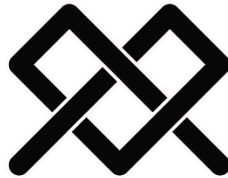


Workshop: Introduction to *Pretzel*



Pretzel

Enabling the seamless connection of plant genotypes to
research and breeding knowledge

Duration: 2 hours

Format: Hands-on with guided exercises

Prerequisites:

- An interest in using digital tools for genomic data analysis
- Your own laptop

Learning Objectives:

- Understand what *Pretzel* does and when to use it
- Load a custom QTL dataset
- Load genotype data and explore allelic diversity
- Search for AGG accessions containing a specific haplotype
- Find a genomic region of interest using BLAST
- How to combine different types of data (genomes, genetic maps, QTLs and genes)



The Australian Grains Genebank Strategic Partnership is a \$30M joint investment between the Victorian State Government and Grains Research and Development Corporation (GRDC) that aims to unlock the genetic potential of plant genetic resources for the benefit of the Australian grain growers. The Partnership is genotypically characterising the AGG plant genetic resource collection and building novel digital tools to improve accessibility and utility of the genebank for industry.

Contents

Introduction to AGG <i>Pretzel</i>	4
What is <i>Pretzel</i> ?.....	4
Why use <i>Pretzel</i> ?	5
<i>Pretzel</i> documentation	6
Workshop Overview	7
Workshop tasks	8
Starting out in <i>Pretzel</i>	8
Task 1: Log in to <i>Pretzel</i> and complete Data Licence Agreement.....	8
<i>Pretzel</i> layout.....	10
Entry points into <i>Pretzel</i>	11
Background information relevant to workshop tasks	12
Upload a QTL to <i>Pretzel</i>	13
Task 2: Prepare a custom QTL upload file.....	13
Task 3: Upload the Excel file to <i>Pretzel</i>	15
Task 4: Visualise the new QTL dataset.....	16
Compare positions of QTLs by comparative mapping	21
Task 5: Load a second QTL dataset	21
Task 6: Load anchored markers for comparative mapping.....	23
Axis Title menu	25
Compare position of a QTL with a reported gene	27
Task 7: Identify a gene of interest using feature search	27
Task 8: Identify the equivalent gene in a different genome based on comparative mapping	29
Explore diversity in a QTL region contributing to plant height	35
Task 9: Load genotype data	35
Task 10: Explore diversity in the region of interest	37
Task 11: Sort AGG accessions in the genotype table based on allele calls.....	40
Task 12: Filter AGG accessions with a defined haplotype of interest	41
Explore diversity around a locus containing a flowering time gene orthologue.....	46
Task 13: Perform a BLAST search	46

Task 14: Explore gene annotations	50
Visualising genotype calls for specific accessions	56
Task 15: Visualise genotype data in QTL region for specific AGG accessions	57
Additional tasks for a challenge	67
References.....	68
Useful URLs	69
Appendix	70
<i>Pretzel</i> data upload templates	70
Overview	70
Metadata sheet	71
Dataset sheet.....	73
Troubleshooting.....	75
Contact Details	76

Introduction to AGG *Pretzel*

What is *Pretzel*?

Pretzel (<https://agg.plantinformatics.io/>) is an open-source, interactive web application designed to enable users to connect plant genotype data to research and breeder knowledge. It allows users to visualise and interact with genetic and genomic data without needing specialised bioinformatics skills.

As genotype data is generated through the Australian Grains Genebank (AGG) Strategic Partnership, the genotype data and related datasets (such as genomes, BLAST databases and marker positions) are being loaded into *Pretzel*.

Pretzel currently contains genotype data and accessory files for wheat, barley, chickpea, lentil and field pea, with lupin being released shortly. Data for faba bean, mungbean, sorghum, oat and Brassica is planned for release in 2025-26.

Pretzel is a flexible platform built to enable users to analyse genetic and genomic data in many ways. Some of the most common uses include:

- Visualise and explore genetic and genomic datasets, including AGG genotype data
- Search for accessions with specific alleles or haplotypes, including trait-linked markers
- Compare marker order across genetic maps and genome assemblies
- Align QTLs, markers and genes across different genome assemblies
- Investigate pan-genomes
- Explore genes underlying QTLs
- Find alternatives for trait-linked markers
- Compare own private data with genotype data for accessions in the AGG

Why use *Pretzel*?

There are many digital tools available, each with their own advantages and disadvantages.

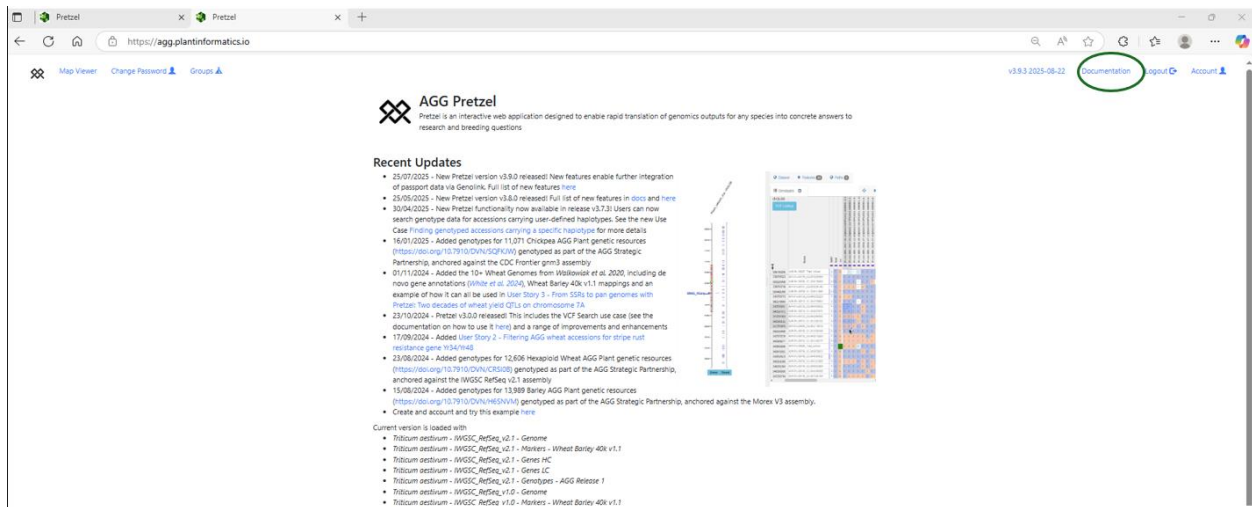
Pretzel does not replace tools such as genome browsers, data repositories or BLAST servers, but is designed to complement these tools.

What makes *Pretzel* different:

- Highly interactive with broad utility
- Access and interact with genotype data for AGG accessions
- Upload custom, private datasets e.g. genetic maps, QTLs, genes, markers, genotype data
- Group permissions enable private data sharing with specific users
- Integrates different types of data, including legacy datasets
- Connects with external tools e.g. Germinate, Gigwa
- Built with industry via co-design and user feedback
- Online documentation and training available to support industry use

Pretzel documentation

Online documentation is available at <https://docs.plantinformatics.io/>.



Three types of documentation are available:

- **Basic Functions**
 - Describes application layout and core features
- **Use Cases**
 - Provides instructions to perform discrete *Pretzel* functions
- **User Stories**
 - Examples of how *Pretzel* can be used to answer complex research and breeding questions

Workshop Overview

Today's workshop will cover:

- Layout of the *Pretzel* GUI
- Uploading a QTL as a private dataset
- Comparing QTLs anchored to different genomes
- Exploring genotypic diversity in a QTL region
- Exploring annotated genes in a QTL interval
- Performing a feature search for a known gene
- Find similar genes across different genomes
- Finding chickpea accessions from the AGG that contain a haplotype of interest
- Comparing allele calls for specific chickpea AGG accessions
- Performing a BLAST search

Optional exercises if time permits include:

- Exploring accession passport data with Genolink

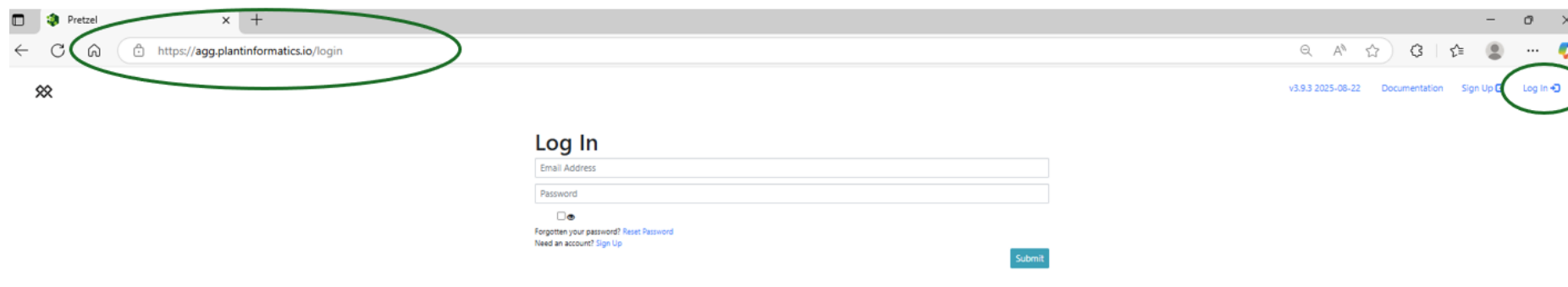
Workshop tasks

Starting out in *Pretzel*

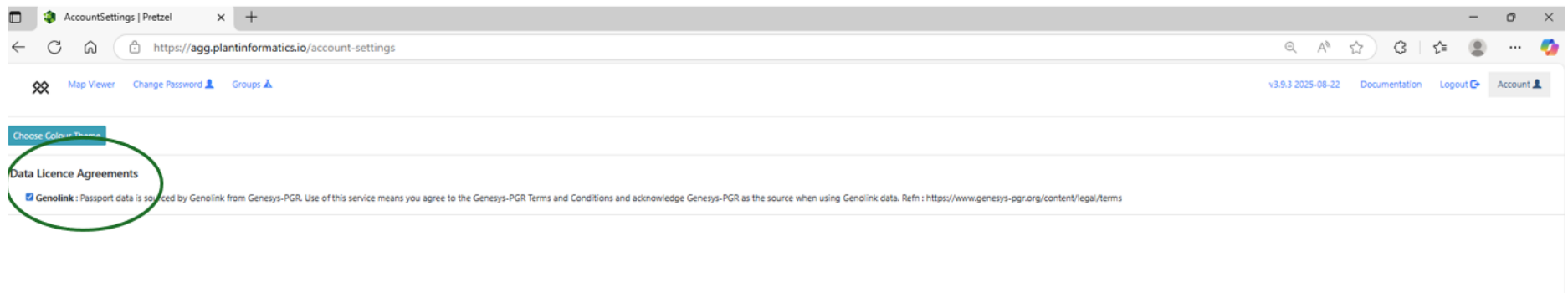
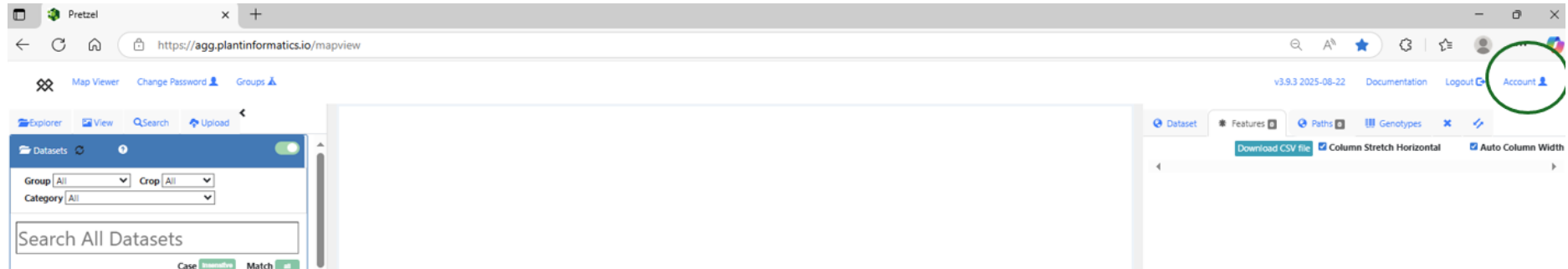
Task 1: Log in to *Pretzel* and complete Data Licence Agreement

If you haven't already requested a *Pretzel* account, click the '**Sign Up**' link.

Otherwise, continue to log in with the '**Log In**' link.



Click on 'Account' then tick the 'Genolink' box to agree to the Genesys-PGR Terms and Conditions .



Pretzel layout

The *Pretzel* screen consists of three main sections:

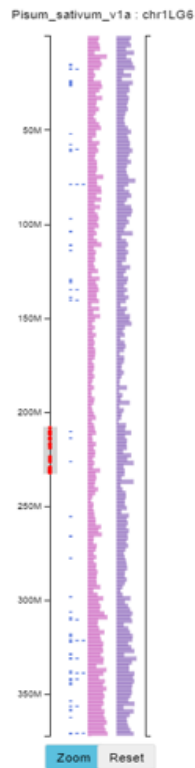
The screenshot displays the Pretzel web application interface, divided into three main sections by vertical dashed orange lines:

- 1 Datasets**: The left sidebar contains a 'Datasets' panel with filters for Group, Crop, and Category. It includes a search bar, tabs for Trait, Ontology, QTL, Genome, Alignment, and VCF, and a list of datasets including 'Cheng_2024' and various Cicer arietinum datasets.
- 2 Canvas**: The central area is a large empty space for data visualization.
- 3 Metadata**: The right sidebar shows the 'Cheng_2024' dataset details, including Namespace, Parent, Created, Last Updated, Public status, Read Only, tags, Group, Group Owner, and API Host. It also displays a 'Meta' section with a list of metadata items.

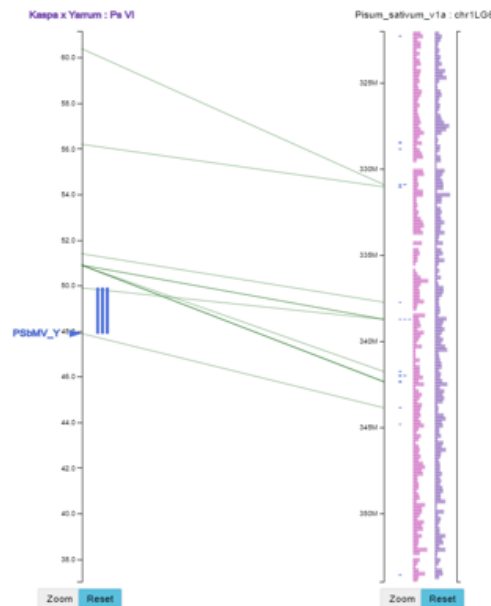
Entry points into *Pretzel*

There are multiple ways for users to figure out which part of the genome to start interacting with.

1
Genomic co-ordinates



2
Feature of interest



3
Gene or marker ID

4
BLAST

Background information relevant to workshop tasks

The exercises undertaken in this workshop are associated with genes involved in phenology and plant height in chickpea. Modern domesticated varieties have a narrow genetic base, which can limit their adaptability and resilience. Globally, there is increasing interest in mining existing diversity in crop wild relatives and diverse germplasm to help meet challenges in food security, sustainability and climate adaptation. However, breeders are often hesitant to utilise wild crops and diverse germplasm due to linkage drag and reduced agronomic performance, since it can be difficult to separate beneficial traits from undesirable ones. For example, genes associated with flowering time have implications for adaption to different climates and growing seasons, and genes associated with plant height can have a big impact on plant architecture and yield. Importantly, the presence of some wild alleles could result in flowering at disadvantageous times, or tall plants that can be impacted by lodging or reduced seed set. Having knowledge of which alleles are present is highly advantageous for selecting the right material for its desired purpose. The tasks in this workshop show users how to look at genetic diversity among AGG accessions around QTLs and candidate genes for flowering time and plant height and show how research knowledge can be used to identify chickpea accessions from the AGG that are more likely to contain wild introgressions. These examples demonstrate how users can make more informed selections of PGRs in research and breeding.

Upload a QTL to *Pretzel*

This task uses data published by Malik et. al. (2023) investigating diversity associated with plant height in chickpea. Information for QTLs on plant height and lignin content was extracted from the publication and entered into an Excel file in advance, with QTL positions reported against the CDC Frontier v1 (kabuli) genome assembly (Varshney et al, 2013). This pre-filled file has been provided to you. Today we will edit this file to ensure the name of the data is unique and then upload it into Pretzel for further investigation. After uploading the custom QTLs, the data will remain private, and you can easily locate the position of the QTLs whenever you log into *Pretzel*.

Task 2: Prepare a custom QTL upload file

1. Rename the *Pretzel* QTL upload file called '**Pretzel_Workshop_chickpea_PH_QTL_Upload-XX.xlsx**', replacing 'XX' with your initials e.g. Pretzel_Workshop_chickpea_PH_QTL_Upload-KF.xlsx
2. Open the upload file and rename the sheet '**QTL|Ca_Malik_PH_QTLs-XX**', replacing 'XX' with your initials e.g. QTL|Ca_Malik_PH_QTLs-KF.
3. Update the dataset name beside the '**Field**' column in the '**Metadata**' sheet.
4. Update the '**displayName**' field '**Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - XX**', replacing 'XX' with your initials.
5. Add your name to the '**Curator**' field.
6. Save the QTL upload file.

Pretzel_Workshop_chickpea_PH_QTL_Upload-XX.xlsx

1. Rename Pretzel QTL upload file name

Field	Value
Species	Cicer arietinum
Crop	Chickpea
parentName	Chickpea CDCFrontierGA_1
displayName	Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - XX
licensing	none
Citation	Malik N, Basu U, Srivastava R, Daware A, Ranjan R, Sharma A, Thakro V, Mahanty JK, Jha UC, Tripathi S, Tyagi AK, Parida SK. 2023. Natural alleles of Mediator subunit genes modulate plant height in chickpea. Plant J. 116:1271-1292. doi: 10.1111/tpj.16423.
Curator	Name
Comments	Genome coordinates for plant height and lignin content QTLs taken from Table 1 of the publication.

2. Rename the 'Field' by incorporating your initials to make it unique, ensuring it matches the worksheet name below

3. Rename 'display name', incorporating your initials to make it unique

4. Add the Dataset curator name

5. Rename worksheet by incorporating your initials, ensuring it matches the 'Field' name above

Metadata QTL|Ca_Malik_PH_QTLs-XX

Task 3: Upload the Excel file to *Pretzel*

By following the steps below, the QTL will be uploaded as a private dataset.

1. Select the 'Upload' tab

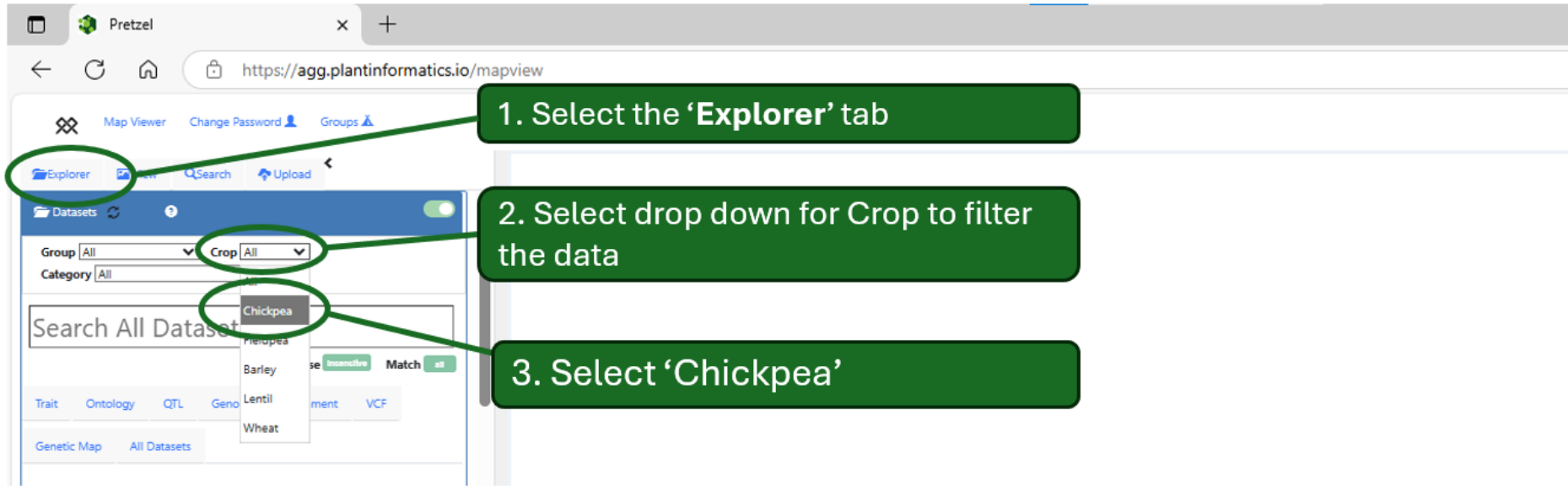
2. Select 'Choose File' to upload the QTL file to Pretzel

3. Success notification confirms uploading of data

The interface shows the Pretzel web application at <https://agg.plantinformatics.io/mapview#left-panel-upload>. The 'Upload' tab is selected in the top navigation bar. The 'Data Specification' panel on the left shows the 'Upload Spreadsheets' section, where the 'Choose File' button is highlighted. A success notification at the bottom of the page states: 'Success Dataset 'Ca_Malik_Pht_QTLs-10F' Uploaded successfully'.

Task 4: Visualise the new QTL dataset

Next we will load the new custom dataset so we can see the location of the QTL on chromosome 4, then adjust the display to colour-code the QTLs according to the type of traits that were defined in the Excel file.



The screenshot shows the Pretzel web application interface. The browser address bar displays <https://agg.plantinformatics.io/mapview>. The interface includes a top navigation bar with links for 'Map Viewer', 'Change Password', and 'Groups'. Below this, a secondary navigation bar contains 'Explorer', 'New', 'Search', and 'Upload'. The 'Explorer' tab is highlighted with a green circle and a callout box. The 'Datasets' section features a 'Group' dropdown set to 'All' and a 'Crop' dropdown set to 'All', both highlighted with green circles and callout boxes. A search bar labeled 'Search All Dataset' is present, with a dropdown menu showing 'Chickpea' as the selected option, also highlighted with a green circle and callout box. Below the search bar, there are tabs for 'Trait', 'Ontology', 'QTL', 'Geno', 'Lentil', 'ment', and 'VCF'. At the bottom, there are links for 'Genetic Map' and 'All Datasets'.

1. Select the 'Explorer' tab
2. Select drop down for Crop to filter the data
3. Select 'Chickpea'

The screenshot shows the Pretzel web application interface. The left panel displays a search results list with a dataset named "Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - KF". The right panel shows the details for the dataset "Ca_Malik_PH_QTLs-KF".

4. Type 'Malik' in the Search bar

Note!
Search has been done in a case insensitive manner and to match all words in search bar

Note!
Filtered datasets will appear in the left panel.

5. Click on the black '+' icon on the left side of dataset name to show available chromosomes

Dataset Details:

- Dataset: Ca_Malik_PH_QTLs-KF
- Namespace: Chickpea_CDCFrontierGA_1
- Parent: Chickpea_CDCFrontierGA_1
- Created: 20/08/2025, 2:05:18 pm
- Last Updated: 20/08/2025, 2:05:18 pm
- Public: false
- Read Only: true
- tags: QTL
- Group: All
- apiHost: https://agg.plantinformatics.io
- Meta: View Edit
- Metadata (8):
 - Species: Cicer arietinum
 - Crop: Chickpea
 - displayName: Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - KF
 - licensing: none
 - Cita: Malik N, Basu U, Srivastava R, Daware A, Ranjan R, Sharma A, Thakro V, Mohanty JK, Jha UC, Tripathi S, Tyagi AK, Parida SK. 2023. Natural alleles of Mediator subunit genes modulate plant height in chickpea. Plant J. 116:1271-1292. doi 10.1111/tpj.16423.
 - Curator: Name
 - Comment: Genome coordinates for plant height and lignin content QTLs taken from Table 1 of the publication.
 - type: QTL

The screenshot displays the Pretzel web application interface. On the left, the 'Datasets' panel shows a list of datasets with a search bar containing 'Malik'. A green callout box points to a '+' button next to 'Ca4', indicating it should be clicked to load the dataset. A second green callout box points to a padlock icon next to the same dataset, indicating it is private. The main area shows a chromosome plot for 'Chickpea_CDCFrontierGA_1 : Ca4' with a vertical axis ranging from 0.0M to 45M. A blue diamond marker is visible on the chromosome. A third green callout box points to the right-hand side of the plot, indicating where to check for the QTL display. On the right, the 'Dataset' panel shows the metadata for 'Ca_Malik_PH_QTLs-KF', including details like Namespace, Parent, Created, Last Updated, Public status, Read Only status, tags, Group, apiHost, and Meta information. A fourth green callout box points to the 'Meta' section, indicating where to check that the metadata is correct.

6. Click on the '+' green button on the left side of Ca4 to load the dataset. It changes to a '-' symbol once loaded.

Padlock icon indicates a private dataset

7. Check that the metadata for the dataset is correct

8. Check that your QTL is displayed on the right-hand side of the chromosome axis

The screenshot below shows how different fields contained in the Excel QTL upload file are incorporated into the *Pretzel* Explorer panel, canvas (i.e. on the chromosome axis), and in the Dataset's metadata section. Refer to the Appendix for more information on the Excel template file for uploading custom QTLs.

The screenshot displays the Pretzel web application interface, showing the Explorer panel, a chromosome map, and the Dataset metadata section. Green arrows point to specific elements, explaining how fields from an Excel QTL upload file are mapped:

- Truncated version of the 'parentName'**: Points to the search bar in the Explorer panel where 'Malik' is entered.
- 'Field' for dataset name**: Points to the dataset name 'Ca_Malik_PH_QTLs-KF' in the Dataset panel.
- 'parentName'**: Points to the 'Parent' field value 'Chickpea_CDCFrontierGA_1' in the Dataset panel.
- 'QTL' tag is derived from the Excel Field name**: Points to the 'tags' field value 'QTL' in the Dataset panel.
- 'displayName'**: Points to the dataset entry 'Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - KF' in the Explorer panel.

The chromosome map shows two blue diamond markers on the Chickpea_CDCFrontierGA_1 chromosome, representing QTLs. The Dataset panel shows the following metadata:

```

Dataset: Ca_Malik_PH_QTLs-KF
Namespace:
Parent: Chickpea_CDCFrontierGA_1
Created: 20/08/2025, 2:05:18 pm
Last Updated: 20/08/2025, 2:18:44 pm
Public: false
Read Only: true
tags: QTL
Group: All
apiHost: https://agg.plantinformatics.io
Meta: View Edit
Metadata {8}
  Species: Cicer arietinum
  Crop: Chickpea
  displayName: Cicer arietinum - CDC_Frontier_v1.0 - QTL - Malik 2023 Plant height - KF
  licensing: none
  Cita: Malik N, Basu U, Srivastava R, Daware A, Ranjan R, Sharma A, Thakro V, Mohanty JK, Jha UC, Tripathi S, Tyagi AK, Parida SK. 2023. Natural alleles of Mediator subunit genes modulate plant height in chickpea. Plant J. 116:1271-1292. doi 10.1111/tpj.16423.
  Curator: Name
  Comments: Genome coordinates for plant height and lignin contents QTLs taken from Table 1 of the publication.
  type: QTL
Dataset Name Scope Interval
  
```

Next the QTL display will be adjusted to colour code the QTLs according to trait type.

9. Click on 'View' tab

10. Select 'Trait'

11. Select 'Show QTLs of all Traits'

12. Check that plant height and lignin content QTLs are displayed in different colours

13. Display a label for each of the plant height (blue) QTLs by first clicking on the QTL (diamond), then click on the arrow that appears on the axis.

Compare positions of QTLs by comparative mapping

Next, we will load a QTL for plant height on chromosome 4 reported in an independent publication by Mugabe et al. (2019) to see whether any of the QTLs from Malik et al (2023) loaded above are detecting the same locus. Since QTLs from the two different publications have been anchored to different genome assemblies, this will be accomplished through a comparative mapping approach using the positions of Illumina Infinium™ Pulsus 30K v1.0 BeadChip markers that have been anchored to both assemblies and loaded into *Pretzel*.

Task 5: Load a second QTL dataset

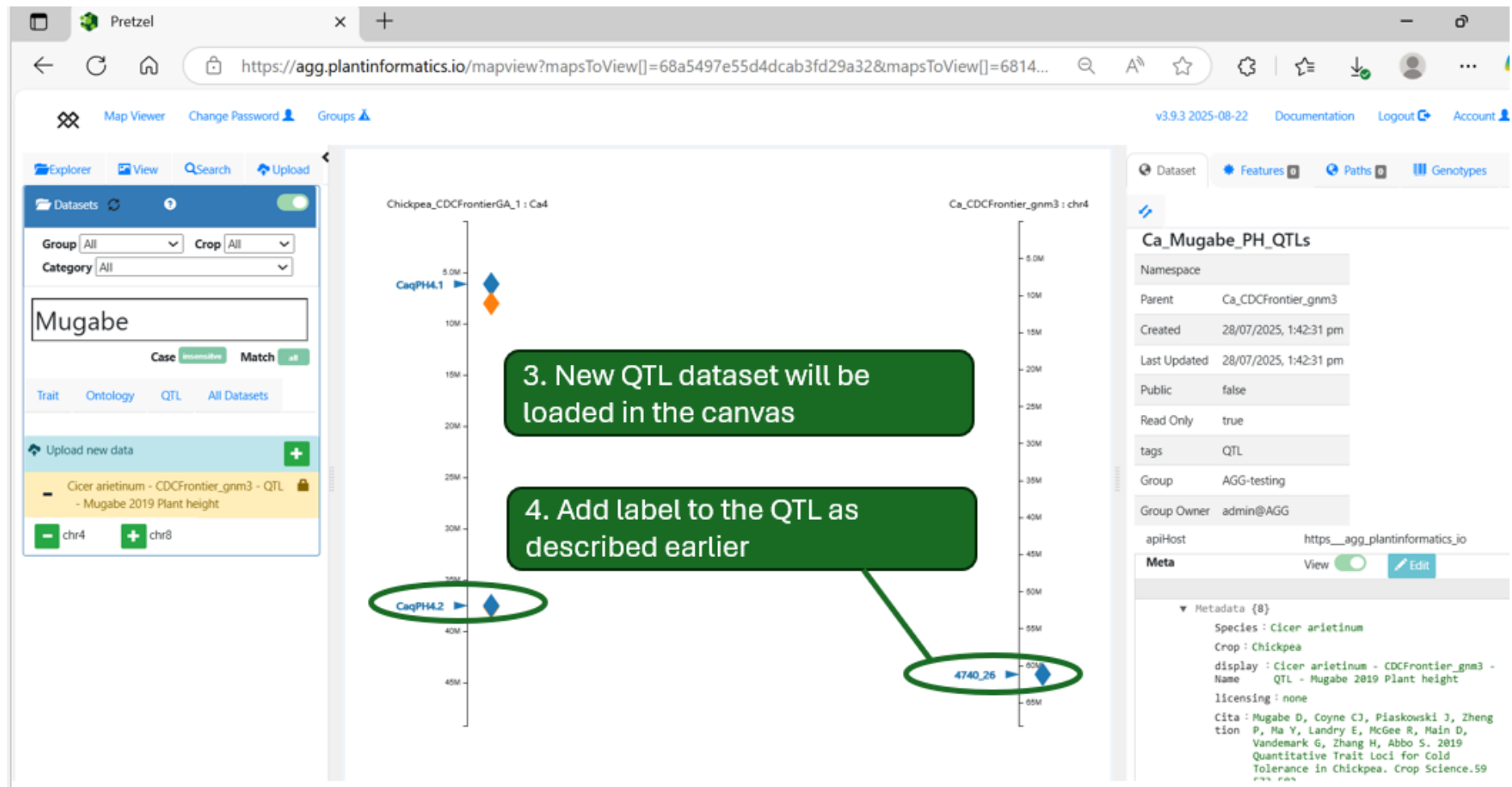
1. Type 'Mugabe' in the Search bar

2. Click on the black '+' icon on the left side of dataset name to show available chromosomes, then click on the '+' green button on the left side of chr4

The screenshot displays the Pretzel web interface. On the left, the 'Datasets' panel shows a search bar with 'Mugabe' entered. Below the search bar, there are filters for 'Group' and 'Crop'. A list of datasets is shown, including 'Cicer arietinum - CDCFrontier_gnm3 - QTL - Mugabe 2019 Plant height'. A green circle highlights a '+' icon next to the dataset name. Below the dataset list, there are buttons for '+ chr4' and '+ chr8'. A green callout box points to the '+' icon and the '+ chr4' button. In the center, a map view shows a chromosome with a blue diamond marker labeled 'CaqPH4.1' and an orange diamond marker labeled 'CaqPH4.2'. On the right, the 'Dataset' panel shows details for 'Ca_Mugabe_PH_QTLs', including parent, created, last updated, public status, read only, tags, group, group owner, api host, and meta data.

Dataset Details:

- Dataset: Ca_Mugabe_PH_QTLs
- Namespace: Ca_CDCFrontier_gnm3
- Parent: Ca_CDCFrontier_gnm3
- Created: 28/07/2025, 1:42:31 pm
- Last Updated: 28/07/2025, 1:42:31 pm
- Public: false
- Read Only: true
- tags: QTL
- Group: AGG-testing
- Group Owner: admin@AGG
- apiHost: https://agg_plantinformatics.io
- Meta: View Edit
- Metadata (8):
 - Species: Cicer arietinum
 - Crop: Chickpea
 - displayname: Cicer arietinum - CDCFrontier_gnm3 - QTL - Mugabe 2019 Plant height
 - me: Plant height
 - licensing: none
 - Cita: Mugabe D, Coyne CJ, Piaskowski J, Zheng P, Ma Y, Landry E, McGee R, Main D, Vandemark G, Zhang H, Abbo S. 2019 Quantitative Trait Loci for Cold Tolerance in Chickpea. Crop Science 59: 573-582. <https://doi.org/10.2135/cropsci2018.06.0504>
 - Curator: Kerrie Forrest - Agriculture Victoria
 - comm: SNP positions for QTLs were identified using BLAST against CDC ent's Frontier gnm3 assembly using SNP flanking sequences for plant height QTLs taken from Table S6 of the publication.
 - type: QTL



Based on the current display, the QTLs **CaqPH4.2** and **4740_26** towards the ends of the chromosomes might be detecting the same locus. We will investigate this further next.

Task 6: Load anchored markers for comparative mapping

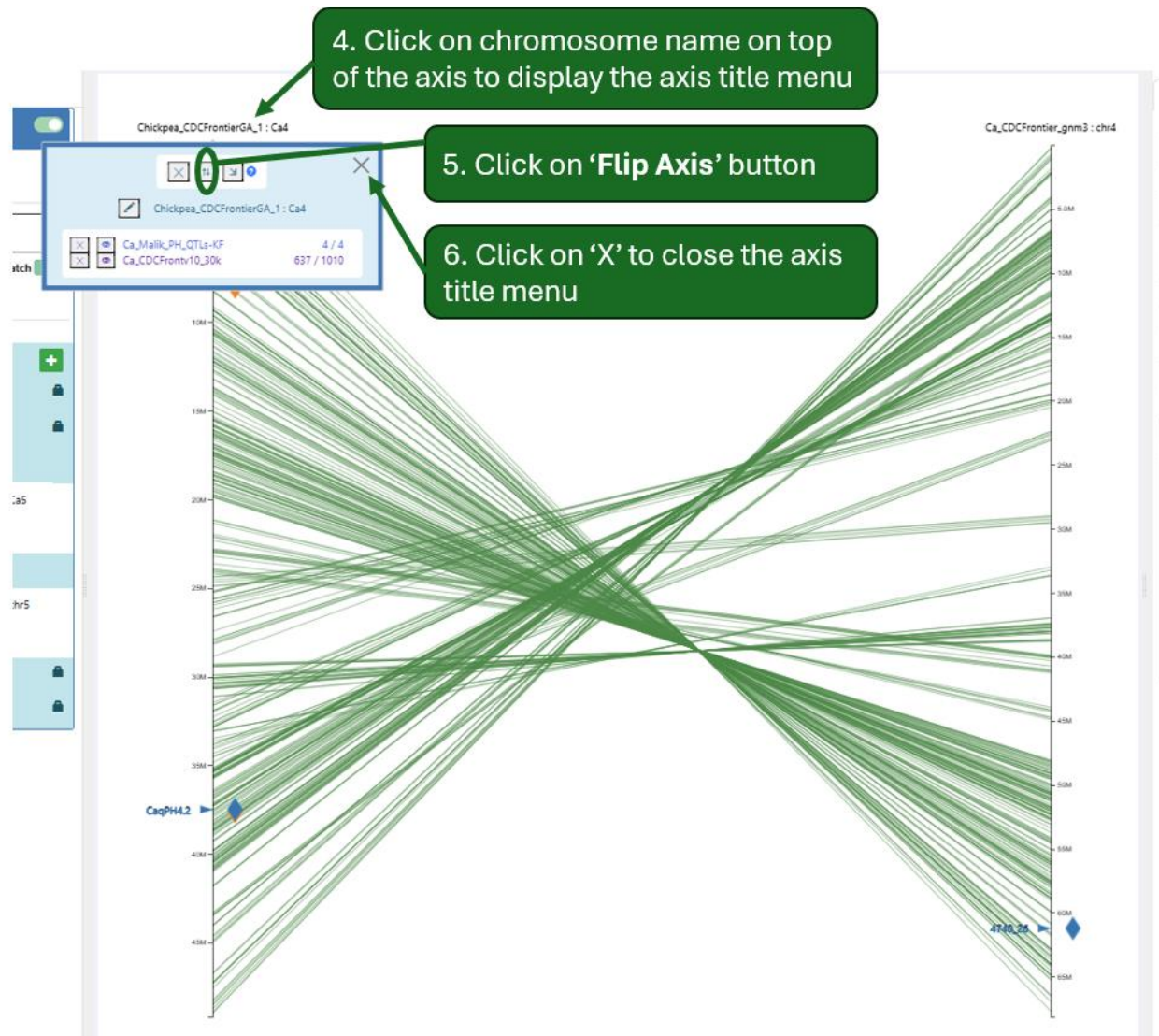
1. Click on the 'Explorer' tab then select 'Chickpea' from the 'Crop' dropdown menu

2. In the Search bar, type 'Markers' to filter the listed datasets

3. Click on the '+' green button on the left side of Ca4 and chr4 to load the 30K marker datasets for CDC_Frontier_v1.0 and CDCFrontier_gnm3, respectively

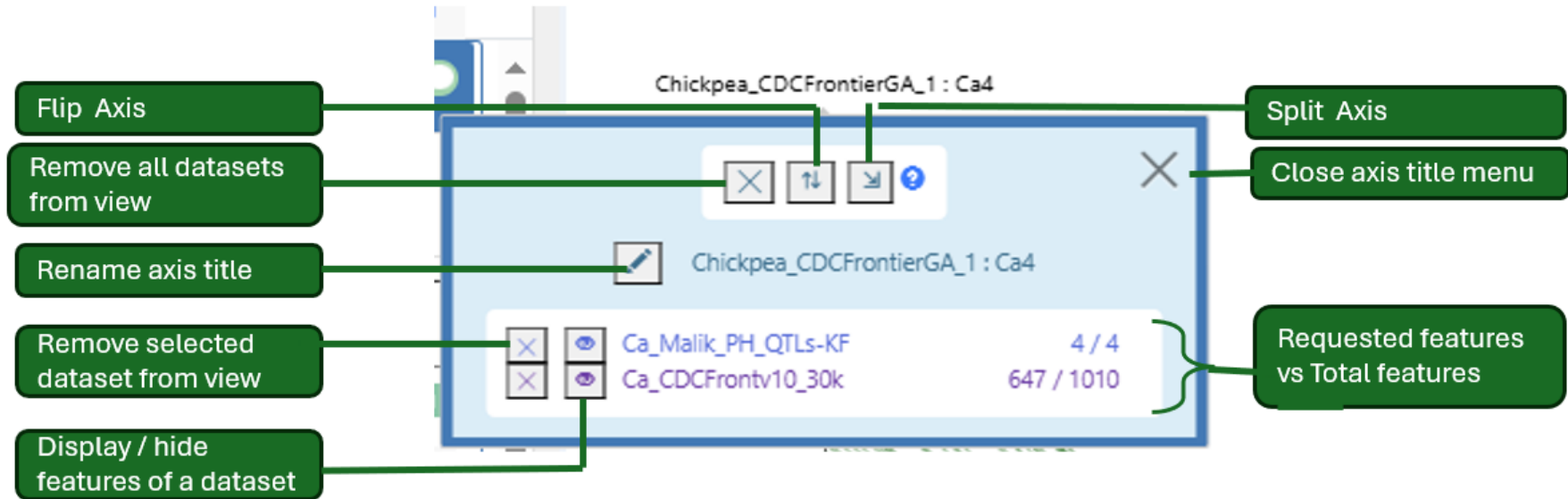
The screenshot shows the Pretzel web interface. The left sidebar has a 'Markers' tab selected. The 'Crop' dropdown menu is set to 'Chickpea'. The search bar contains 'Markers'. The list of datasets shows 'Cicer arietinum - CDC_Frontier_v1.0 - Markers - Pulses30k v1.0' and 'Cicer arietinum - CDCFrontier_gnm3 - Markers - Pulses30k v1.0'. The central map view shows a chromosome map with markers. The right sidebar shows the details for the 'Ca_CDCFront-gnm3_30k' dataset, including its parent, creation date, last updated date, public status, read-only status, and a list of metadata.

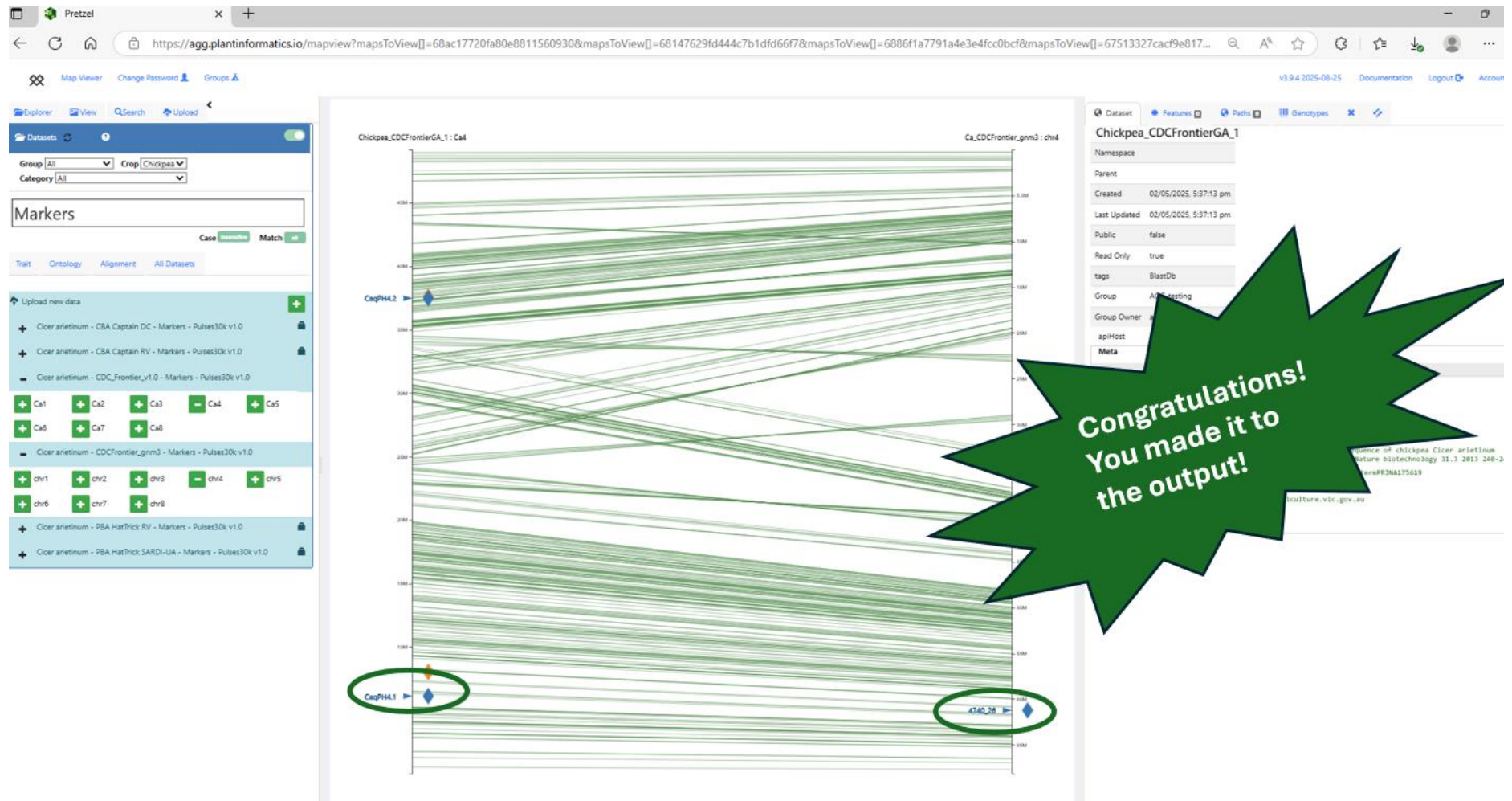
A green line drawn between the axes indicates that a marker with the same name is anchored to each of the two different versions of chromosome 4 displayed. In this case, marker ordering between the two versions of chromosome 4 is reversed, as indicated by the many green lines crossing over. One of the axes needs to be flipped. We will flip CDC Frontier v1.



Axis Title menu

The information below describes functionality that becomes available when the Axis Title menu is opened, and information about the loaded datasets.





We have now successfully compared the positions of QTLs from two different publications, one set anchored to CDC Frontier v1 (Varshney et al. 2013) and the other anchored to an independent genome assembly CDC Frontier gnm3 (Cook, 2022). After taking into account the different orientations of the chromosome sequences, it is now clear that QTLs **CaqPH4.1** and **4740_26** are likely detecting the same locus, while **CaqPH4.2** is detecting a different locus.

Compare position of a QTL with a reported gene

In this example, we will search for a gene called CaMED23 reported to be involved in plant height and compare its position to the published QTL regions being investigated. The gene ID '**Ca_15125**' in the CDC Frontier v1 genome annotation was reported in Malik et. al. (2023).

Task 7: Identify a gene of interest using feature search

1. Click on the 'Search' tab

Note!
The QTLs already labelled in the display will also appear in the Search field

2. Type 'Ca_15125' in the 'Features Input' field within 'Feature Search'

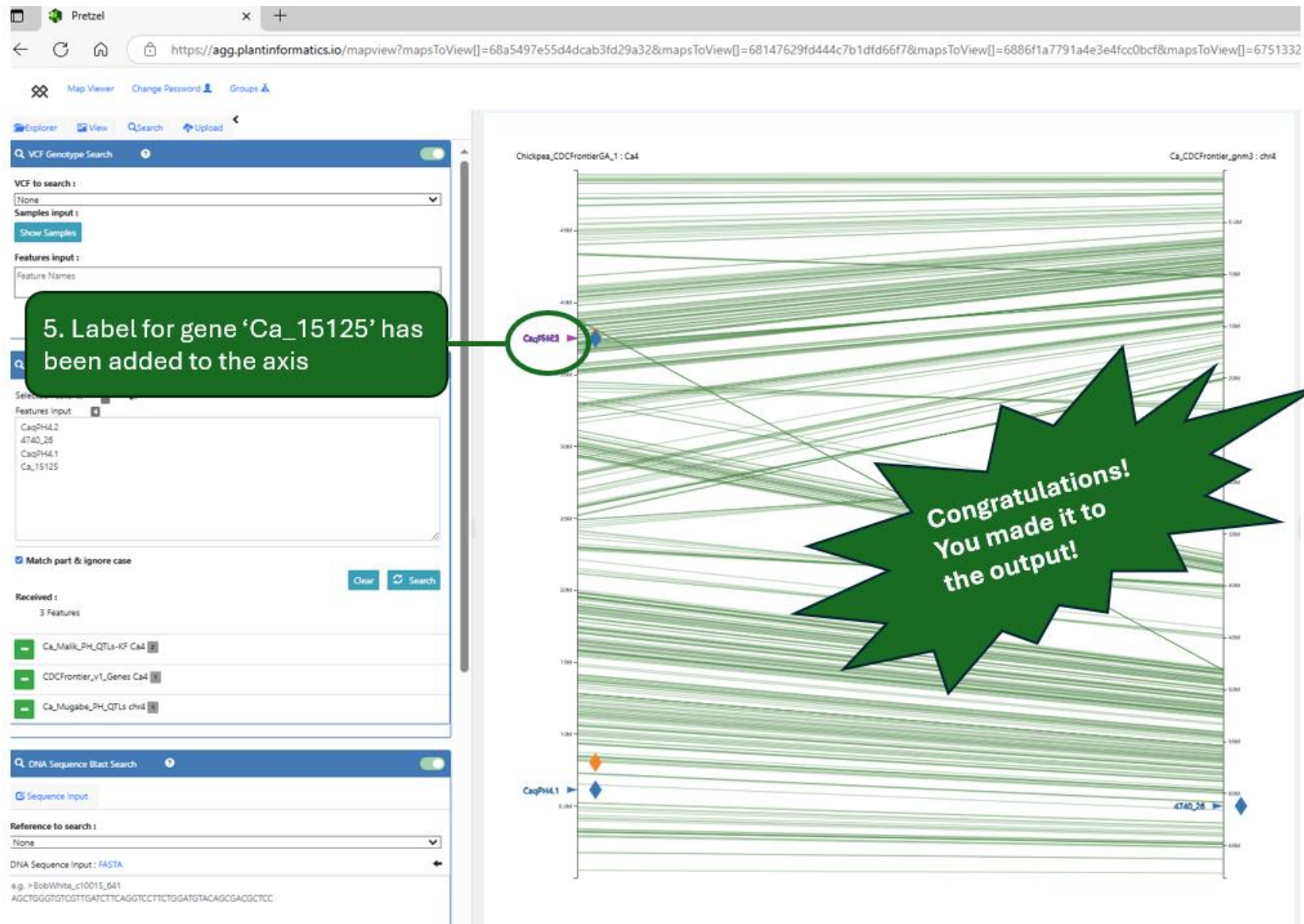
3. Click on 'Search'

Note!
All Pretzel datasets containing a feature with the ID 'Ca_15125' are listed below.

4. Click on the '+' button to load the dataset 'CDCFrontier_v1_Genes Ca4' to display the position of Ca_15125

Received:
3 Features

- Ca_Mugabe_PH_QTLs chr4
- + CDCFrontier_v1_Genes Ca4
- Ca_Mugabe_PH_QTLs chr4



The position of gene Ca_15125 in chromosome Ca4 of the CDC Frontier v1 assembly is now indicated with a purple arrow and label. The gene is located very close to QTL CaqPH4.2 from Malik et. al. (2023).

Task 8: Identify the equivalent gene in a different genome based on comparative mapping

Next we will identify the annotated gene for CaMED23 in a newer chickpea genome, CDC Frontier gnm3 (Cook, 2022). This will be accomplished through comparative mapping using Infinium™ Pulses 30K v1.0 BeadChip (Illumina) markers that have been anchored to both assemblies.

1. With your mouse pointer, click and drag along the axis to highlight the portion of the chromosome surrounding the Ca_15125 gene. Selected features will turn red.

2. Remove QTL labels by clicking on the arrow then on the diamond.

3. Click on the 'Zoom' button

Chickpea_CDCFrontierGA_1

Namespace

Parent

Created 02/05/2025, 5:37:13 pm

Last Updated 02/05/2025, 5:37:13 pm

Public false

Read Only true

tags BlastDb

Group AGG-testing

Group Owner admin@AGG

apiHost https://agg_plantinformatics.io

Meta View Edit

▼ Metadata (10)

- displayName: Cicer arietinum - CDC_Frontier_v1.0 - Genome
- species: Cicer arietinum
- crop: Chickpea
- DOI: https://doi.org/10.1038/nbt.2491
- cita: Varshney, Rajeev K., et al. Draft genome sequence of chickpea Cicer arietinum provides a resource for trait improvement. Nature biotechnology 31.3 2013 240-246.
- Data source: https://www.ncbi.nlm.nih.gov/bioproject/termPRJ NA175619
- uploaded by: Agriculture Victoria
- Contact: gabriel.keeble@gagnereagriculture.vic.gov.au
- Licensing: None
- type: Genome

Estimated Features in

Brush	Zoom	Dataset
2	4	Ca_Malik_PH_QTLs-KF
1010	1010	Chickpea_CDCFrontier_v10_30k
256.4	3469	CDCFrontier_v1_Genes

Map Viewer
Change Password
Groups

v3.9.4 2025-08-25
Documentation
Logout
Account

VCF Genotype Search

VCF to search :
None

Samples input :
[Show Samples](#)

Features input :
Feature Names

[Search](#)

Feature Search

Selected Features: 116

Features Input: 1

Ca_15125

☒ Match part & ignore case

[Clear](#) [Search](#)

Received :
1 Features

[-](#) CDCFrontier_v1_Genes Ca4

DNA Sequence Blast Search

[Sequence Input](#)

Chickpea_CDCFrontierGA_1:Ca4

Ca_CDCFrontier_gnm3:chr4

Zoom [Reset](#)

Ca_CDCFrontier_gnm3

Namespace

Parent

Created 05/12/2024, 3:59:19 pm

Last Updated 05/12/2024, 3:59:19 pm

Public true

Read Only true

tags BlastDb

apiHost https://agg.plantinformatics.io

Meta [View](#) [Edit](#)

▼ Metadata {11}

```

displayname: Cicer arietinum - CDCFrontier_gnm3 - Genome
species: Cicer arietinum
citati: Genome assembly v3 for genotype CDCFrontier from
on the Cook Lab, UC-Davis
Data: https://data.legumeinfo.org/Cicer/arietinum/genomes/CDCFrontier_gnm3_QTOP/
Uploaded by: Agriculture Victoria
Contact: gabriel.keeble-gagnereagriculture.vic.gov.au
Licensing: None
type: Genome
Crop: Chickpea
__comment__: TBC
Category: core
          
```

DNA sequence for brushed interval Lookup Region limit is 100kb

Estimated Features in

Brush	Zoom	Dataset
0	1	Ca_Mugabe_PH_QTLs
1007	1007	Chickpea_CDCFrontier_gnm3_30k

4. Repeat steps 1-3 to zoom in on equivalent markers in the gnm3 chromosome on the right-hand side

The screenshot displays the Chickpea Genome Explorer interface. On the left, the 'Explorer' tab is selected. Below it, the search bar contains 'Genes gnm3'. Under the 'Upload new data' section, the dataset 'Cicer arietinum - CDCFrontier_gnm3 - Genes' is listed with a black '+' icon next to it. Below this, a grid of chromosome selection buttons (chr1 to chr8) is shown, with 'chr4' highlighted. In the center, a genomic track view shows the 'Ca_CDCFrontier_gnm3 : chr4' track with a blue diamond marker at approximately 37.50M. On the right, a detailed view of the 'CDCFrontier_g3_Genes' dataset is shown, including its metadata and a list of genes.

5. Click on the 'Explorer' tab

6. Type 'Genes gnm3' in the Search bar

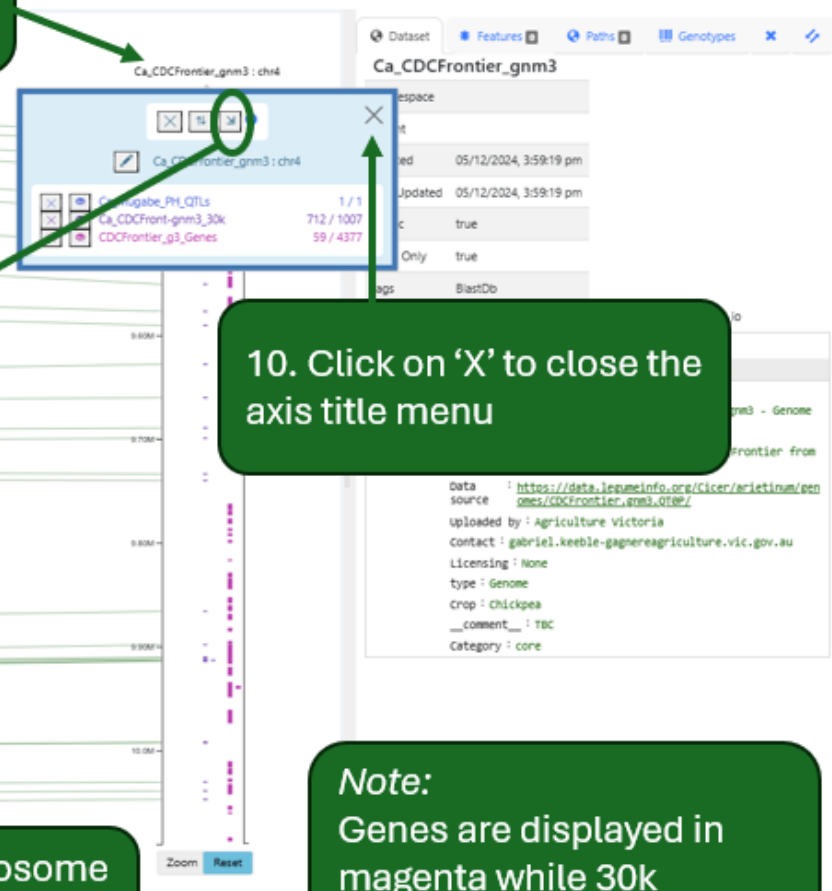
7. Click on the black '+' icon on the left side of dataset name 'Cicer arietinum - CDCFrontier_gnm3 - Genes' then load chr4



8. Click on chromosome name on top of the axis to display the axis title menu

9. Click on 'Split Axis' button to show features i.e. markers with genotype calls and genes

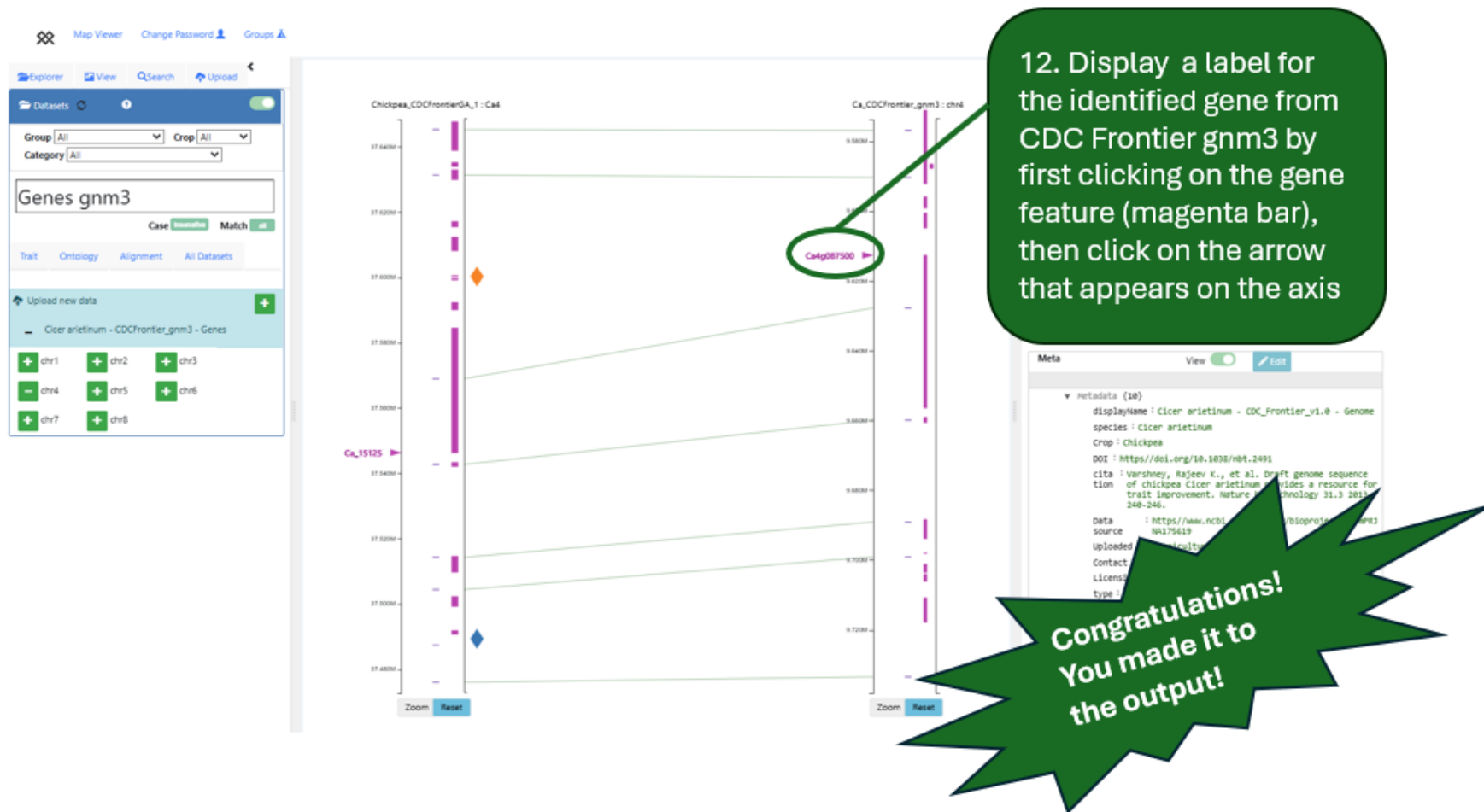
11. Repeat steps 7-9 for chromosome Ca4 from CDC Frontier v1 (left-hand side of canvas)



10. Click on 'X' to close the axis title menu

Note:
Genes are displayed in magenta while 30k markers are purple.

Continue to zoom in on both axes as required until you can see the start and stop position of Ca_15125 in CDC Frontier v1 and the equivalent gene in CDC Frontier gnm3. Next, we will add a label to the candidate gene in CDC Frontier gnm3 genome annotation to see the gene ID.



By comparing the relative positions of annotated genes in the two genome assemblies, we have successfully identified gene ID **Ca4g087500** in CDC Frontier gnm3 as the equivalent gene of **Ca15125** in v1.

We will now close the Ca4 chromosome for CDC Frontier v1 to simplify the display for downstream tasks.

13. Click on chromosome name on top of the axis to display the axis title menu

14. Click on the 'X' to remove all datasets for CDCFrontierGA1 from view

The screenshot shows the Pretzel web application interface. The left sidebar contains a 'Datasets' section with filters for 'Group' (All) and 'Category' (All). Below this is a search bar for 'Genes gnm3' and a 'Case' section with 'Browse' and 'Match' buttons. The 'Upload new data' section shows a list of datasets, including 'Ciclar arletinum - CDCFrontier_gnm3 - Genes'. The central map view displays a chromosome map with tracks for 'Chickpea_CDCFrontierGA_1: Ca4'. The right sidebar shows a legend with a table of datasets:

Dataset	Count
Ca_Malik_PH_QT1s-KF	4 / 4
Ca_CDCFront10_30k	717 / 1010
CDCFrontier_v1_Genes	686 / 3469

The map view shows several chromosome tracks with labels like 'Ca_15125', 'CaqP44.2', and 'Ca4g087500'. The bottom of the map has 'Zoom' and 'Reset' buttons.

Explore diversity in a QTL region contributing to plant height

Next, we will load a vcf file containing genotypic data for chickpea accessions in the AGG and explore diversity of allele calls around the QTL CaqPH4.2 and annotated genes that contribute to plant height on chromosome 4.

Task 9: Load genotype data

1. Click the 'Reset' at the bottom of the axis and zoom in until you see a similar number of markers displayed surrounding gene Ca4g087500 as shown below.

2. Click on the 'Explorer' tab

3. Type 'Genotypes Raw' in the Search field

4. Click on the black '+' icon on the left side of dataset name 'AGG Raw Release 2' to show available chromosomes and add chr4 genotypic data

Note: New feature track in pink has now been added to the axis

Ca_CDCFrontier_gnm3 : chr4

Ca_CDCFrontier_gnm3

Namespace

Parent

Created 05/12/2024, 3:59:19 pm

Last Updated 05/12/2024, 3:59:19 pm

Public true

Read Only true

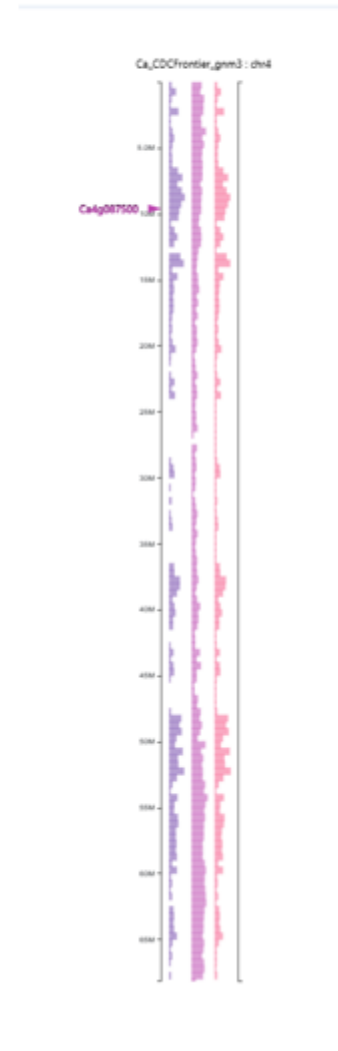
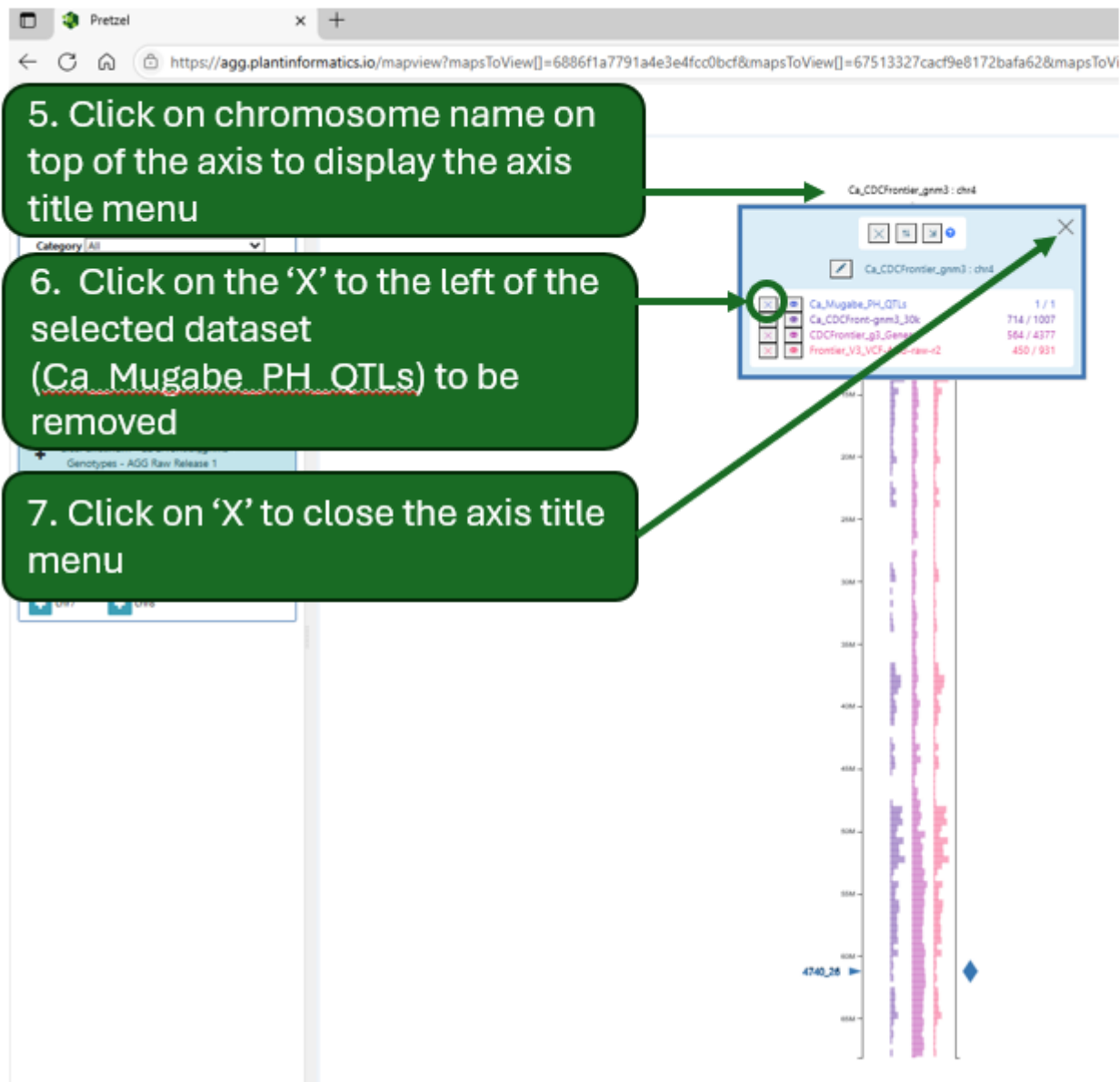
tags BlastDb

apiHost <https://agg.plantinformatics.io>

Meta View Edit

Metadata (11)

```
displayname: Cicer arietinum - CDCFrontier_gnm3 - Genome
species: Cicer arietinum
citation: Genome assembly v3 for genotype CDCFrontier from the Cook Lab, UC-Davis
data source: https://data.legumefdb.org/Cicer/arietinum/genomes/CDCFrontier\_gnm3.GTF.gz
uploaded by: Agriculture Victoria
contact: gabriel.keenle-gagereagriculture.vic.gov.au
licensing: None
type: Genome
crop: Chickpea
comment: TBC
category: core
```



Task 10: Explore diversity in the region of interest

This task will involve zooming in and selecting markers that you wish to view genotype calls for surrounding the candidate gene.

The screenshot displays the Pretzel genome browser interface with three numbered callouts illustrating the workflow:

- 1. With your mouse pointer, click and drag along the axis to highlight a portion of the chromosome surrounding the gene**: A green bracket on the chromosome axis highlights a region around the *CdcFrontier_g3* gene.
- 2. Click on the 'Genotypes' tab to display base pair positions, features and Ref/Alt nucleotides for the selected markers**: The 'Genotypes' tab is selected in the top navigation bar.
- 3. Click on the cog wheel to display the Genotypes menu**: A green circle highlights the cogwheel icon in the top right corner of the Genotypes panel.

The interface includes a left sidebar with 'Genotypes Raw' and 'Upload new data' options, a central chromosome view, and a right panel showing a table of markers and their associated data.

Ch	Position	Marker	Ref	Alt	Genotype
chr4	948044	AVR-Ca-01875-04-037780482	A	G	
chr4	9512413	AVR-Ca-01874-04-037725337	A	T	
chr4	9520774	AVR-Ca-01873-04-037688148	A	G	
chr4	9578931	AVR-Ca-01872-04-037645286	A	T	
chr4	9595421	AVR-Ca-01871-04-037613362	A	G	
chr4	9627788	AVR-Ca-01870-04-037588948	A	G	
chr4	9659900	AVR-Ca-01869-04-037542976	A	G	
chr4	9693901	AVR-Ca-01868-04-037514242	A	A	
chr4	9699024	AVR-Ca-01867-04-037504342	C	T	
chr4	9733418	AVR-Ca-01866-04-037475961	T	C	
chr4	9758783	AVR-Ca-01865-04-037469173	A	T	

Markers that overlap annotated genes are indicated by the coloured (magenta) boxes in the 'CDCFrontier_g3_Genes' column. Note that while the nucleotides in the 'Ref' and 'Alt' columns are defined, the actual genotype calls are not yet loaded.

After displaying the Genotypes menu, you should select only a subset of samples to display e.g. 100. Attempting to load thousands of samples will result in slow performance.

4. See the total number of samples in the vcf file

5. Select a subset of samples from the list (click top sample in list, scroll down, then press **'Shift'** on your keyboard while clicking on another sample lower down the list)

6. Check you have selected ~100 samples. (The number of samples selected to be displayed is indicated in brackets)

7. Click on the **'VCF Lookup'** button to load genotype calls

8. Click on the **'X'** button to close the window

Note!
The name of the vcf file loaded is shown in this tab

Note!
The selected samples will appear in the lower box. You can remove unwanted samples by highlighting them and pressing **'Backspace'** on your keyboard.

9. Adjust the height of the header row by hovering your pointer here until the cursor changes to a double-ended arrow, then drag upwards or downwards.

10. Adjust the width of the genotype matrix by hovering your pointer here until the cursor changes to a double-ended arrow, then click and drag to the left.

Congratulations! You made it to the output!

Task 11: Sort AGG accessions in the genotype table based on allele calls

In this task, you will learn how to sort accessions from left to right in the '**Genotypes**' table, based on the allele you select for a chosen marker. You can sort samples by the '**Ref**' or '**Alt**' columns.

The screenshot shows the Pretzel web interface. On the left, there's a sidebar with filters for 'Group' (All), 'Crop' (Chickpeas), and 'Category' (All). Below this is the 'Genotypes Raw' section with 'Case' and 'Match' buttons. The main area displays a map view of the 'Ca_CCDChromat_gm3 - chr4' and a large table of genotype data. The table has columns for 'Chromosome', 'Position', 'Name', 'Ref', and 'Alt'. The 'Ref' and 'Alt' columns are highlighted in green, indicating the selected alleles. A green callout box with an arrow points to the 'Ref' and 'Alt' columns, explaining the sorting process.

1. Click on a 'Ref' or 'Alt' allele to sort samples in the genotype table left-to-right based on the marker selected.

2. Watch how samples are re-ordered when you try clicking on different 'Ref' and 'Alt' alleles.

Note!
Selected alleles appear green. To remove selection of the allele, click on the allele a second time.

Task 12: Filter AGG accessions with a defined haplotype of interest

Using data from a research project, haplotypic diversity in this region containing a locus for plant height was explored in wild chickpea (*Cicer reticulatum* and *Cicer echinospermum*) as well as cultivated chickpea (*Cicer arietinum*). It was discovered that some haplotypes (i.e. series of genotype calls) were common among wild chickpea accessions while other haplotypes were common among cultivated chickpea accessions. Some of the most frequently observed haplotypes are shown in the table below. Our hypothesis is that we can identify accessions from the AGG with wild chickpea introgressions around the plant height gene by filtering accessions to display based on specific haplotypes.

				Haplotypes predominant in wild					Haplotypes predominant in cultivated			
Marker Name	Position	REF	ALT	1	2	3	4	5	6	7	8	9
AVR-Ca-01872.04-037645268	9576931	G	T	REF	REF	REF	REF	REF	REF	REF	REF	REF
AVR-Ca-01871.04-037631362	9590421	G	A	REF	REF	REF	REF	REF	ALT	REF	ALT	./.
AVR-Ca-01870.04-037568948	9627768	A	G	ALT	ALT	ALT	ALT	ALT	ALT	REF	ALT	ALT
AVR-Ca-01869.04-037542676	9659900	G	A	ALT	ALT	ALT	ALT	ALT	ALT	REF	ALT	./.
AVR-Ca-01868.04-037514243	9689081	G	A	REF	REF	REF	REF	REF	REF	REF	ALT	REF
AVR-Ca-01867.04-037504342	9699024	C	T	REF	REF	REF	ALT	REF	REF	REF	REF	REF
AVR-Ca-01865.04-037475951	9733418	T	C	ALT	REF	REF	REF	./.	REF	REF	REF	REF
AVR-Ca-01864.04-037469173	9739763	T	C	ALT	REF	REF	REF	ALT	REF	REF	REF	REF

First, we will do a Feature search of the markers IDs defining the haplotypes above to make them easier to find in the chromosome.

1. Click on the 'Search' tab

2. Copy and paste the marker names from the provided text file 'Pretzel_workshop_plant_height_haplotye_markers.txt' into the 'Feature Search' box

3. Click Search

4. Zoom axis to achieve desired marker resolution

5. The selected markers (purple arrows) will be displayed along the chromosome axis

Note: To simplify the display, the feature tracks can be collapsed from view by clicking the 'Split Axis' button in the Axis Menu

In this case, we will seek to identify chickpea samples in the AGG that have the wild haplotype '1' by selecting the specified combination of REF and ALT allele calls shown below for the 8 markers.

7. Click on the cog wheel icon on the 'Genotypes' tab to bring up the vcf menu

6. Click a series of 'Ref' and/or 'Alt' alleles for the series of markers shown, reflecting haplotype '1'.

Chr	MAF	Ref	Alt	AGG1006598C HIC1-B00011-2-12	AGG1006611C HIC1-B00011-2-20	AGG1006722C HIC1-B00011-2-83	AGG1006726C HIC1-B00011-2-85	AGG1006709C HIC1-B00011-2-73	AGG1006655C HIC1-B00011-2-41	AGG1006679C HIC1-B00011-2-52	AGG1001510C HIC1-B00010-2-84	AGG1001511C HIC1-B00010-2-85	AGG1001515C HIC1-B00010-2-86	AGG1001516C HIC1-B00010-2-87
95 chr4	0.10	G	T	.	.	.	.	0	.	.	0	0	0	0
95 chr4	0.40	G	A	.	.	.	1	0	.	.	0	0	0	0
96 chr4	0.1	A	G	2	2	2	2	2	2	2	2	2	2	2
96 chr4	0.2	G	A	2	2	2	2	2	2	2	0	0	0	0
96 chr4	0.1	G	A	0	0	0	0	0	0	0	0	0	0	0
96 chr4	0.09	C	T	1	1	1	.	2	0	0	0	0	0	0
97 chr4	0.13	T	C	.	0	0	0	0	0	0	0	0	0	0
97 chr4	0.15	T	C	1	.	.	.	1	0	0	0	0	0	0

8. If samples are still listed in the lower box from the previous VCF Lookup, click into the box, select all samples with 'Ctrl' + 'A', then clear them with the 'Backspace' button on the keyboard.

Select Samples for Dataset

List of samples from VCF (Total : 12884)

☐ Filter by defined Haplotype (8 SNPs) ☐ Match exact alleles

Filter by Name: Filter available samples by name

AGG1001208CHIC1-B00002-5-11
AGG1001214CHIC1-B00002-5-12
AGG1001216CHIC1-B00002-5-13
AGG1001219CHIC1-B00002-5-14
AGG1001222CHIC1-B00002-5-15

Selected samples to be requested in VCF Lookup (165)

AGG55155CHIC1-B00013-6-03
AGG55158CHIC1-B00013-6-05
AGG55161CHIC1-B00013-6-08
AGG55166CHIC1-B00013-6-13
AGG55167CHIC1-B00013-6-14

Clear Passport data via Genolink

VCF Lookup

IMPORTANT: select only a subset of the filtered samples. Loading thousands of samples will result in slow *Pretzel* performance.

The screenshot shows the 'Select Samples for Dataset' window in Pretzel. The window has tabs for 'Datasets', 'Lookup', 'Result', and 'Controls'. The 'Lookup' tab is active, showing a list of samples from a VCF file (Frontier_V3_VCF-AGG-raw-r2) with 486 filtered samples. The 'Filter by defined Haplotype' checkbox is checked, and the 'Match exact alleles' checkbox is also checked. A list of sample IDs is shown, with a subset of 165 samples selected for the VCF Lookup. The 'VCF Lookup' button is highlighted. A green callout box at the top right indicates to click the 'X' button to close the window.

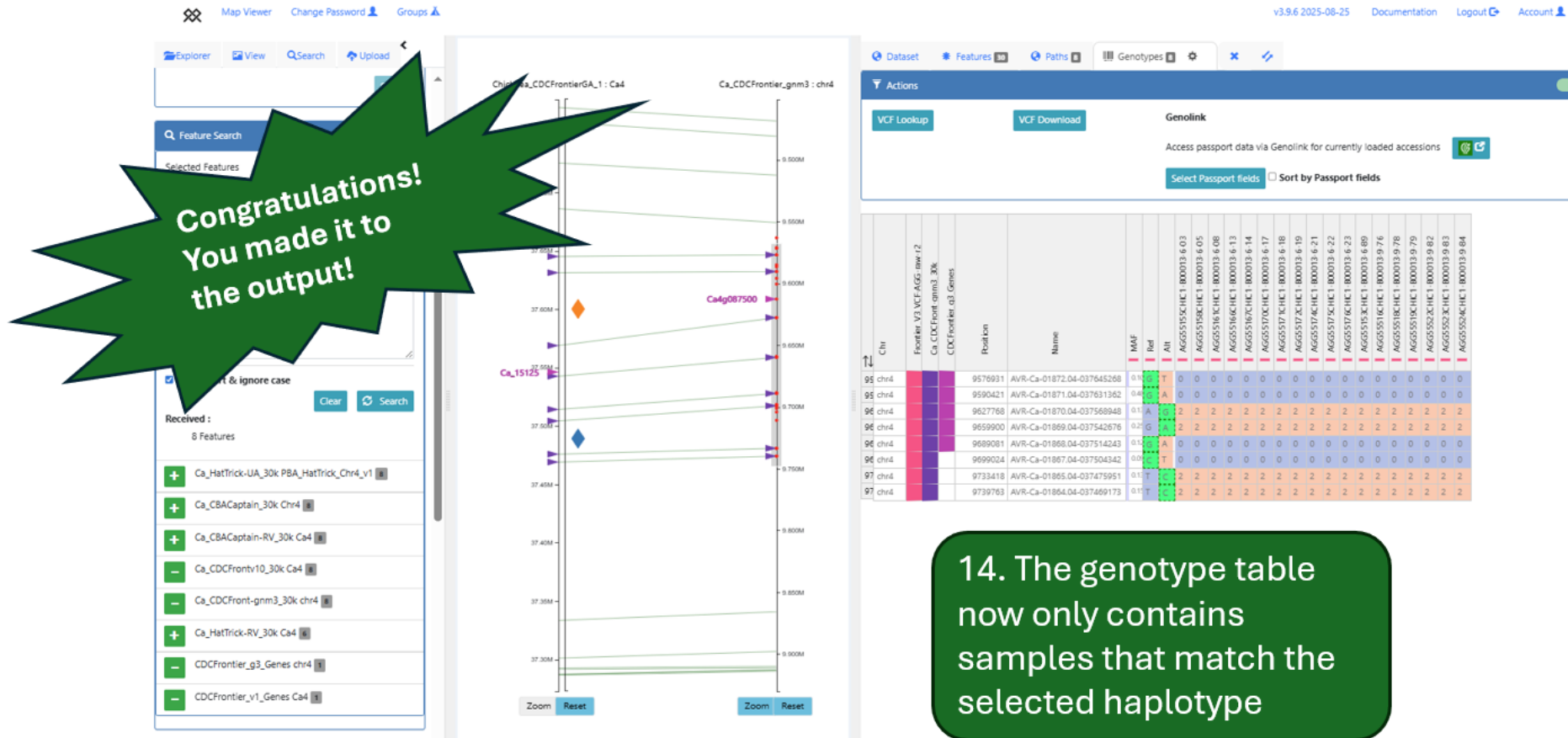
9. Tick the 'Filter by defined Haplotype' box

10. Tick the 'Match exact alleles' box to avoid heterozygous genotype calls

11. Select a subset of samples from the list

12. Click the 'VCF Lookup' button

13. Click on the 'X' button to close the window



We have now successfully identified chickpea accessions in the AGG with the wild haplotype '1'.

Explore diversity around a locus containing a flowering time gene orthologue

The next task will demonstrate how you can use BLAST to find a region of interest in the genome to investigate in *Pretzel*. Here, we will aim to find a gene in chickpea that is orthologous to Medicago gene MtFTa1 that was reported by Laurie et. al. (2011) to be involved in flowering time.

Task 13: Perform a BLAST search

Start by restarting your *Pretzel* session (e.g. re-enter the URL <https://agg.plantinformatics.io/>), since this is the quickest way to remove loaded datasets from view. Then, perform BLAST Search using the nucleotide sequence in the provided file 'Flowering_time_HQ721813.txt'. Alternatively, you can copy the sequence from NCBI with the link <https://www.ncbi.nlm.nih.gov/nuccore/HQ721813.1?report=fasta>.

1. Click on the 'Map Viewer' tab after resetting the URL

AGG Pretzel
Pretzel is an interactive web application designed to enable rapid translation of genomics outputs for any species into concrete answers to research and breeding questions.

Recent Updates

- 25/07/2025 - New Pretzel version v3.9.0 released! New features enable further integration of passport data via Genolink. Full list of new features [here](#)
- 25/05/2025 - New Pretzel version v3.8.0 released! Full list of new features in [docs](#) and [here](#)
- 30/04/2025 - New Pretzel functionality now available in release v3.7.3! Users can now search genotype data for accessions carrying user-defined haplotypes. See the new Use Case [Finding genotyped accessions carrying a specific haplotype](#) for more details
- 16/01/2025 - Added genotypes for 11,071 Chickpea AGG Plant genetic resources (<https://doi.org/10.7910/DVN/5QPKW>) genotyped as part of the AGG Strategic Partnership, anchored against the CDC Frontier gnm3 assembly
- 01/11/2024 - Added the 10+ Wheat Genomes from Walkowiak et al. 2020, including de novo gene annotations ([White et al. 2024](#)), Wheat Barley 40k v1.1 mappings and an example of how it can all be used in [User Story 3 - From SSRs to pan-genomes with Pretzel](#). Two decades of wheat yield QTLs on chromosome 7A
- 23/10/2024 - Pretzel v3.0.0 released! This includes the VCF Search use case (see the documentation on how to use it [here](#)) and a range of improvements and enhancements
- 17/09/2024 - Added User Story 2 - [Filtering AGG wheat accessions for stripe rust resistance gene Yr34/Yr45](#)
- 23/08/2024 - Added genotypes for 12,606 Hexaploid Wheat AGG Plant genetic resources (<https://doi.org/10.7910/DVN/134Q2W>) genotyped as part of the AGG Strategic Partnership

← ↻ 🏠 <https://agg.plantinformatics.io/mapview#left-panel-feature-search>

🔗 Map Viewer Change Password 👤 Groups 🧑

🔍 Explorer 📄 View 🔍 Search

Feature Search

Selected Features 0 ←

Features Input 0

Feature (Marker/Gene) Names
(whitespace or comma separated)

☒ Match part & ignore case

Clear 🔍 Search

DNA Sequence Blast Search

☒ Sequence Input

Reference to search :
Cicer arietinum - CDCFrontier_gnm3 - Genome

DNA Sequence Input : FASTA

>HQ721813.1 Medicago truncatula cultivar Jester Flt1 mRNA, complete cds
 ATGGCTGGTAGCAGTAGGAATCCACTAGCTGTAGGGCGTGAATAGGGGA
 TGTAAAGACTCATTGAAA
 ATTCCATTCTCTTCGAGTGACCTATGGTAATAGGGATGTGAATAATGGTTGT
 GAGCTCAAACCTTCTCA
 AATTGGAAATCAACCGAGAGTGAGTGTGGTGGAAACGATCTCAGAAACC
 TCTACACCTAGTTATGGTG
 GATCCTGATTCACCTAGCCCAAGTAACCCCACTTTTAAGGAGTACCTTCACT
 GGTGGTGACTGATATTC
 CAGGAACCACTGAAGTCACTTCGGCAATGAGGTGTAAATTATGAAAGG
 CCACGACCACTTCAGGGAT
 CCATCGTTTCGTGTTTGTCTTATCCGCTCAACAGTGTAGACAAAGGGTTTAT
 GCTCCAGGATGGCGACAA
 AATTTCAACACAAGAGAATTGCTGAACCTACAATCTTGATCACCTGTT
 GCTGCTGCTTCTTCAATT

Clear 🔍 Search

Optional Inputs

2. Click on the 'Search' tab

3. Scroll down to 'DNA Sequence Blast Search' section

4. Select the reference as 'Cicer arietinum – CDC Frontier_gnm3 – Genome'

5. Paste the FASTA sequence in the DNA Sequence input space

6. Click 'Search'

Reference to search :

Cicer arietinum - CDCFrontier_gnm3 - Genome

DNA Sequence Input : FASTA

complete cds
 ATGGCTGGTAGCAGTAGGAATCCACTAGCTGTAGGGCGTGAATAGGGGA
 TGTAAAGACTCATTGAAA
 ATTCCATTCTCTTCGAGTGACCTATGGTAATAGGGATGTGAATAATGGTTGT
 GAGCTCAAACCTTCTCA
 AATTGGAAATCAACCGAGAGTGAGTGTGGTGGAAACGATCTCAGAAACC

Note!
 BLAST parameters can be modified by scrolling down further

GCTGCTGCTTCTTCAATT
 GTCAGAGGGAGAGTGGCTCTGGAGGAAGAACCTTTAGATAA

Clear

Search

Optional Inputs

Rows : 500

Table Result Limit

0

Length of Hit

0

% Identity

0

% Coverage

🔍 External Lookup

When you click on the BLAST output sheet, each chromosome with a BLAST hit is automatically loaded, with the position of the BLAST hits marked with an arrow and label. Note that the label is taken from the FASTA sequence identifier.

The screenshot displays the Agg.plantinformatics.io web application. The browser address bar shows the URL: [https://agg.plantinformatics.io/mapview?mapsToView\[\]=67513327cacf9e3df3bafa61&mapsToView\[\]=blastResults_Ca_CDCFrontier_gnm3-chr3](https://agg.plantinformatics.io/mapview?mapsToView[]=67513327cacf9e3df3bafa61&mapsToView[]=blastResults_Ca_CDCFrontier_gnm3-chr3). The interface includes a top navigation bar with links for Map Viewer, Change Password, and Groups. A sidebar on the left contains a search bar and a table of BLAST results. A green callout box labeled "7. Select the 'Blast Output' tab" points to the "Blast Output (16:00:44)" tab in the sidebar. Another green callout box labeled "Note! Position of the BLAST hit on chromosome 3" points to a specific row in the table. The main area shows a chromosome map for "Ca_CDCFrontier_gnm3 : chr3" with a vertical axis ranging from 1000 to 8000. A blue arrow points to a specific location on the chromosome, labeled "HQ721813.1". The right sidebar contains a "Dataset" section with a "Download CSV file" button and checkboxes for "Column Stretch Horizontal" and "Auto Column W".

7. Select the 'Blast Output' tab

Note! Position of the BLAST hit on chromosome 3

view	query ID	subject ID	% Identity	length
☒	HQ721813.1	chr3	99.790	64

← ↻ 🏠 <https://agg.plantinformatics.io/mapvie>

🔗 Map Viewer Change Password 👤 Groups 📁

🔍 Explorer 📄 View 🔍 Search 📄 Upload

☑ Match part

DNA Sequence Blast Search ⓘ

☑ Sequence Input → Blast Output (16:00:44) ✕

view	query ID	subject ID	% identity	length of HSP (hit)
☑	HQ721813.1	chr3	93.750	64

8. Click on this icon to expand the BLAST results table to a separate window

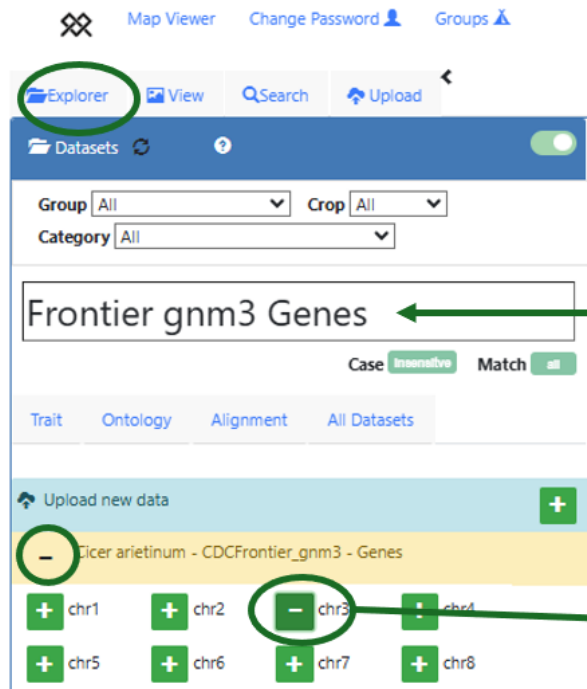
Congratulations!
You made it to the output!

view	query ID	subject ID	% identity	length of HSP (hit)	# mismatches	# gaps	query start	query end	subject start	subject end	e-value	score	query length	subject length
☑	HQ721813.1	chr3	93.750	64	4	0	202	265	12312882	12312819	1.83e-18	97.1	531	85388044

9. Click on the cross icon to close the BLAST results table.

Task 14: Explore gene annotations

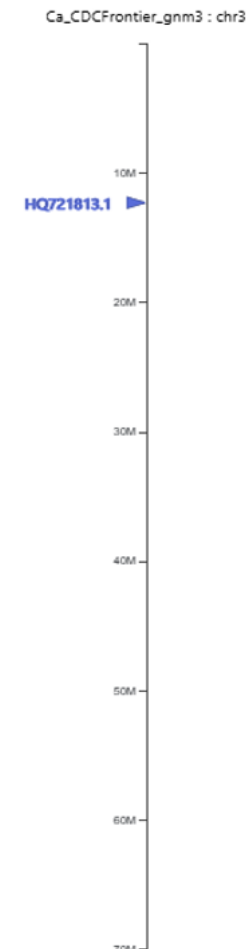
In this example, we will find a chickpea gene that corresponds to the Medicago MtFTa1 gene, then review chickpea gene annotations in the Feature Table to find a candidate gene contributing to flowering time in chickpea.



1. Go to the 'Explorer' tab

2. Type 'Frontier gnm3 Genes' into the search bar

3. Select the '+' icon to load the genes from CDC Frontier on chr3



4. Click on chromosome name on top of the axis to display the axis title menu

5. Click on 'Split Axis' button to show position of genes

6. Click on 'X' to close the axis title menu

Ca_CDCFrontier_gnm3 : chr3

blastResults_Ca_CDCFrontier_gnm3 1 / undefined
CDCFrontier_g3_Genes 0 / 3925

7. With your mouse pointer, click and drag along the axis to highlight a portion of the chromosome surrounding the BLAST hit.

8. Click the 'Zoom' button

Ca_CDCFrontier_gnm3 : chr3

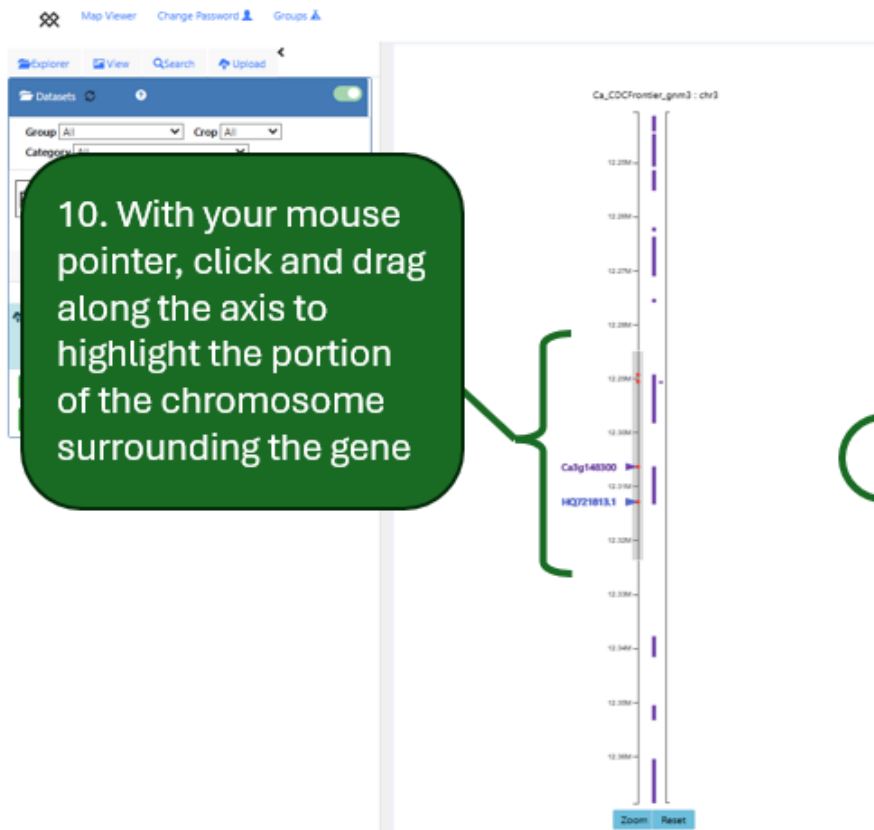
HQ721813.1

Zoom

Repeat the zoom step above until you can see the position of the BLAST hit with the closest annotated gene, as shown below. Then, we can identify the gene ID with a label.

The screenshot shows a genomic map viewer interface. On the left, there is a sidebar with various controls: a 'Map Viewer' tab, 'Change Password' and 'Groups' links, and a 'Datasets' section with filters for 'Group' (All), 'Crop' (All), and 'Category' (All). Below these are tabs for 'Trait', 'Ontology', 'Alignment', 'Genome', and 'All Datasets'. A list of datasets is shown, including 'Cicer arietinum - CDC_Frontier_v1.0 - Genes' and 'Cicer arietinum - CDCFrontier_gnm3 - Genes'. The main map area displays a vertical chromosome with gene features represented by purple bars. A green oval highlights a specific gene feature labeled 'Ca3g148300' and 'HQ721813.1'. A green callout box with white text provides instructions: '9. Click on the gene feature so the arrow shows up, then click on the arrow to add a label for the candidate gene.'

You can now see that the BLAST hit for the Medicago MtFTa1 gene sequence (**HQ721813**) matches the position of chickpea gene **Ca3g148300** from the CDC Frontier gnm3 assembly.



11. Click on the 'Features' tab. The number in the tab shows the number of features selected

12. Click and drag the handle to the left to make the Features Table wider

Block	Feature	Position	End	Strand	
blastResults_Ca_CDCFrontier_gnm3:chr3	HQ721813.1	12,312,882	12,312,819		
2	CDCFrontier_g3_Geneschr3	Ca3g148300	12,306,387	12,313,281	-
3	CDCFrontier_g3_Geneschr3	Ca3g148200	12,290,598	12,290,848	+
4	CDCFrontier_g3_Geneschr3	Ca3g148100	12,289,321	12,298,224	-

13. Toggle the tick box for 'Column Stretch Horizontal' and adjust column widths with the mouse to get the text to wrap for easier reading.

Dataset Features 4 Paths 0 Genotypes

Download CSV file ☐ Column Stretch Horizontal

	Block	Feature	Position	End	Strand	Additional Information
2	blastResults_Ca_CDCFrontier_gnm3:chr3	HQ721813.1	12312882	12312819		
	CDCFrontier_g3_Genes:chr3	Ca3g148300	12306367	12313261	-	Dbxref=Gene3D G3DSA 3.90.280.10,InterPro IPR001858,InterPro IPR008914,PANTHER PTHR11362,Pfam PF01161,Prosite PS01220,Superfamily SSF49777;Note=protein FLOWERING LOCUS T-like [Glycine max]%3B IPR008914 (Phosphatidylethanolamine-binding protein PEBP);geneFamily=legfed_v1_0.L_0FWGQD
3	CDCFrontier_g3_Genes:chr3	Ca3g148200	12290598	1229084	+	Note=putative ribonuclease H protein At1g65750-like [Glycine max];geneFamily=legfed_v1_0.L_005JJR
4	CDCFrontier_g3_Genes:chr3					symbol=MtrunA17_Chr7g0252791;Dbxref=Gene3D G3DSA 3.90.280.10,InterPro IPR001858,InterPro IPR008914,PANTHER PTHR11362,Pfam PF01161,Prosite PS01220,Superfamily SSF49777;Note=protein FLOWERING LOCUS T-like [Glycine max]%3B IPR008914 (Phosphatidylethanolamine-binding protein PEBP);geneFamily=legfed_v1_0.L_0FWGQD

14. View gene annotations

Congratulations! You made it to the output!

We can see the Prosite information describes gene Ca3g148100 as 'FLOWERING LOCUS T-like', suggesting we have correctly identified the chickpea gene in the CDC Frontier gnm3 gene annotation file that is orthologous to Medicago MtFTa1.

We can now close the BLAST result from view.

The screenshot shows the 'Search' tab of a genomic explorer. It contains three search panels:

- VCF Genotype Search:** Includes a 'VCF to search' dropdown (set to 'None'), a 'Samples input' section with a 'Show Samples' button, and a 'Features input' section with a text box containing 'Feature Names' and a 'Search' button.
- Feature Search:** Includes a 'Selected Features' section with a count of 4, a 'Features Input' section with a text box containing 'Ca3g148300', a 'Match part & ignore case' checkbox (checked), and 'Clear' and 'Search' buttons.
- DNA Sequence Blast Search:** Includes a 'Sequence Input' section with a 'Blast Output (11x46:51)' tab, and a results table. The table has columns: 'view', 'query ID', 'subject ID', '% Identity', and 'length o'. The first row is partially visible with values: '1', '10112-1', 'chr3', '99.130', and '99'. A green circle highlights the 'view' checkbox, which is currently checked.

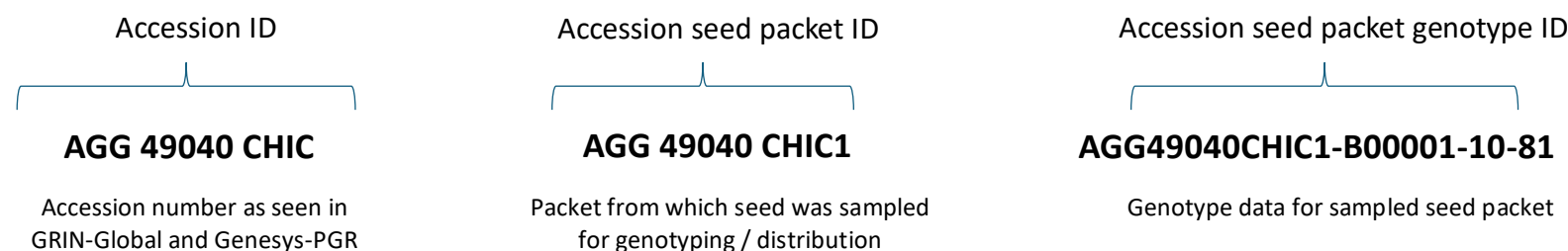
A green callout box with white text points to the 'view' checkbox, stating: "15. In the 'Search' tab, scroll down to 'DNA Sequence Blast Search' and remove the tick to remove the BLAST result from the axis".

Visualising genotype calls for specific accessions

Using data from a research project, haplotypic diversity surrounding a gene for flowering time was explored in wild chickpea (*Cicer reticulatum* and *Cicer echinospermum*) as well as cultivated chickpea (*Cicer arietinum*). It was discovered that some haplotypes (i.e. series of genotype calls) were common among wild chickpea accessions while other haplotypes were common among cultivated chickpea accessions. Some of the most frequently observed haplotypes are shown in a table below in Task 15. Our hypothesis is that we can differentiate accessions in the AGG that carry wild introgressions versus cultivated alleles around the flowering time gene using haplotype information.

In this exercise, we will search for some known varieties of chickpea and some hybrids (wild introgression lines) to assess whether they contain haplotypes typically seen in wild or cultivated chickpea around a flowering time gene. AGG sample identifiers are provided in the electronic file 'Pretzel_workshop_Task15_AGG_chickpea_sample_list.txt'.

The Genotype ID provided uniquely identifies the DNA sample that was genotyped in the AGG Strategic Partnership for that accession.



Accession IDs can be found from several sources:

- AGG GRIN-Global database (<https://ausgenebank.agriculture.vic.gov.au/gringlobal/search>)
- Genolink (<https://genolink.plantinformatics.io/>)
- Genesys-PGR (<https://www.genesys-pgr.org/>)
- Publications

Task 15: Visualise genotype data in QTL region for specific AGG accessions

The screenshot displays a genomic data visualization tool. On the left, there are three search panels: 'VCF Genotype Search', 'Feature Search', and 'DNA Sequence Blast Search'. The 'VCF Genotype Search' panel has a 'VCF to search:' dropdown set to 'None', a 'Samples input:' field with a 'Show Samples' button, and a 'Features input:' field with a 'Search' button. The 'Feature Search' panel has a 'Selected Features' list with 'Ca3g148300' and a 'Features Input' field with a 'Search' button. The 'DNA Sequence Blast Search' panel has a 'Sequence Input' field and a 'Blast Output (11:46:51)' field. Below these panels is a table with columns: view, query ID, subject ID, % identity, and length o. The table contains one row:

view	query ID	subject ID	% identity	length o
☐	HQ721813.1	chr3	93.750	64

In the center, a chromosome map shows a vertical axis with positions from 12,240,000 to 12,420,000. A purple bar represents the chromosome, and a red dot marks the position of 'Ca3g148300'.

On the right, a table displays genomic data. The table has columns: Block, Feature, Position, End, Strand, and Additional Information. The table contains four rows:

Block	Feature	Position	End	Strand	Additional Information
blastResults_Ca_CDCFrontier_gnm3:chr3	HQ721813.1	12312882	12312819		
2	CDCFrontier_g3_Genes:chr3	Ca3g148100	12289321	12298224	-
3	CDCFrontier_g3_Genes:chr3	Ca3g148200	12290598	12290846	+
4	CDCFrontier_g3_Genes:chr3	Ca3g148300	12306367	12313261	-

A green arrow points to the 'Reset' icon at the bottom of the chromosome map, with the text: '1. Click on the 'Reset' icon to reset the chromosome axis'.

2. Go to the 'Explorer' tab

3. Type 'Frontier gnm3 Raw' into the filter

4. Select the '+' icon to load raw genotype data on chr3

Dataset: Frontier_V3_VCF-AGG-raw-r2

Namespace: Ca_CDCFrontier_gnm3

Parent: Ca_CDCFrontier_gnm3

Created: 31/07/2025, 4:14:45 pm

Last Updated: 31/07/2025, 4:14:45 pm

Public: true

Read Only: true

tags: VCF view Genotype

apiHost: https://agg.plantinformatics.io

Meta: View Edit

Metadata (11)

- Crop: Chickpea
- displayN: Cicer arietinum - CDCFrontier_gnm3 - Genotypes
- ame: - AGG Raw Release 2
- species: Cicer arietinum
- DOI: https://doi.org/10.7910/DVN/ECQ4WC
- Desc: Raw anchored genotype data VCF files for chickpea
- ript: Plant Genetic Resources from the Australian Grains
- ion: Genebank genotyped with the Illumina Infinium
- Pulses 30K v1.0 BeadChip. Anchored genotypes are
- defined against Cicer arietinum genotype
- CDCFrontier genome assembly v3
- CDCFrontier.gnm3.QT0P. Release 250505.
- Data source: Australian Grains Genebank Strategic Partnership
- uploaded by: Agriculture Victoria
- Contact: Gabriel Keeble-Gagnere
- Licensing: CC BY 4.0
- name: 250505_AGG-Chickpea_Raw_Anchored-CDCFrontier_gnm3
- type: VCF

Get VCF Status Prepare Dataset Histograms

Show detail

5. With your mouse pointer, click and drag along the axis to highlight the portion of the chromosome surrounding the gene

6. Click on the cog wheel icon in the 'Genotypes' tab

Ca₂CDCFrontier_gnm3 : chr3

Zoom Reset

Dataset Features Paths Genotypes

Actions

VCF Lookup VCF Download

Chr	Frontier VCF	Barcode	CDCFrontier	Position	Name	MAF	Ref	Alt
118	chr3			11847470	AvR-Ca-00971.03-026868214	0.4	T	G
118	chr3			11878565	AvR-Ca-00970.03-026833644	0.4	T	C
128	chr3			12861815	AvR-Ca-00969.03-025880939	0.3	T	C
129	chr3			12922894	AvR-Ca-00968.03-025790529	0.0	C	T
130	chr3			13000988	AvR-Ca-00967.03-025712157	0.0	G	A
130	chr3			13022575	AvR-Ca-00966.03-025690569	0.0	T	G
138	chr3			13847944	AvR-Ca-00965.03-024852336	0.0	C	T
138	chr3			13875982	AvR-Ca-00964.03-024824337	0.0	A	G
138	chr3			13891881	AvR-Ca-00963.03-024807932	0.0	C	T
140	chr3			14009901	AvR-Ca-00962.03-024688407	0.4	G	A
141	chr3			14183822	AvR-Ca-00961.03-024520962	0.3	C	T

7. Type or paste an Accession ID into the 'Filter by Name' box. If the accession is contained in the VCF, the Genotype ID will be displayed in the box below.

Frontier_V3_VCF-AGG-raw-r2

List of samples from VCF (Total : 1)

☐ Filter by defined Haplotype ?

Filter by Name AGG 51643 CHIC (1)

AGG51643CHIC2-B00004-8-14

Selected samples to be requested in VCF Lookup

Sample Names

Clear Passport data via Genolink

VCF Lookup

8. Click on the Genotype ID in the upper box to add the sample to the lower 'Selected samples to be requested in VCF Lookup' box

Frontier_V3_VCF-AGG-raw-r2

Select Samples for

List of samples from VCF (Total : 1)

☐ Filter by defined Haplotype ?

Filter by Name AGG51643CHIC2-B00004-8-14 (1)

AGG51643CHIC2-B00004-8-14

Selected samples to be requested in VCF Lookup (1)

AGG51643CHIC2-B00004-8-14

Clear Passport data via Genolink

VCF Lookup

9. Select 'VCF Lookup'

Ca_CDCFrontier_gnm3 : chr3



Dataset Features 188 Paths 0 Genotypes 23

Actions

VCF Lookup VCF Download Genolink

Access passport data via Genolink for currently loaded accessions

Select Passport fields ☐ Sort by Passport fields

Chr	Frontier V3 VCF AGG 181 12 blastResults_Ca_CDCFrontier_	CDCFrontier g3 Genes	Position	Name	MAF	Ref	All	AGG51643CHC2-1000048-14
105 chr3			10594959	AVR-Ca-00975.03-027833123	0.00	C	T	0
106 chr3			10637683	AVR-Ca-00976.03-027876086	0.00	A	G	0
106 chr3			10690629	AVR-Ca-00977.03-027929809	0.00	C	T	0
107 chr3			10794564	AVR-Ca-00978.03-028035126	0.3	A	G	2
107 chr3			10795303	AVR-Ca-00979.03-028035865	0.00	C	T	0
113 chr3			11324333	AVR-Ca-00974.03-027997784	0.40	A	C	2
118 chr3			11847388	AVR-Ca-00972.03-026868404	0.40	T	C	2
118 chr3			11847478	AVR-Ca-00971.03-026868314	0.40	T	G	2
118 chr3			11878585	AVR-Ca-00970.03-026833644	0.40	T	C	2
128 chr3			12861815	AVR-Ca-00969.03-025880039	0.30	T	C	2
129 chr3			12922894	AVR-Ca-00968.03-025790529	0.00	C	T	0
130 chr3			13000988	AVR-Ca-00967.03-025712157	0.00	G	A	0
130 chr3			13022575	AVR-Ca-00966.03-025690569	0.00	T	G	0
138 chr3			13847944	AVR-Ca-00965.03-024852336	0.00	C	T	0
138 chr3			13875982	AVR-Ca-00964.03-024824337	0.00	A	G	2
138 chr3			13891881	AVR-Ca-00963.03-024807932	0.00	C	T	2
140 chr3			14009901	AVR-Ca-00962.03-024688407	0.30	G	A	2
141 chr3			14183822	AVR-Ca-00961.03-024520962	0.30	C	T	2
145 chr3			14537790	AVR-Ca-00960.03-024159068	0.00	G	A	0
146 chr3			14681706	AVR-Ca-00959.03-024013204	0.40	C	T	2
147 chr3			14729191	AVR-Ca-00958.03-023965230	0.30	T	C	0
151 chr3			15108801	AVR-Ca-00957.03-023581415	0.00	G	A	0
152 chr3			15221087	AVR-Ca-00956.03-023468600	0.10	A	G	2

10. Genotype data
can now be
visualised for the
desired accession

Since we already have the Genotype IDs for chickpea samples of interest, we can directly copy and paste them into the lower box. Copy and paste Genotype IDs from the electronic file '**Pretzel_workshop_Task15_AGG_chickpea_sample_list.txt**' as shown below.

Frontier_V3_VCF-AGG-raw-r2

Select Samples for Dataset

List of samples from VCF (Total : 12884)

☐ Filter by defined Haplotype

Filter by Name Filter available samples by name (1)

AGG51643CHIC2-B00004-8-14

Selected samples to be requested in VCF Lookup (5)

- AGG49039CHIC1-B00011-6-57
- AGG49042CHIC1-B00001-10-83
- AGG48845CHIC1-B00012-10-44
- AGG51137CHIC1-B00012-12-77
- AGG48844CHIC1-B00012-10-43

Clear Passport data via Genolink

VCF Lookup

11. Next, paste Genotype IDs directly from the provided file into the lower 'Selected samples' box

12. Select 'VCF Lookup'

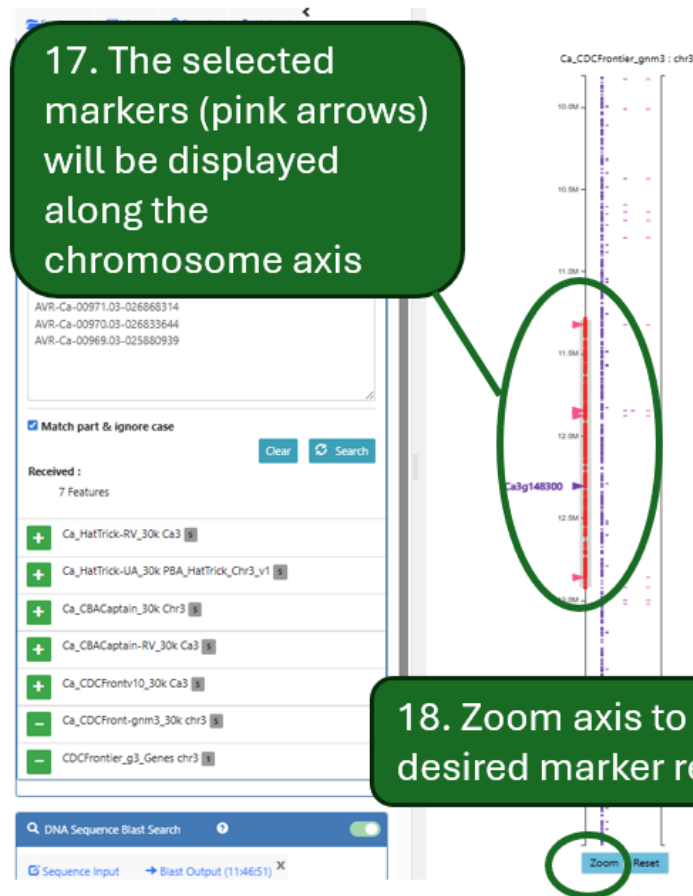
The screenshot displays the Genomics Workbench interface. On the left, the 'Feature Search' tab is active, showing a list of features. A green circle highlights the 'Ca_CDCFront-gnm3_30k chr3' feature. A green arrow points to the 'Search' button. In the center, a green box contains the text: '13. Click on the 'Search' tab', '14. Copy and paste the marker names from the provided text file 'Pretzel_workshop_Task15_flowering_time_markers.txt' into the 'Feature Search' box', and '15. Click Search'. On the right, a table of marker data is shown, with columns for Chr, Position, Name, and various marker types. A green circle highlights the 'Ca_CDCFront-gnm3_30k chr3' feature in the table.

13. Click on the 'Search' tab

14. Copy and paste the marker names from the provided text file 'Pretzel_workshop_Task15_flowering_time_markers.txt' into the 'Feature Search' box

15. Click Search

16. Load 30k markers 'CDCFront-gnm3_30k chr3' dataset



18. Zoom axis to achieve desired marker resolution

Dataset Features Paths Genotypes

Actions

VCF Lookup VCF Download

Genolink

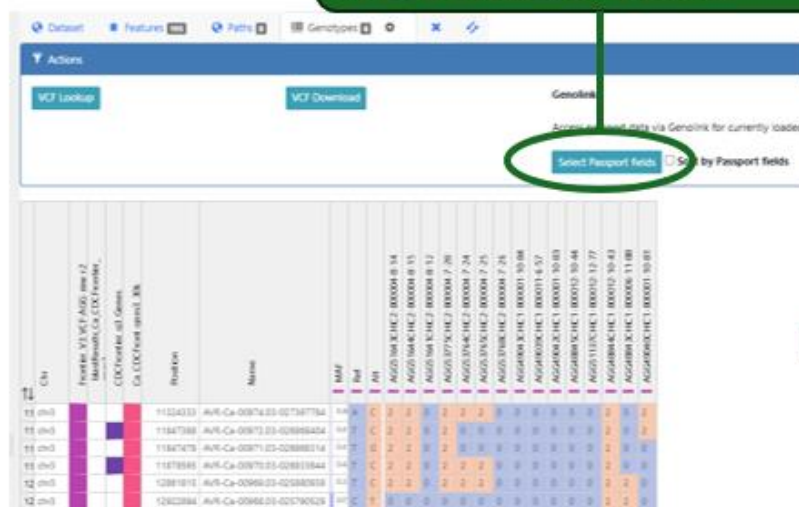
Access passport data via Genolink for currently loaded accessions

Select Passport fields Sort by Passport fields

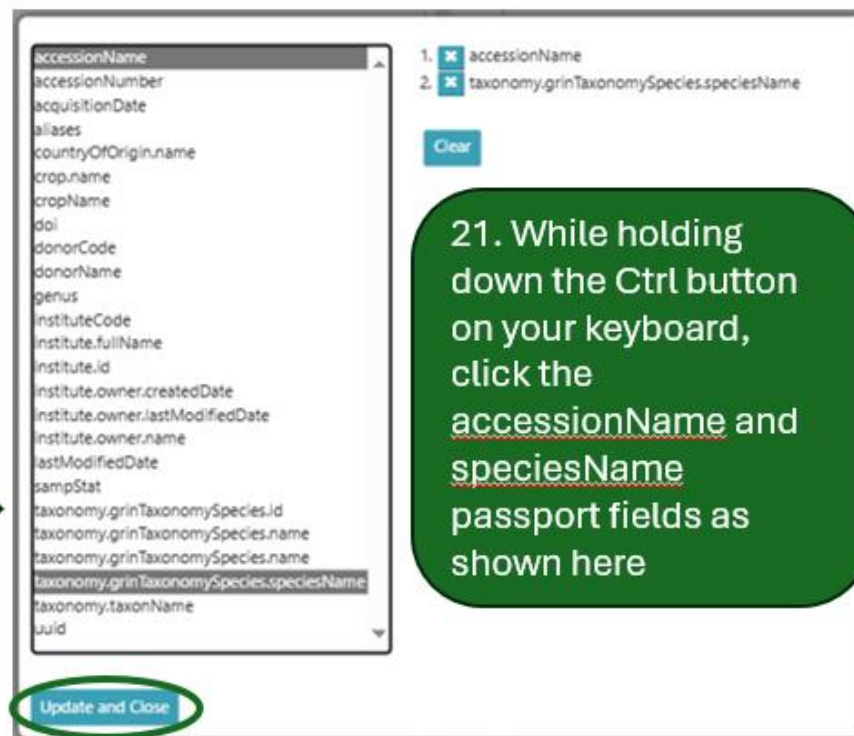
Chr	Position	Name	MAF	Ref	Alt
11 chr3	11324333	AVR-Ca-00974.03-027397784	0.44	A	C
11 chr3	11847388	AVR-Ca-00972.03-026866404	0.44	T	C
11 chr3	11847478	AVR-Ca-00971.03-02666314	0.44	T	C
11 chr3	11878565	AVR-Ca-00970.03-026833644	0.44	T	C
12 chr3	12861815	AVR-Ca-00969.03-025880939	0.33	T	C
12 chr3	12922894	AVR-Ca-00968.03-025790529	0.60	C	T

19. Genotype data can now be visualised for the selected accessions

20. Click on 'Select Passport fields' button



21. While holding down the Ctrl button on your keyboard, click the accessionName and speciesName passport fields as shown here



22. Click on 'Update and Close'

23. Adjust the passport data row height by hovering your pointer here until the cursor changes to a double-ended arrow, then drag upwards or downwards.

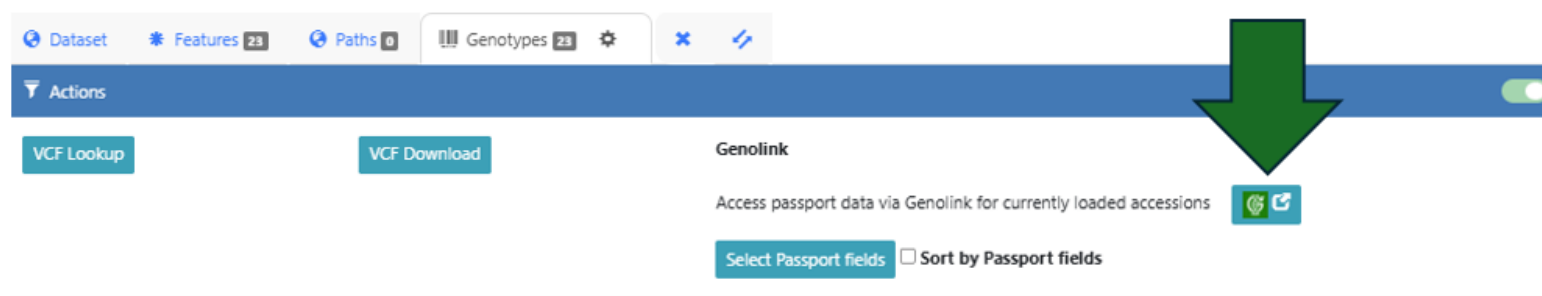
	Chr	Genome	Accession	Name
11	chr3			Frontier_V3.VCF.AGS.raw.r2
11	chr3			blastResults_Ca_CDC.Frontier.gmm
11	chr3			CDC.Frontier.g3_Genes
11	chr3			Ca_CDC.Front.gmm3_30k
11	chr3			Position
11	chr3			Name
11	chr3			MAF
11	chr3			Ref
11	chr3			Alt
11	chr3			Cv. Hatrick xC. reticulatum (Cudi E
11	chr3			Cv. Hatrick xC. reticulatum (Cudi E
11	chr3			Cv. Hatrick xC. reticulatum (Cudi E
11	chr3			UCD_09946_03_019013
11	chr3			UCD_09945_03_018988
11	chr3			UCD_09945_03_018989
11	chr3			UCD_09947_03_018992
11	chr3			Yorker
11	chr3			Almaz
11	chr3			PBA Boundary
11	chr3			PBA Hat Trick
11	chr3			PBA Drummond
11	chr3			PBA Slasher
11	chr3			PBA Pistol
11	chr3			Flipper

				Haplotypes predominant in wild				Haplotypes predominant in cultivated			
Marker Name	Position	REF	ALT	1	2	3	4	5	6	7	8
AVR-Ca-00974_03-027397784	11,324,333	A	C	ALT	ALT	ALT	ALT	REF	ALT	REF	ALT
AVR-Ca-00972_03-026868404	11,847,388	T	C	ALT	REF	ALT	ALT	REF	REF	ALT	ALT
AVR-Ca-00971_03-026868314	11,847,478	T	G	ALT	REF	ALT	REF	REF	REF	ALT	ALT
AVR-Ca-00970_03-026833644	11,878,565	T	C	ALT	ALT	ALT	ALT	REF	REF	ALT	ALT
AVR-Ca-00969_03-025880939	12,861,815	T	C	ALT	ALT	./.	ALT	REF	REF	REF	REF

Based on these results, we can find chickpea hybrid lines that do and do not contain wild-like haplotypes at the flowering time locus. Interestingly, PBA Slasher contains a wild-type haplotype, suggesting it may have wild chickpea in its pedigree.

Additional tasks for a challenge

- Search for chickpea accessions in the AGG that contain wild haplotype 2 around the plant height gene in Task 12, then click on the link to Genolink to view passport data for these accessions derived from Genesys PGR.



- Use Genolink (<https://genolink.plantinformatics.io/>) to find the AGG barcode for the chickpea accession '**Almaz**' that has been genotyped and load AGG genotype data in *Pretzel*. Load genotypes for both plant height (chr X) and flowering time (chr Y) in single genotype table by selecting markers on both chromosomes.

References

Cook D. (2022) Genome assembly for CDC Frontier - CDCFrontier.gnm3.QTOP

(<https://data.legumeinfo.org/Cicer/arietinum/genomes/CDCFrontier.gnm3.QTOP/>)

Laurie RE, Diwadkar P, Jaudal M, Zhang L, Hecht V, Wen J, Tadege M, Mysore KS, Putterill J, Weller JL, Macknight RC. (2011) The Medicago FLOWERING LOCUS T homolog, MtFTa1, is a key regulator of flowering time. *Plant Physiol.* 156:2207-2224. doi: 10.1104/pp.111.180182

Malik N, Basu U, Srivastava R, Daware A, Ranjan R, Sharma A, Thakro V, Mohanty JK, Jha UC, Tripathi S, Tyagi AK, Parida SK. (2023) Natural alleles of Mediator subunit genes modulate plant height in chickpea. *Plant J.* 116:1271-1292. doi: 10.1111/tpj.16423

Mugabe D, Coyne CJ, Piaskowski J, Zheng P, Ma Y, Landry E, McGee R, Main D, Vandemark G, Zhang H, Abbo S. (2019) Quantitative Trait Loci for Cold Tolerance in Chickpea. *Crop Science.* 59: 573-582. doi: 10.2135/cropsci2018.08.0504

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, Nelson W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CL, Singh NP, Lichtenzweig J, Gali KK, Rubio J, Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR. (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat Biotechnol.* 31:240-246. doi: 10.1038/nbt.2491

Useful URLs

- *Pretzel*
<https://agg.plantinformatics.io/>
- *Pretzel* documentation
<https://docs.plantinformatics.io/>
- Australian Grains Genebank (AGG) website
<https://agriculture.vic.gov.au/crops-and-horticulture/australian-grains-genebank>
- Australian Grains Genebank (AGG) online portal - GRIN-Global Database
<https://ausgenebank.agriculture.vic.gov.au/gringlobal/search>
- Genesys-PGR
<https://www.genesys-pgr.org/>
- Genolink
<https://genolink.plantinformatics.io/>
- AGG Dataverse for accessing genotype data
<https://dataverse.harvard.edu/dataverse/australiangrainsgenebank>
- NCBI Genbank Nucleotide site
<https://www.ncbi.nlm.nih.gov/nucleotide/>

Appendix

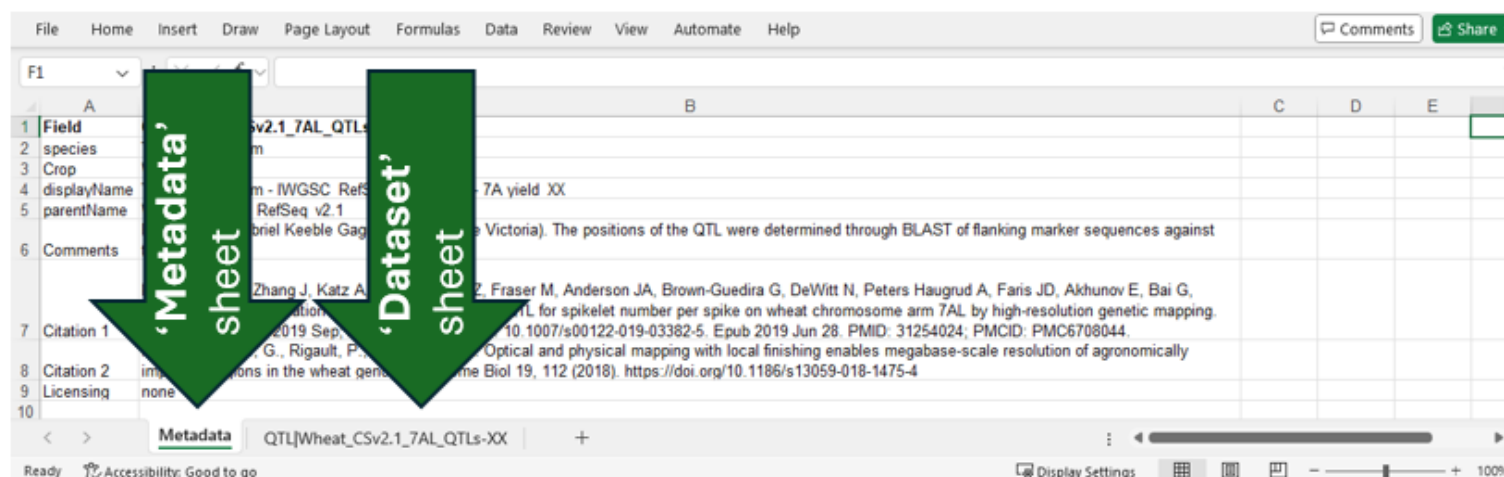
Pretzel data upload templates

Overview

Data can be uploaded into *Pretzel* using several methods, but the Excel upload method is recommended for most users. Three types of data (alignments, genetic maps and QTLs) can be independently uploaded to *Pretzel* as custom datasets, as described below. Note that it is also possible to upload genomes, BLAST databases and vcf files with assistance from the *Pretzel* development team.

Data type	Dataset Name Format	Description
Alignment	Alignment DataName	Refers to a feature with a defined position or interval, except for a QTL. This is commonly used for markers or genes, which are typically anchored to a specified genome assembly. Units are in base pairs (bp).
Map	Map DataName	Used for uploading genetic maps, where units for the defined position are in centiMorgans (cM).
QTL	QTL DataName	Used for uploading QTLs. The defined interval positions can relate to either a genome assembly (where coordinates are in bp), or a genetic map (where coordinates are in cM),

The Excel template file consists of two sheets, the '**Metadata**' sheet for defining parameters and providing meta data, and the '**Dataset**' sheet that contains data you wish to upload into *Pretzel*.



Metadata sheet

The figure and table below describe the fields in the **'Metadata'** sheet for uploading a custom QTL dataset into *Pretzel*. The mandatory fields are used by *Pretzel* and are case sensitive; they must match the relevant dataset in *Pretzel* exactly to link datasets effectively e.g. to link a QTL to a genome assembly. The optional parameters below are used to display meta data in the **'Dataset'** panel to align with FAIR (Findable, Accessible Interoperable, Re-usable) data standards. We recommend using the parameters below as best practise. Additional optional columns can be added if desired.

The figure below shows where the mandatory parameters in the **'Metadata'** sheet get incorporated into *Pretzel*.

The screenshot displays the Pretzel web interface at [https://agg.plantinformatics.io/mapview?mapsToView\[\]=blastResults_Wheat_IWGSC_RefSeq_v2.1-Chr7A&mapsToView\[\]=bl...](https://agg.plantinformatics.io/mapview?mapsToView[]=blastResults_Wheat_IWGSC_RefSeq_v2.1-Chr7A&mapsToView[]=bl...). The interface includes a 'Map Viewer' section on the left with filters for 'Group' and 'Crop', a search bar containing '7A QTL', and a list of datasets. A green arrow points from the 'shortName' of the 'parentName' in the metadata sheet to the 'Wheat_IWGSC_RefSeq_v2.1 : Chr7A' label in the map viewer. Another green arrow points from the 'Field' for dataset name in the metadata sheet to the 'Wheat_CSv2.1_7AL_QTLs-KF' label in the dataset panel. A third green arrow points from the 'parentName' in the metadata sheet to the 'Parent' field in the dataset panel. A fourth green arrow points from the 'QTL' tag in the metadata sheet to the 'tags' field in the dataset panel. A fifth green arrow points from the 'displayName' in the metadata sheet to the 'displayN' field in the dataset panel. A sixth green arrow points from the 'Meta data' in the metadata sheet to the 'Meta' section in the dataset panel. The 'Meta' section shows a list of metadata fields: species, Crop, displayN, ame, Comm, ents, Cit, and ati.

'shortName' of the 'parentName'

'Field' for dataset name

'parentName'

The 'QTL' tag is derived from the Excel Field name

'displayName'

Meta data

Wheat_IWGSC_RefSeq_v2.1 : Chr7A

Wheat_CSv2.1_7AL_QTLs-KF

Namespace

Parent Wheat_IWGSC_RefSeq_v2.1

Created 21/05/2025, 11:41:51 am

Last Updated 21/05/2025, 11:41:51 am

Public false

Read Only true

tags QTL

Group All

apiHost https://agg.plantinformatics.io

Meta View Edit

▼ Metadata {8}

species : Triticum aestivum

Crop : wheat

displayN : Triticum aestivum - IWGSC_RefSeq_v2.1

ame : - QTL - 7A yield

Comm : Prepared by Gabriel Keeble Gagnere

ents : Agriculture Victoria. The positions of the QTL were determined through BLAST of flanking marker sequences against the genome.

Cit : Kuzay S, Xu Y, Zhang J, Katz A, Pearce S, Su Z, Fraser M, Anderson JA, Brown-

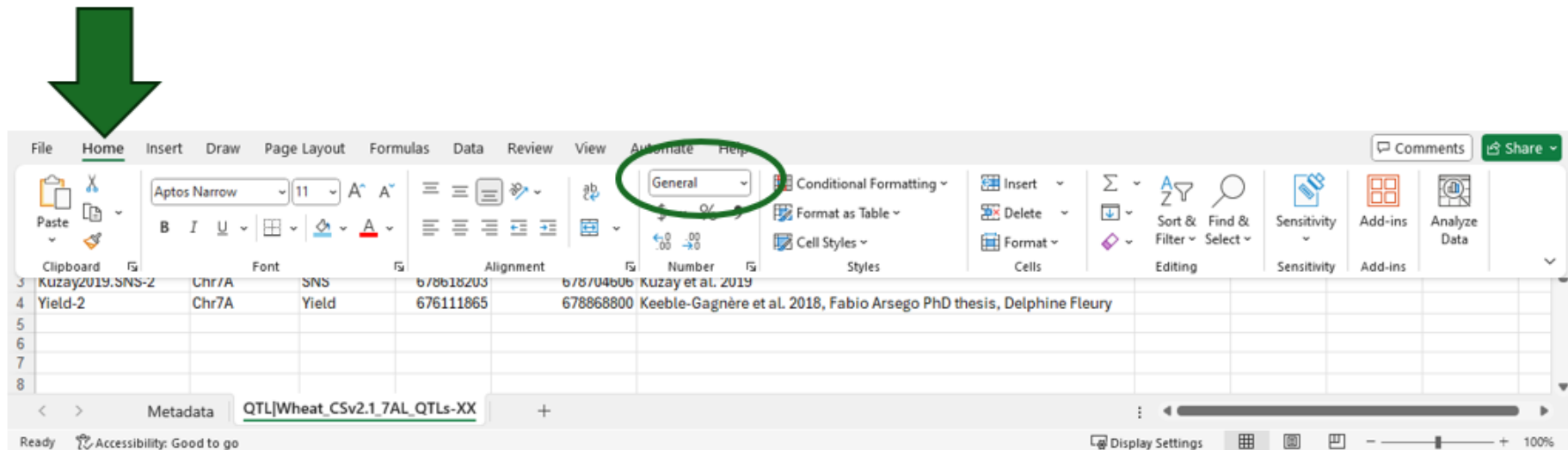
ati : -

Parameter	Details	Usage		
		QTL	Alignment	Map
Field	Name of the custom dataset to be loaded into <i>Pretzel</i>. It is structured as the type of dataset on the LHS, a pipe symbol ' ' and the unique name of the dataset on the RHS. E.g. 'QTL Dataset_XX', 'Alignment Dataset_XX' or 'Map Dataset_XX' This dataset name must be unique, it cannot match any other datasets loaded in <i>Pretzel</i>, including datasets loaded by other users. Note that this dataset name is NOT used in the left-hand side Explorer panel, it is only displayed in the 'Dataset' panel (see Figure below). This must exactly match the dataset name on the RHS of the pipe symbol in the Excel worksheet name i.e. the portion in bold e.g. QTL Wheat_CSv2.1_7AL_QTLs-KF The name of the dataset is limited to 31 characters (including e.g. 4 characters used for 'QTL ') due to limitations set by MS Excel on the sheet name. Take care to ensure there are no spaces.	mandatory	mandatory	mandatory
Species	Taxonomy e.g. <i>Triticum aestivum</i>	mandatory	mandatory	mandatory
Crop	Used to filter the datasets in the 'Explorer' panel based on the crop type e.g. Wheat	mandatory	mandatory	mandatory
displayName	Complete, descriptive name that shows up in the left-hand 'Explorer' panel	mandatory	mandatory	mandatory
parentName	Name of the genome assembly or genetic map that the feature (e.g. QTL, marker, gene) is being anchored to. This must exactly match the dataset name in <i>Pretzel</i> .	mandatory	mandatory	not applicable
shortName	Short dataset name displayed at the top of the chromosome axis in <i>Pretzel</i> . Required only for a genome or genetic map dataset.	not applicable	not applicable	mandatory
platform	Type of technology platform used to generate data e.g. 90k, 40k, SSRs, etc. If multiple types of markers were used e.g. in the case of a consensus map, then the use of 'multiple' is recommended.	not applicable	not applicable	mandatory
licensing	E.g. CC BY 4.0 (or 'none' if not applicable)	optional	optional	optional
Citation	Reference to a manuscript, 'personal communication (Researcher name)', 'unpublished' etc.	optional	optional	optional
DOI	DOI links the uploaded data to the source e.g. manuscript	optional	optional	optional
Comments	Longer description of the dataset, and/or information on the study, or how the data was compiled. This could include: - parents and filial generation of a mapping population - information about the germplasm (e.g. diverse landraces) - trait information - analysis and/or filtering steps taken	optional	optional	optional
Contact	Person / organisation who uploaded the data (e-mail address encouraged)	optional	optional	optional

Dataset sheet

The fields in the '**Dataset**' sheet are very specific to the type of data being uploaded. The names of the mandatory parameters in the LHS column below are case sensitive.

It is important that the numbers entered into the '**Start**', '**End**' and '**Position**' fields are formatted correctly. In excel, highlight the relevant cells and set the format to 'General', ensuring they do not contain any commas or thousand separators e.g. '678380602' not '678,380,602'.



Additional, optional columns can be added as desired to capture other information e.g. human-readable gene descriptions, gene identifiers.

Parameter	Details	Usage		
		QTL	Alignment	Map
Name	Name of the QTL or feature	mandatory	mandatory	not applicable
Marker	Name of the markers in the genetic map	not applicable	not applicable	mandatory
Chromosome	Name of the chromosome the feature is anchored to. You must check the name of the chromosome in the relevant dataset in <i>Pretzel</i> for the ' parentName '. E.g. if the chromosome in <i>Pretzel</i> is called 'Chr7A', then the use of e.g. '7A' will fail to anchor the feature to the chromosome.	mandatory	mandatory	mandatory
Trait	Name of the trait being investigated	mandatory	not applicable	not applicable
Start	Start position of the QTL. If you have anchored the QTL to a chromosome in a genome assembly, this should be the number in basepairs. If you have anchored the QTL to a genetic map, then the number should be in cM.	mandatory	mandatory	not applicable
End	End position of the QTL. If the QTL is associated with a SNP marker, then the Start and End positions should be the same number.	mandatory	mandatory	not applicable
Position	Position of the marker in the genetic map	not applicable	not applicable	mandatory
Ontology	Crop ontology term from https://cropontology.org/ e.g. CO_321:0000902. The data contained in this field enables <i>Pretzel</i> to colour-code QTLs based on crop ontology terms.	optional	not applicable	not applicable
Reference	Reference to a published manuscript / 'personal communication (Researcher name)', 'unpublished' etc. It can be especially helpful to list a reference for each entry if data has been compiled from multiple sources.	optional	optional	optional
Comments	Longer description of the dataset, and/or information on the study, or how the data was compiled.	optional	optional	optional

Troubleshooting

If you encounter a bug and *Pretzel* is no longer responding as expected, reset your session. The easiest way to do this is put your cursor in the URL field of your browser and delete characters back to the main URL: <https://agg.plantinformatics.io/>, then press 'Enter' on your keyboard.

If any software bugs are causing problems, please report them to the *Pretzel* development team: info@plantinformatics.io.

Contact Details

For requests to set up a group training session:

Kerrie Forrest - Senior Research Scientist

Agriculture Victoria (AgriBio, Bundoora)

kerrie.forrest@agriculture.vic.gov.au

For questions about data management, digital tools and Program 1 of the AGG Strategic Partnership:

Gabriel Keeble-Gagnere - Science Program Leader – Australian Grains Genebank

Agriculture Victoria (AgriBio, Bundoora)

gabriel.keeble-gagnere@agriculture.vic.gov.au

For questions about the Australian Grains Genebank and Program 2 of the AGG Strategic Partnership:

Sally Norton - Genebank Curator and Research Leader of the Australian Grains Genebank

Agriculture Victoria (Horsham)

sally.norton@agriculture.vic.gov.au