Connecting Germinate to Pretzel Step by Step

Australian Grains Genebank Strategic Partnership

December 2023

GRDC
GRAINS RESEARCH & DEVELOPMENT CORPORATION

AGRICULTURE VICTORIA

Australian Grains Genebank Strategic Partnership

Vision: To unlock the genetic potential of the Australian Grains Genebank (AGG) to accelerate cereal, oilseed and pulse crop improvement for the benefit of Australian grain growers

1 Genetically characterise plant genetic resources

Goal

To use DNA genotyping to characterise plant genetic resources warehoused within the AGG.

2 Digital and genomic innovation

Goal

To establish web-based data visualisation and analysis tools to make crop biodiversity within the AGG highly accessibility to researchers and breeders.

3 International genebank community engagement

Goal

To participate in international initiatives to advance best-practice genebank management and contribute to data resource and platform establishment.

4 Communities of Practice

Goal

To drive a high level of awareness for the importance of plant genetic resource conservation and their role in sustaining national productivity targets into the future.

5 Investment in infrastructure and intellectual capacity

Goal

To attract infrastructure funding and develop intellectual capacity to future proof the AGG as a key national innovation asset for global food security and climate change adaptation.

Context

- Germinate stores genotype call data in an HDF file separate from the database, and provides an API endpoint to return genotype call data as a single file per dataset. This result contains all chromosomes of the dataset.
- To enable the *Pretzel* workflow, Agriculture Victoria has added endpoints to request data for a given sample in given interval within a chromosome, enabling targeted requests to be made from *Pretzel* to Germinate.

Germinate Endpoints Used

Already existing in Germinate

- get a list of genotype datasets
  endpoint : brapi/v2/maps

- get a list of linkagegroups (chromosomes) of each dataset
  endpoint : brapi/v2/maps/{mapDbId}/linkagegroups

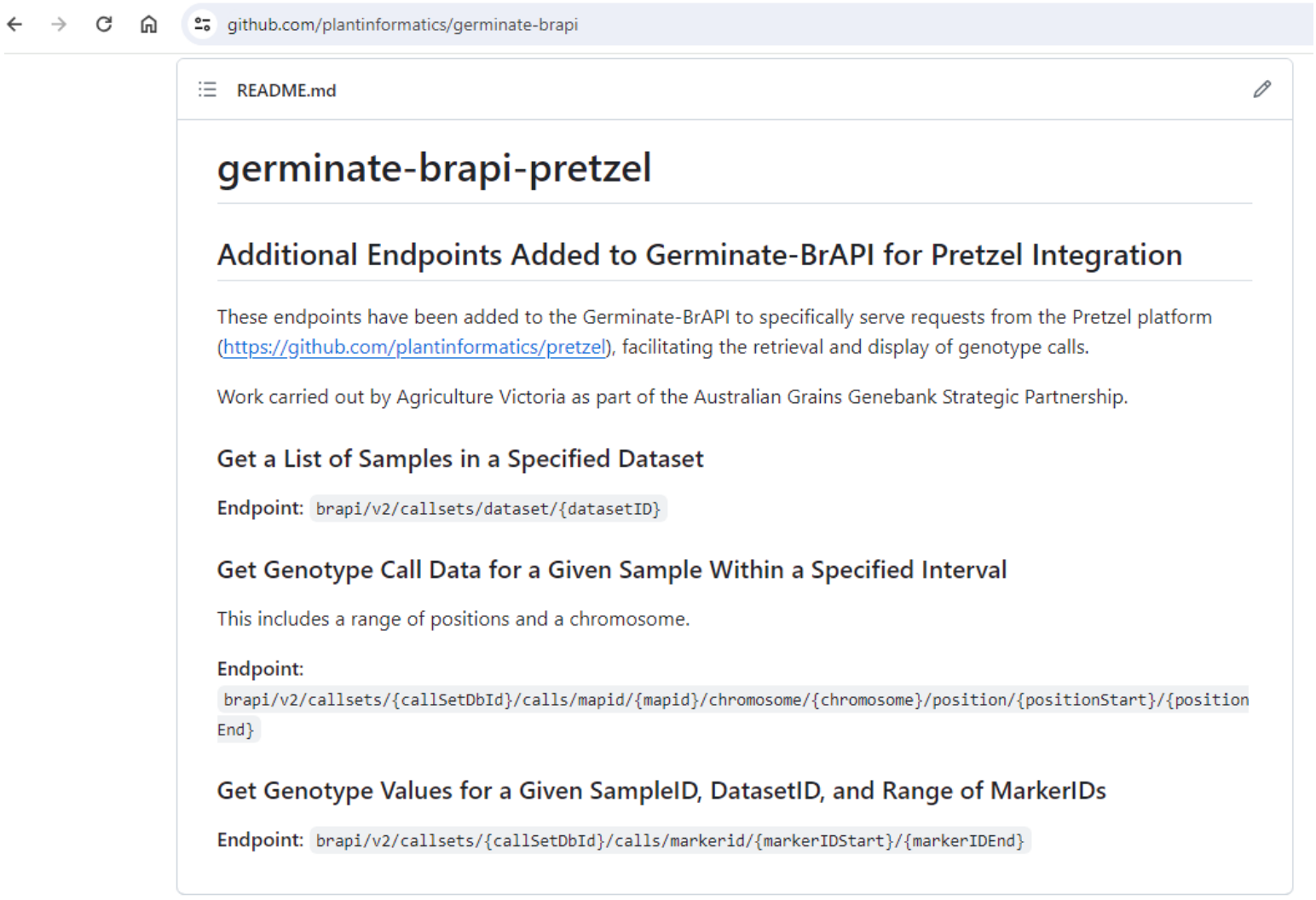
Additional endpoints added by AVR

- get a list of samples in a specified dataset
  endpoint : brapi/v2/callsets/dataset/{datasetID}

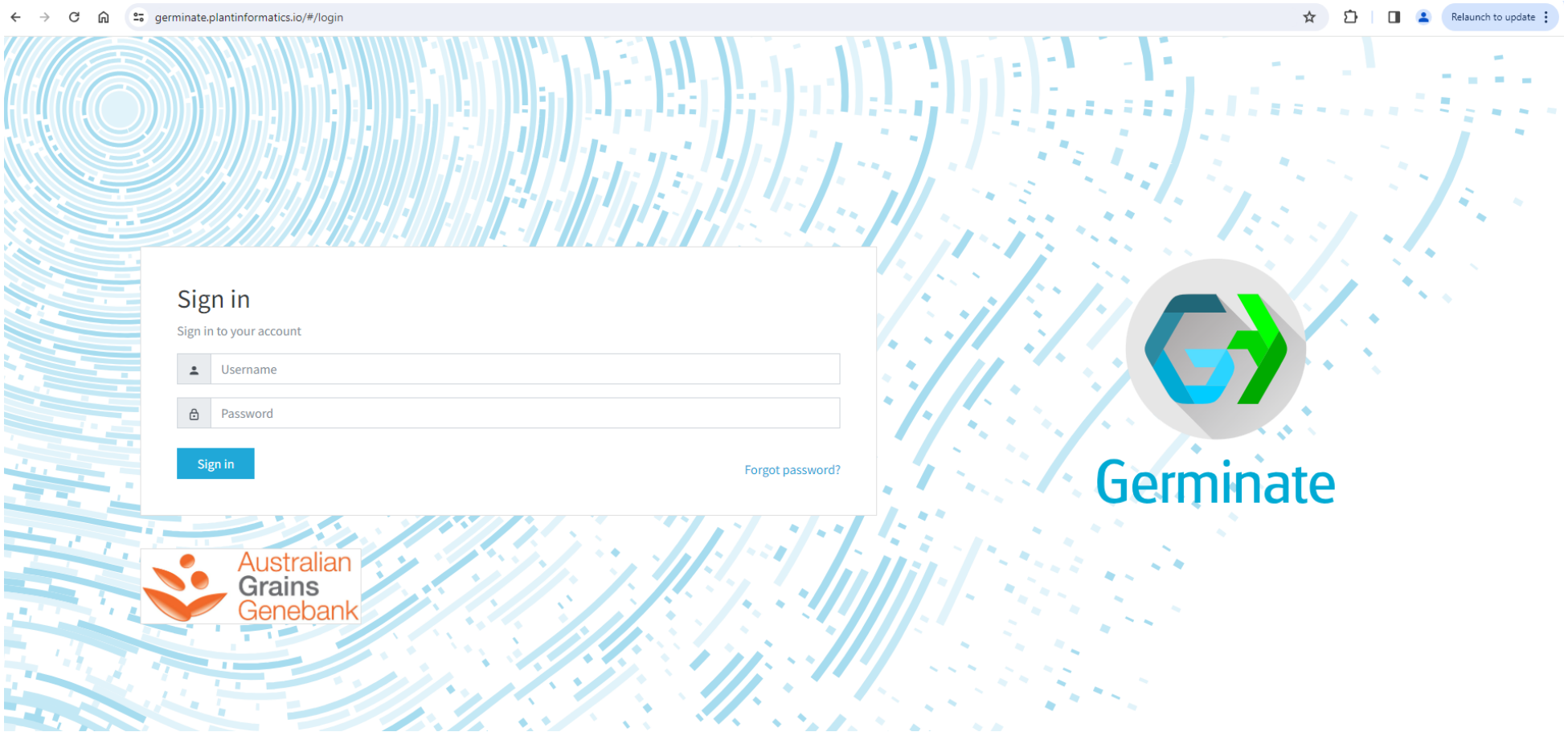
- get genotype call data for a given sample, within a specified interval (range of positions + chromosome)
  endpoint:
  brapi/v2/callsets/{callSetDbId}/calls/mapid/{mapid}/chromosome/{chromosome}/position/{positionStart}/{positionEnd}

- get genotype values for a given sampleID, datasetID, and range of markerIDs.
  endpoint:
  brapi/v2/callsets/{callSetDbId}/calls/markerid/{markerIDStart}/{markerIDEnd}

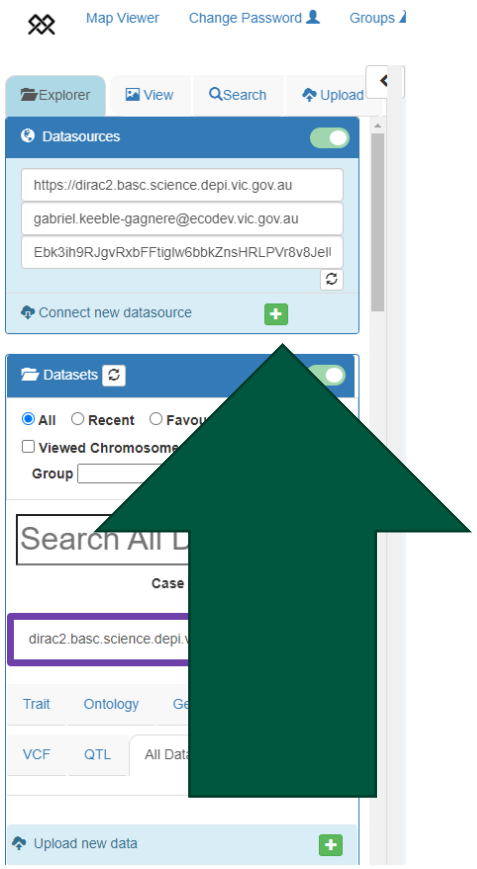
Modified Germinate Backend Is On Github

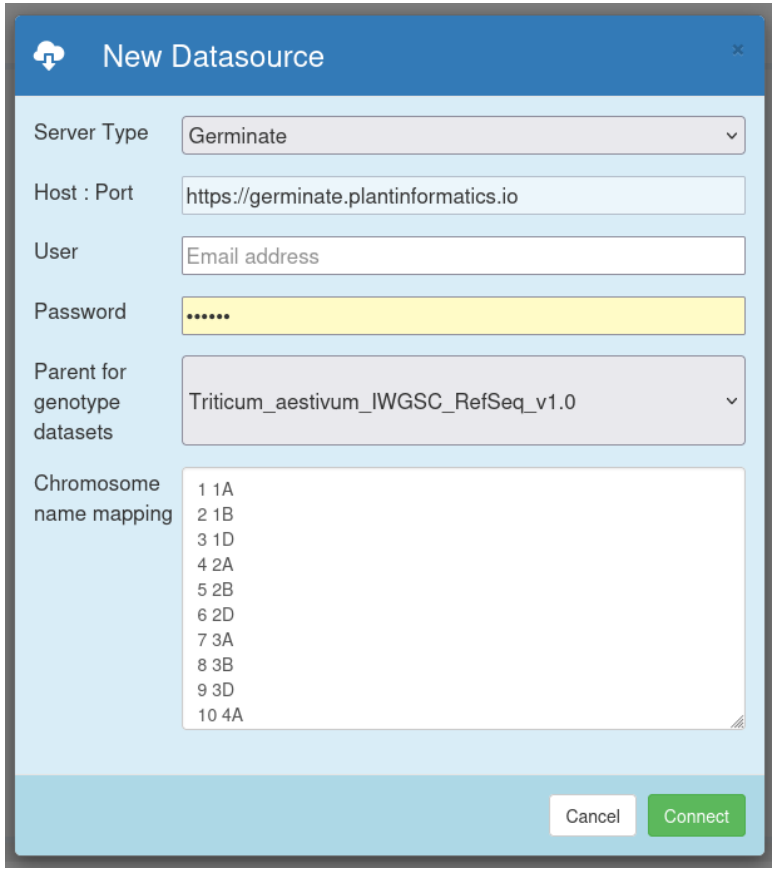


Germinate Server Running Modified Code



1) Add Germinate Server As A Datasource



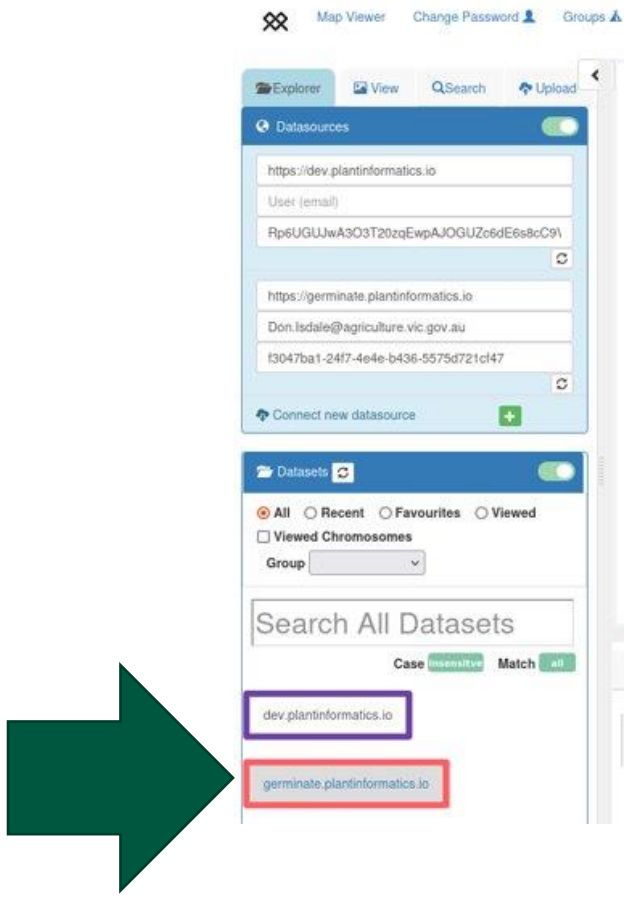


Needs to be a Germinate server running the forked code

User must define dataset to use as parent (the genome against which the genotype data is defined)

User can define mapping of chromosomes to correspond to existing chromosome names

2) User Selects The Datasource To View

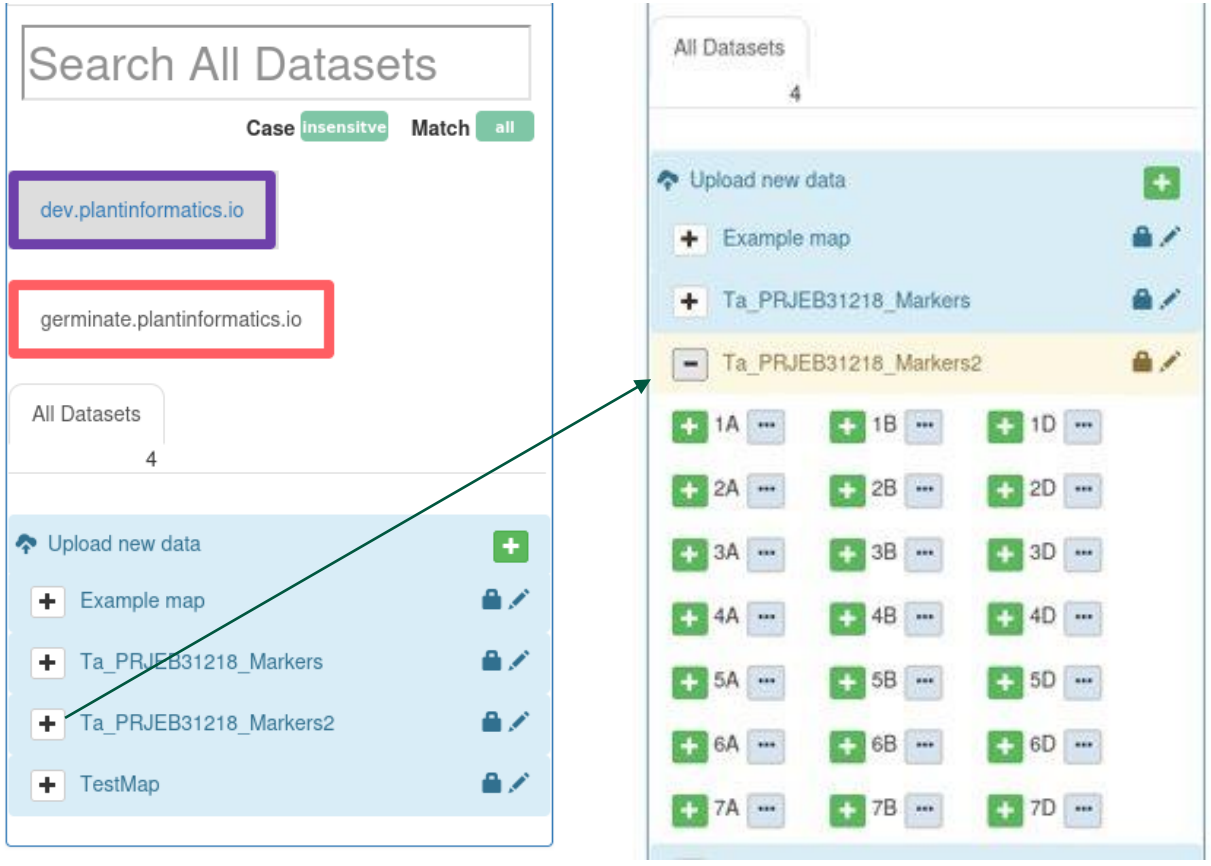


Authenticated connection to Germinate server is listed in Datasources

User can select the Germinate Datasource to view the list of datasets available

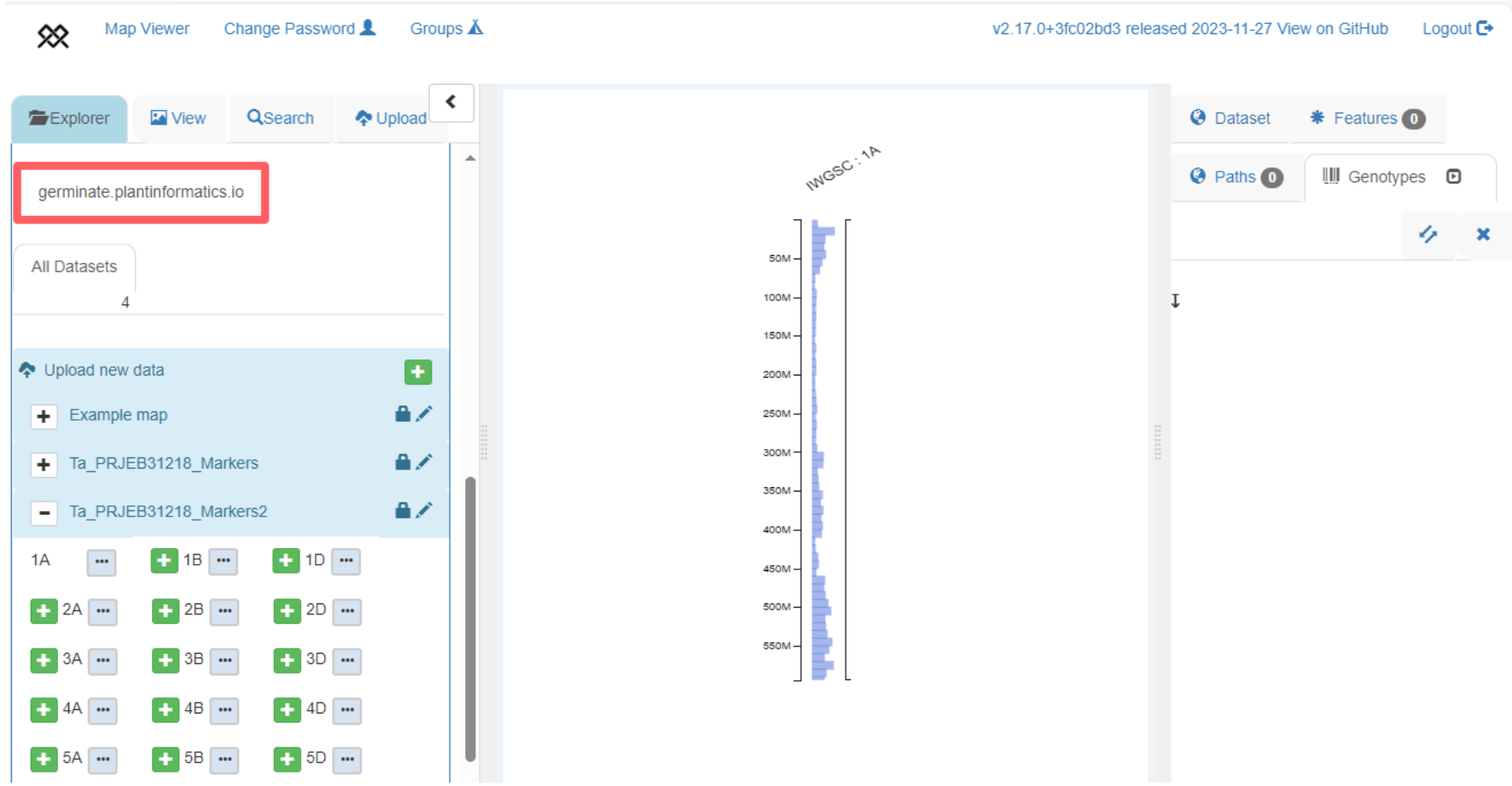
endpoint : brapi/v2/maps

3) Dataset Explorer Lists Datasets Of The Selected Source



Clicking on the dataset name displays the list of chromosomes which can be viewed
Clicking on the green + for a chromosome will display it in the graph (*note these are the chromosomes as defined in the parent genome*)

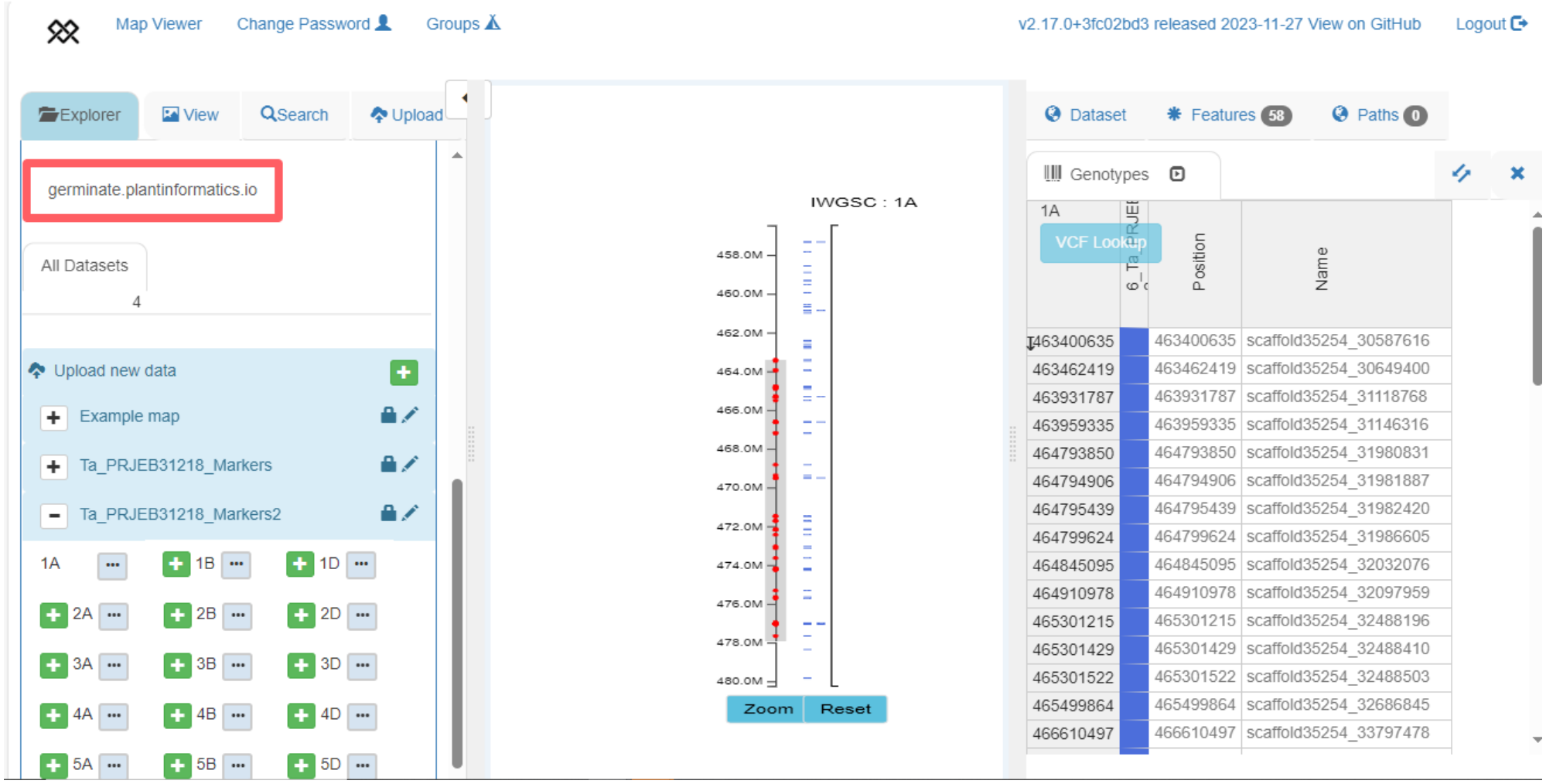
4) *Pretzel* Receives Marker Information From Germinate



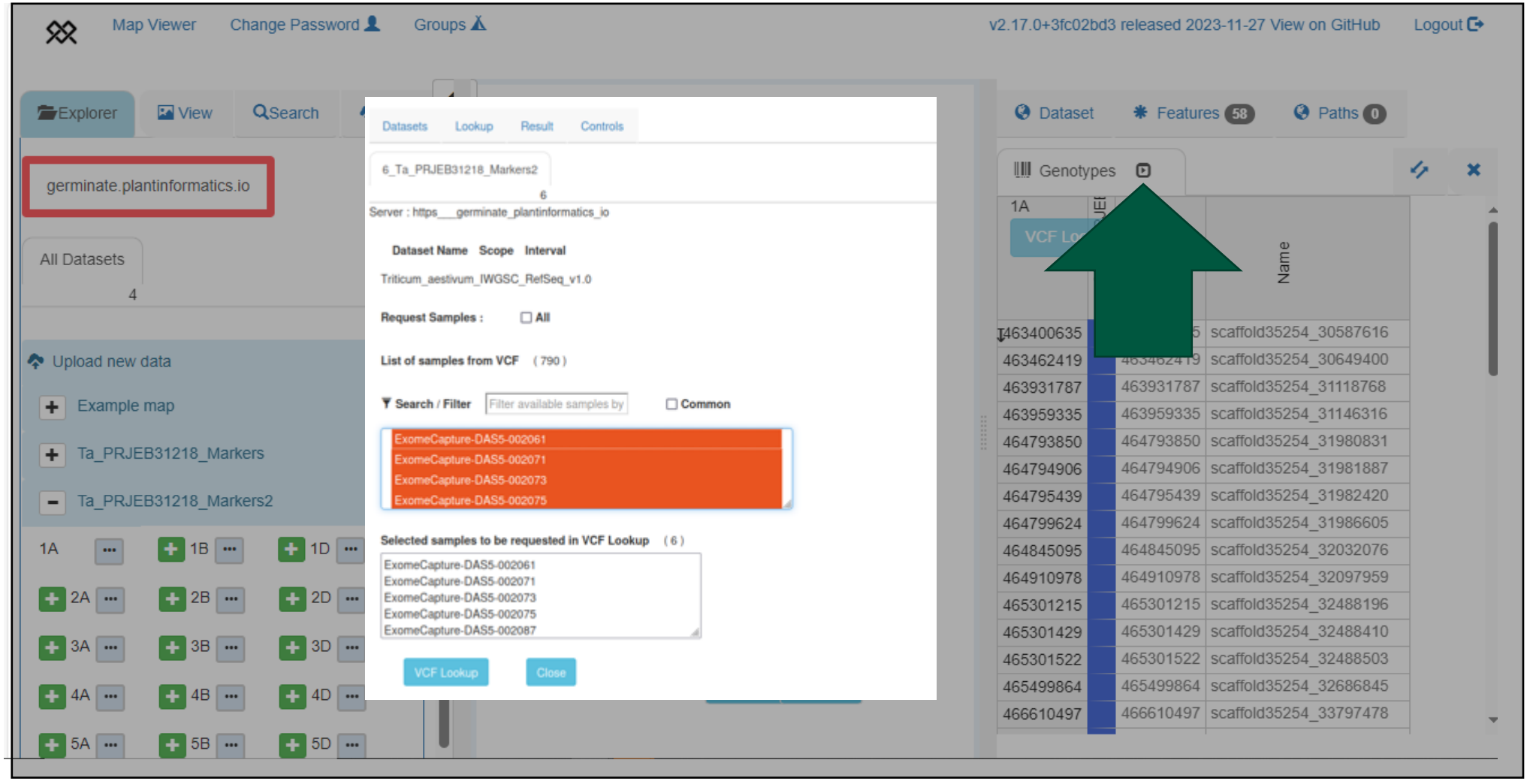
** Pretzel receives all markers for the chromosome*

- endpoint: brapi/v2/callsets/{callSetDbId}/calls/mapid/{mapid}/chromosome/{chromosome}/position/{positionStart}/{positionEnd}

5) Brushing The Axis Displays Marker Names In Right Panel

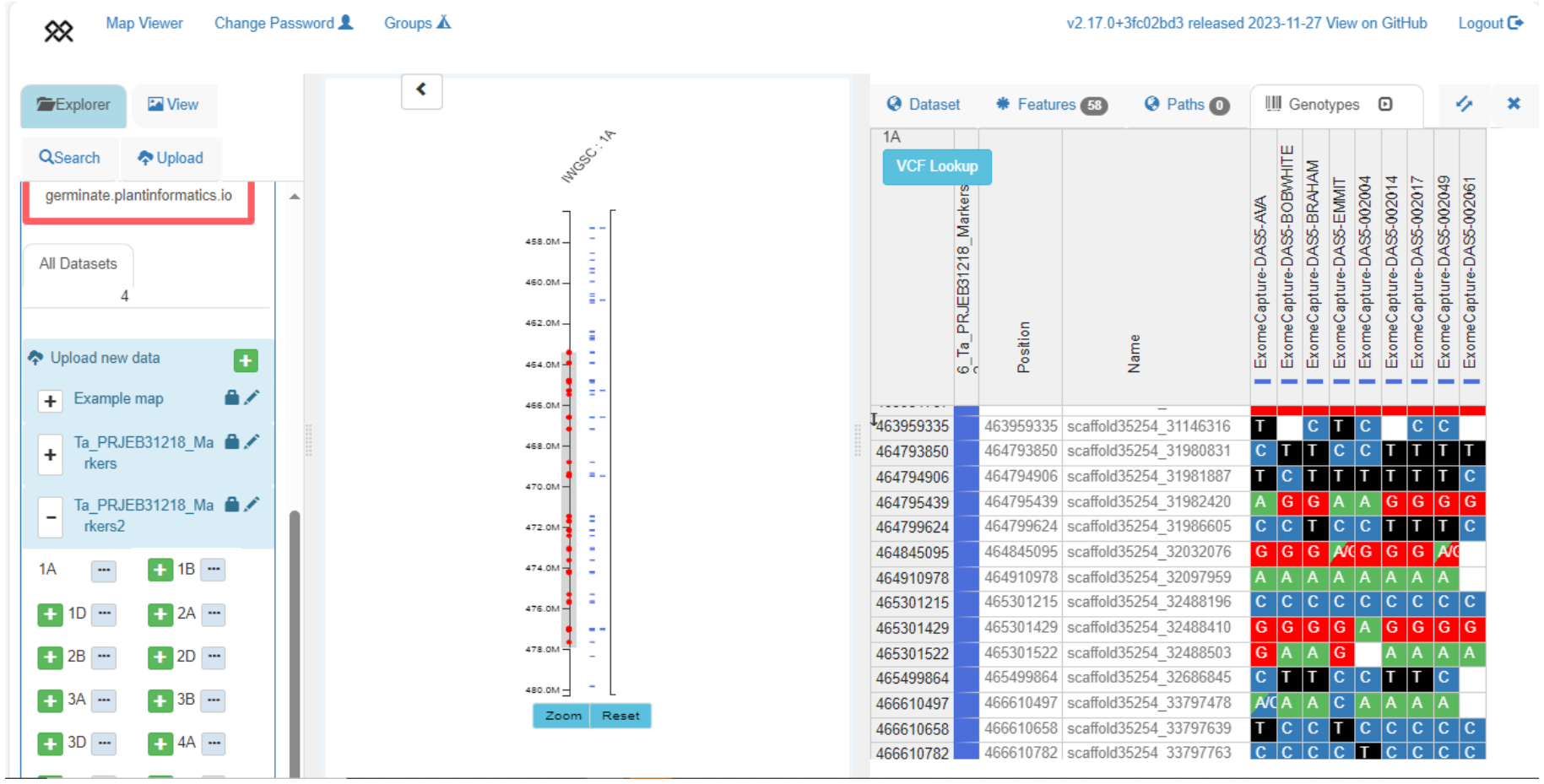


6) Sample Selection



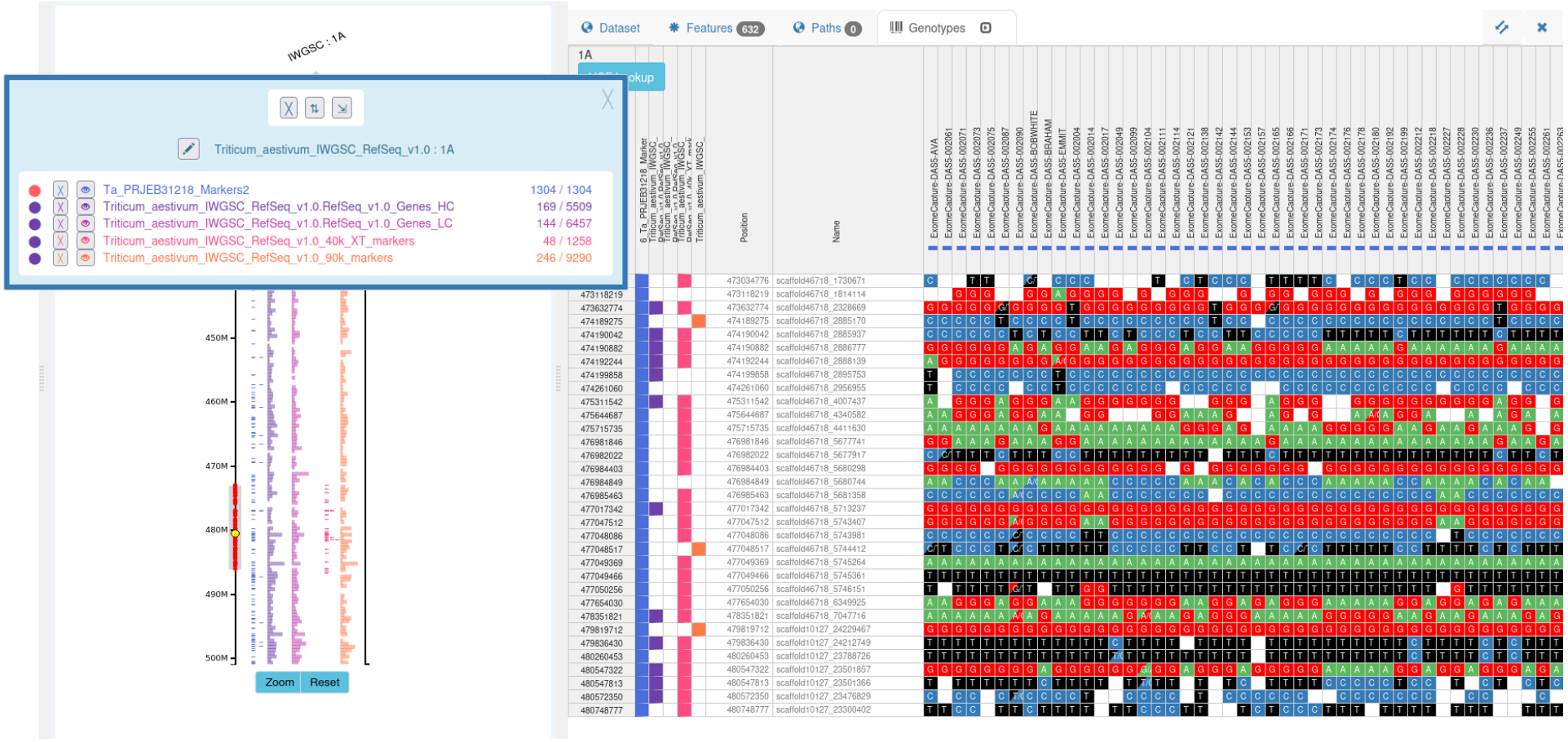
- endpoint : brapi/v2/callsets/dataset/{datasetID}

7) *Pretzel* Requests Genotypes For Each Selected Sample

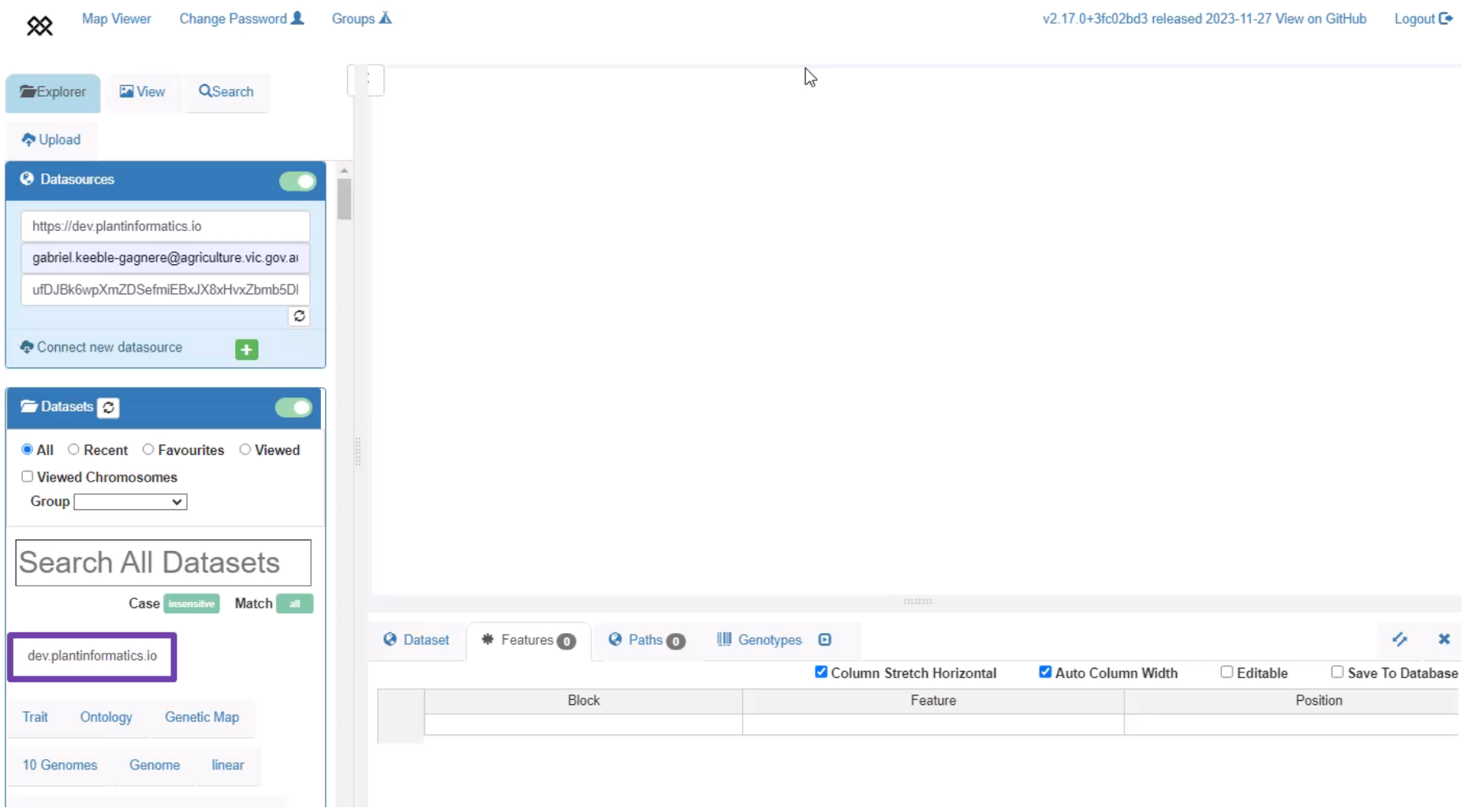


- endpoint: brapi/v2/callsets/{callSetDbId}/calls/mapid/{mapid}/chromosome/{chromosome}/position/{positionStart}/{positionEnd}

Integrating Germinate Genotype Data With *Pretzel* Data



Animation Of The Workflow



Summary

- The added endpoint does not modify any existing interface or functionality, so the *Pretzel* branch of Germinate could be used by existing Germinate servers
- This would enable Germinate users to view their Germinate data using the *Pretzel* GUI, and
- enable *Pretzel* users to integrate genotype data stored in Germinate with other *Pretzel* datasets.