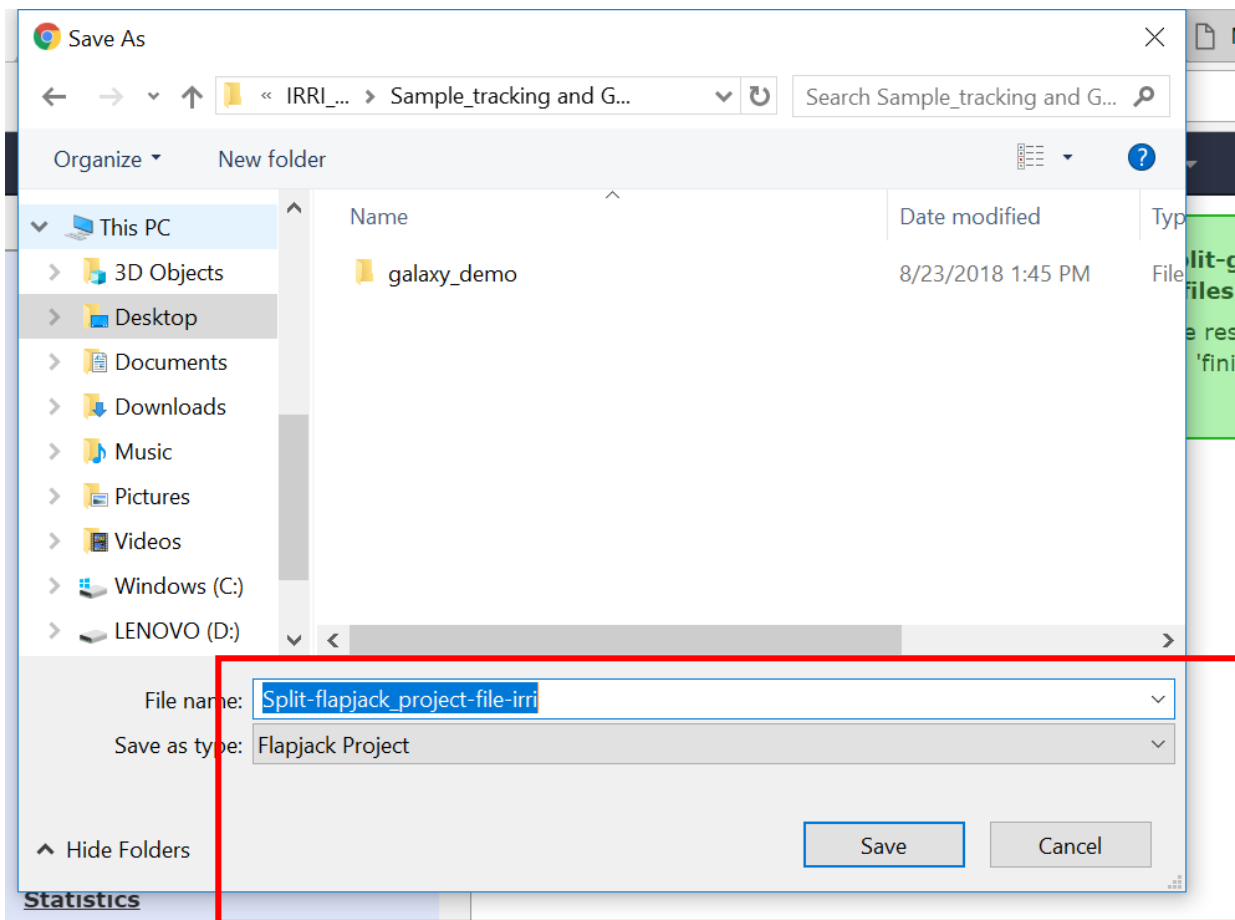
Project created

got entry: null

Project

Display in Flapjack: View (Help)

Click on Download icon and save file



Split populations

Split-flapjack\_project-file-irri.flapjack - Flapjack - 1.18.06.29

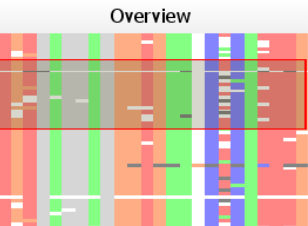
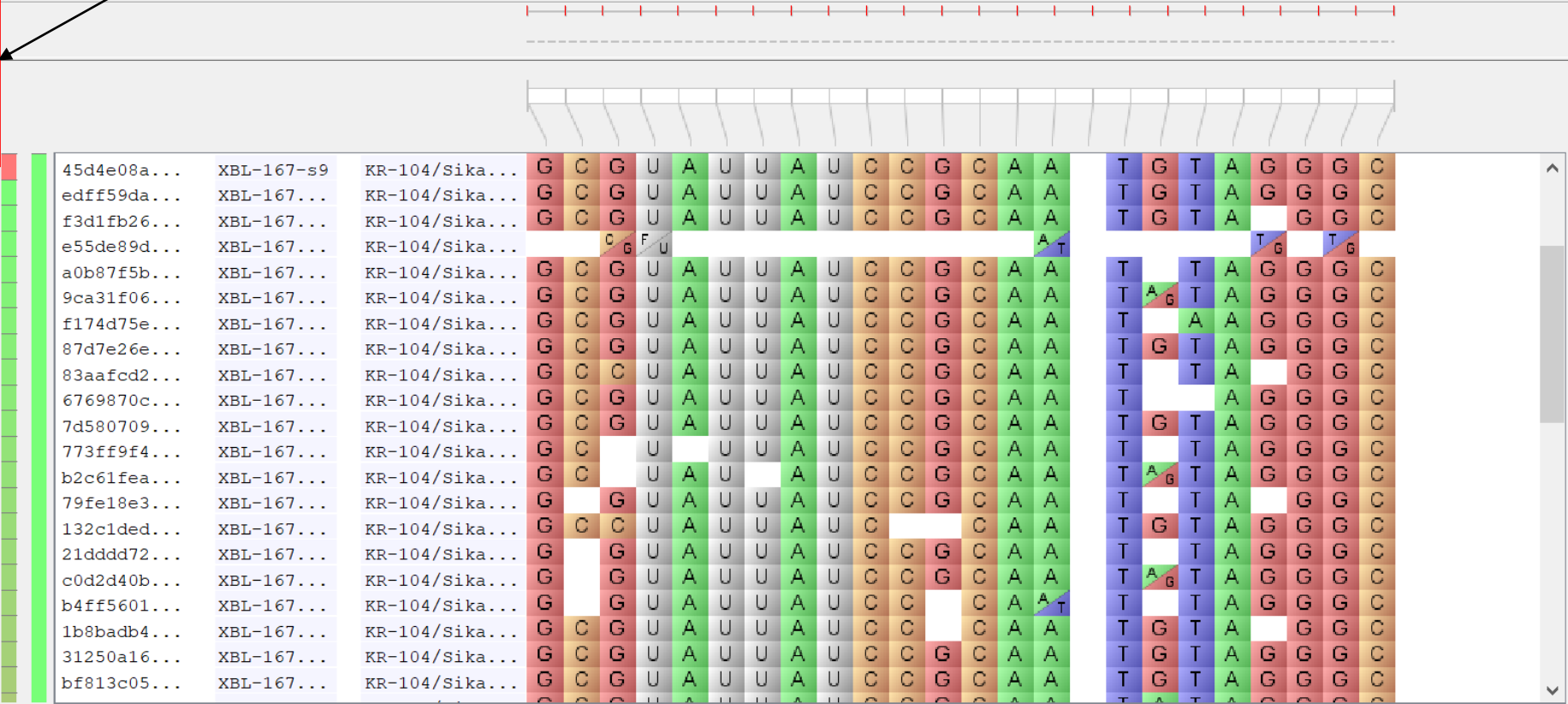
File Edit View Visualization Analysis Data Help

New Project Open Project Import Data Find Genotypes Chromosomes

**Data Sets**

- xbl-2018-P3 60x24
  - Trait Data
  - Default View
- RRN-2018-P1 40x24
  - Trait Data
  - Default View

Chromosome: All Chromosomes 60 lines, 24 markers, length: 23



Line:  
Marker:  
Genotype:

Zoom:

Flapjack Tip: Quickly track locations of interest by right-clicking on the display and selecting 'Bookmark location'

22x29, 4C, 7T, 102.00MB



Next step: Flapjack analysis

<https://media.readthedocs.org/pdf/flapjack/latest/flapjack.pdf>