

Connecting Germinate to Pretzel Step by Step

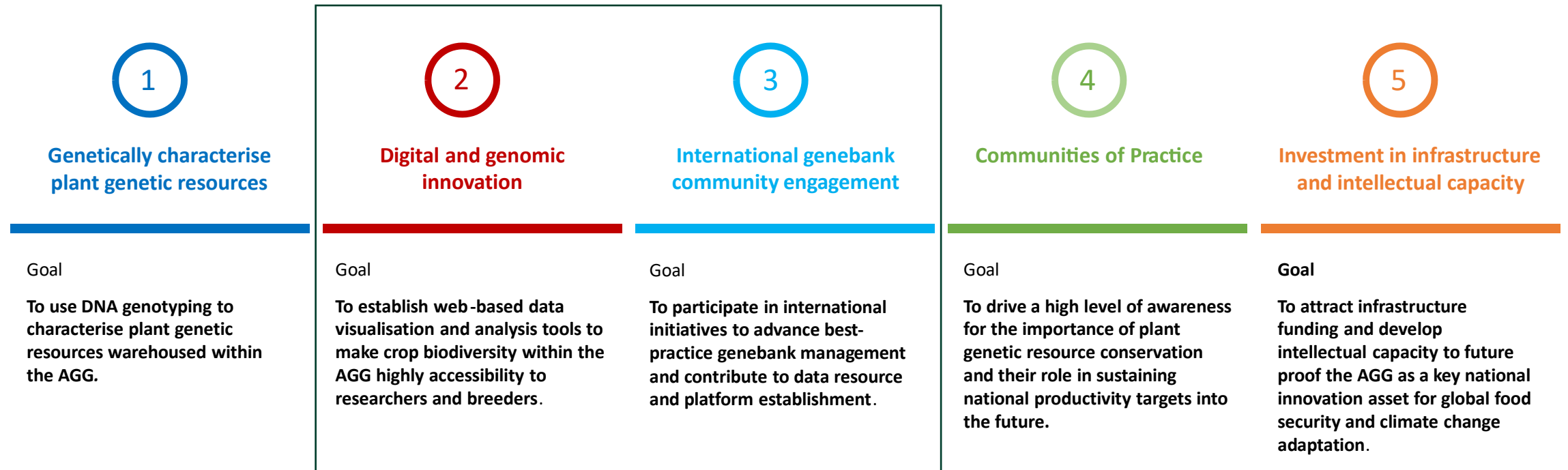
Australian Grains Genebank
Strategic Partnership

December 2023



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



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

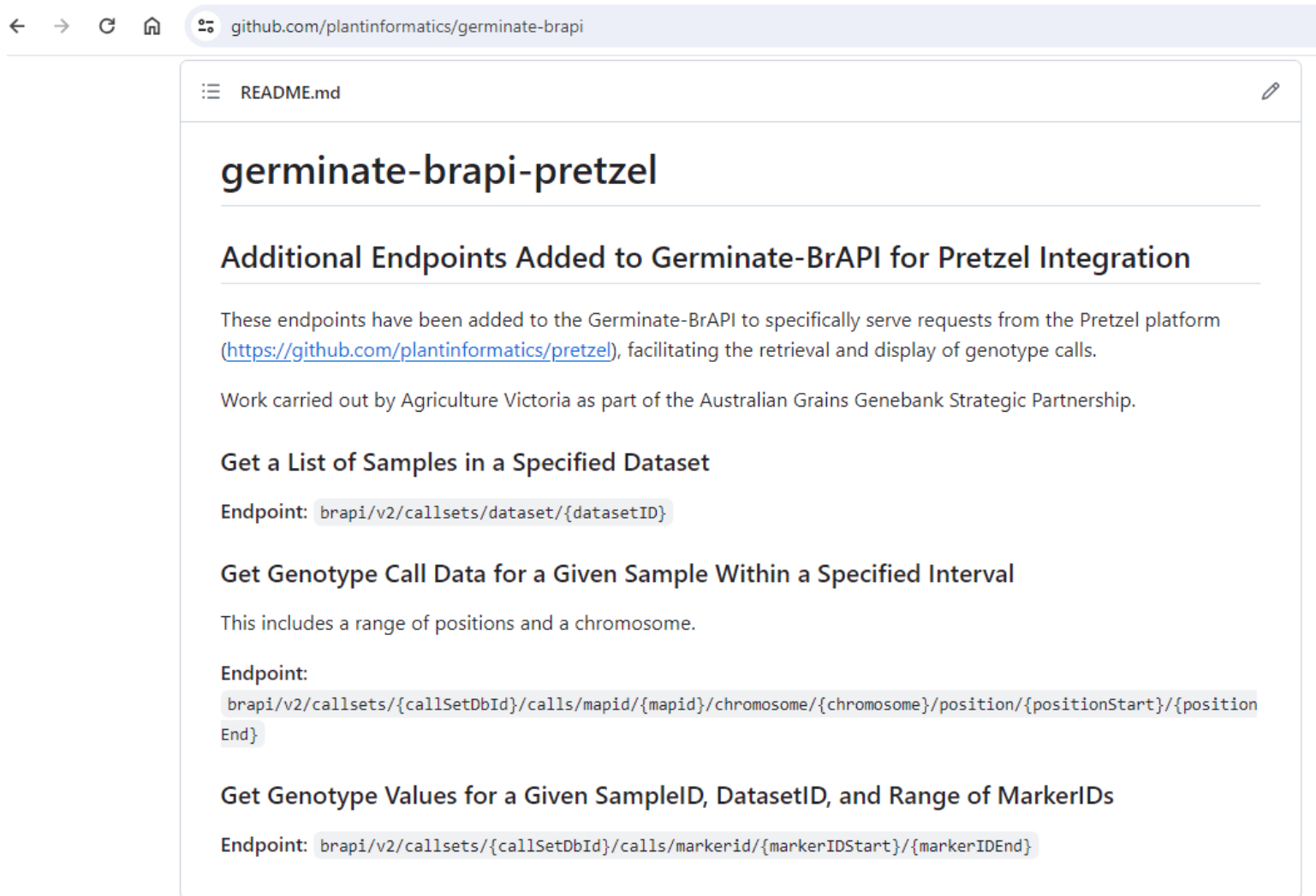
Already existing in Germinate

- 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



The image shows a browser window displaying the GitHub README for the 'germinate-brapi-pretzel' repository. The browser's address bar shows the URL 'github.com/plantinformatics/germinate-brapi'. The README content includes a title, a subtitle, a description of the endpoints, and three specific endpoint descriptions with their respective URLs.

← → ↻ 🏠 🔍 github.com/plantinformatics/germinate-brapi

☰ README.md ✎

germinate-brapi-pretzel

Additional Endpoints Added to Germinate-BrAPI for Pretzel Integration

These endpoints have been added to the Germinate-BrAPI to specifically serve requests from the Pretzel platform (<https://github.com/plantinformatics/pretzel>), facilitating the retrieval and display of genotype calls.

Work carried out by Agriculture Victoria as part of the Australian Grains Genebank Strategic Partnership.

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


This includes a range of positions and a 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}`

Germinate Server Running Modified Code

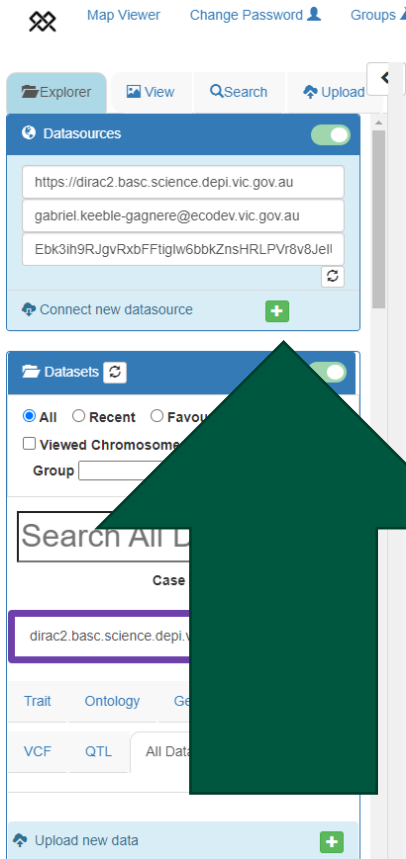


The image shows a web browser window displaying the login page for Germinate. The browser's address bar shows the URL `germinate.plantinformatics.io/#/login`. The page features a background of light blue concentric circles and a central logo consisting of a blue and green hexagon with a white arrow pointing right. Below the logo, the word "Germinate" is written in a blue sans-serif font. On the left side, there is a white "Sign in" form with the following elements:

- Title: "Sign in"
- Subtitle: "Sign in to your account"
- Username field: A text input box with a person icon on the left and the label "Username".
- Password field: A text input box with a lock icon on the left and the label "Password".
- Sign in button: A blue button with the text "Sign in".
- Forgot password link: A link labeled "Forgot password?" located to the right of the sign in button.

In the bottom left corner, there is a logo for the "Australian Grains Genebank" featuring three stylized orange grains above the text.

1) Add Germinate Server As A Datasource

A screenshot of a 'New Datasource' dialog box. The dialog has a blue header with a cloud icon and the title 'New Datasource'. The form contains the following fields:

- Server Type:** A dropdown menu with 'Germinate' selected.
- Host : Port:** A text input field containing 'https://germinate.plantinformatics.io'.
- User:** A text input field with the placeholder text 'Email address'.
- Password:** A text input field with a yellow background and a masked password '.....'.
- Parent for genotype datasets:** A dropdown menu with 'Triticum_aestivum_IWGSC_RefSeq_v1.0' selected.
- Chromosome name mapping:** A list box containing the following items: 1 1A, 2 1B, 3 1D, 4 2A, 5 2B, 6 2D, 7 3A, 8 3B, 9 3D, 10 4A.

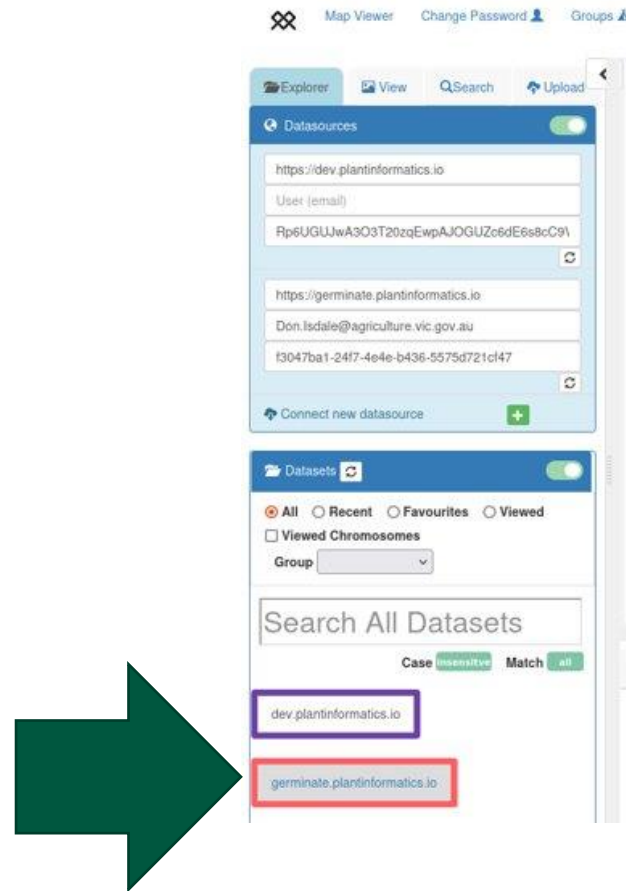
At the bottom of the dialog are two buttons: 'Cancel' and 'Connect'.

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 Databases

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

endpoint : [brapi/v2/maps](https://brapi.v2/maps)

3) Dataset Explorer Lists Datasets Of The Selected Source

The screenshot displays the Dataset Explorer interface. On the left, there is a search bar labeled "Search All Datasets" with filters for "Case insensitive" and "Match all". Below the search bar, two source URLs are listed: "dev.plantinformatics.io" (highlighted with a purple box) and "germinate.plantinformatics.io" (highlighted with a red box). An arrow points from the red box to the right-hand panel. The right-hand panel shows a list of datasets under the heading "All Datasets" (with a count of 4). The datasets listed are: "Upload new data" (with a green plus icon), "Example map" (with a lock and edit icon), "Ta_PRJEB31218_Markers" (with a lock and edit icon), and "Ta_PRJEB31218_Markers2" (with a lock and edit icon). Below these, a grid of chromosome options is shown, including 1A, 1B, 1D, 2A, 2B, 2D, 3A, 3B, 3D, 4A, 4B, 4D, 5A, 5B, 5D, 6A, 6B, 6D, 7A, 7B, and 7D, each with a green plus icon and a three-dot menu icon.

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

The screenshot displays the Pretzel web interface. At the top, there are navigation links for 'Map Viewer', 'Change Password', and 'Groups', along with the version 'v2.17.0+3fc02bd3 released 2023-11-27' and a 'Logout' button. The main interface is divided into three panels. The left panel, titled 'Explorer', shows a search bar with 'germinate.plantinformatics.io' entered and highlighted in a red box. Below it, there are options for 'All Datasets' (4), 'Upload new data', and a list of datasets including 'Example map', 'Ta_PRJEB31218_Markers', and 'Ta_PRJEB31218_Markers2'. A grid of chromosome markers (1A-5D) is visible at the bottom. The middle panel shows a vertical chromosome map for 'IWGSC: 1A' with a scale from 50M to 550M. The right panel shows a 'Dataset' view with 'Features' (0), 'Paths' (0), and 'Genotypes' (0) tabs, along with edit and delete icons.

** 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

The screenshot shows a web-based genomic data viewer. The top navigation bar includes "Map Viewer", "Change Password", and "Groups". The version information is "v2.17.0+3fc02bd3 released 2023-11-27 View on GitHub Logout".

The left sidebar contains an "Explorer" panel with a search bar containing "germinate.plantinformatics.io". Below it, there are "All Datasets" (4) and an "Upload new data" section with options for "Example map", "Ta_PRJEB31218_Markers", and "Ta_PRJEB31218_Markers2". A grid of chromosome markers (1A-5D) is visible at the bottom of the sidebar.

The main map view displays "IWGSC : 1A" with a vertical axis from 458.0M to 480.0M. A brush highlights a region from approximately 463.4M to 466.6M. A "VCF Lookup" button is overlaid on the map.

The right panel shows a table of marker data for the selected region. The table has columns for "Position" and "Name".

Position	Name
463400635	scaffold35254_30587616
463462419	scaffold35254_30649400
463931787	scaffold35254_31118768
463959335	scaffold35254_31146316
464793850	scaffold35254_31980831
464794906	scaffold35254_31981887
464795439	scaffold35254_31982420
464799624	scaffold35254_31986605
464845095	scaffold35254_32032076
464910978	scaffold35254_32097959
465301215	scaffold35254_32488196
465301429	scaffold35254_32488410
465301522	scaffold35254_32488503
465499864	scaffold35254_32686845
466610497	scaffold35254_33797478

6) Sample Selection

The screenshot displays the web interface for germinate.plantinformatics.io. The main navigation bar includes 'Map Viewer', 'Change Password', and 'Groups'. The top right corner shows the version 'v2.17.0+3fc02bd3 released 2023-11-27' and a 'Logout' button. The left sidebar contains an 'Explorer' panel with a search bar and a list of datasets, including 'Ta_PRJEB31218_Markers2'. A modal window titled '6_Ta_PRJEB31218_Markers2' is open, showing the dataset name 'Triticum_aestivum_IWGSC_RefSeq_v1.0' and a list of samples from a VCF file. A search filter is applied, showing a list of selected samples: ExomeCapture-DAS5-002061, ExomeCapture-DAS5-002071, ExomeCapture-DAS5-002073, and ExomeCapture-DAS5-002075. A 'VCF Lookup' button is visible at the bottom of the modal. On the right, a 'Genotypes' table is displayed, showing a list of scaffolds and their corresponding coordinates. A large green arrow points to the 'Name' column of the table.

Genotype	Coordinate	Name
1A	463400635	scaffold35254_30587616
	463462419	scaffold35254_30649400
	463931787	scaffold35254_31118768
	463959335	scaffold35254_31146316
	464793850	scaffold35254_31980831
	464794906	scaffold35254_31981887
	464795439	scaffold35254_31982420
	464799624	scaffold35254_31986605
	464845095	scaffold35254_32032076
	464910978	scaffold35254_32097959
	465301215	scaffold35254_32488196
	465301429	scaffold35254_32488410
	465301522	scaffold35254_32488503
	465499864	scaffold35254_32686845
	466610497	scaffold35254_33797478

- endpoint : [brapi/v2/callsets/dataset/{datasetID}](https://brapi.v2/callsets/dataset/{datasetID})

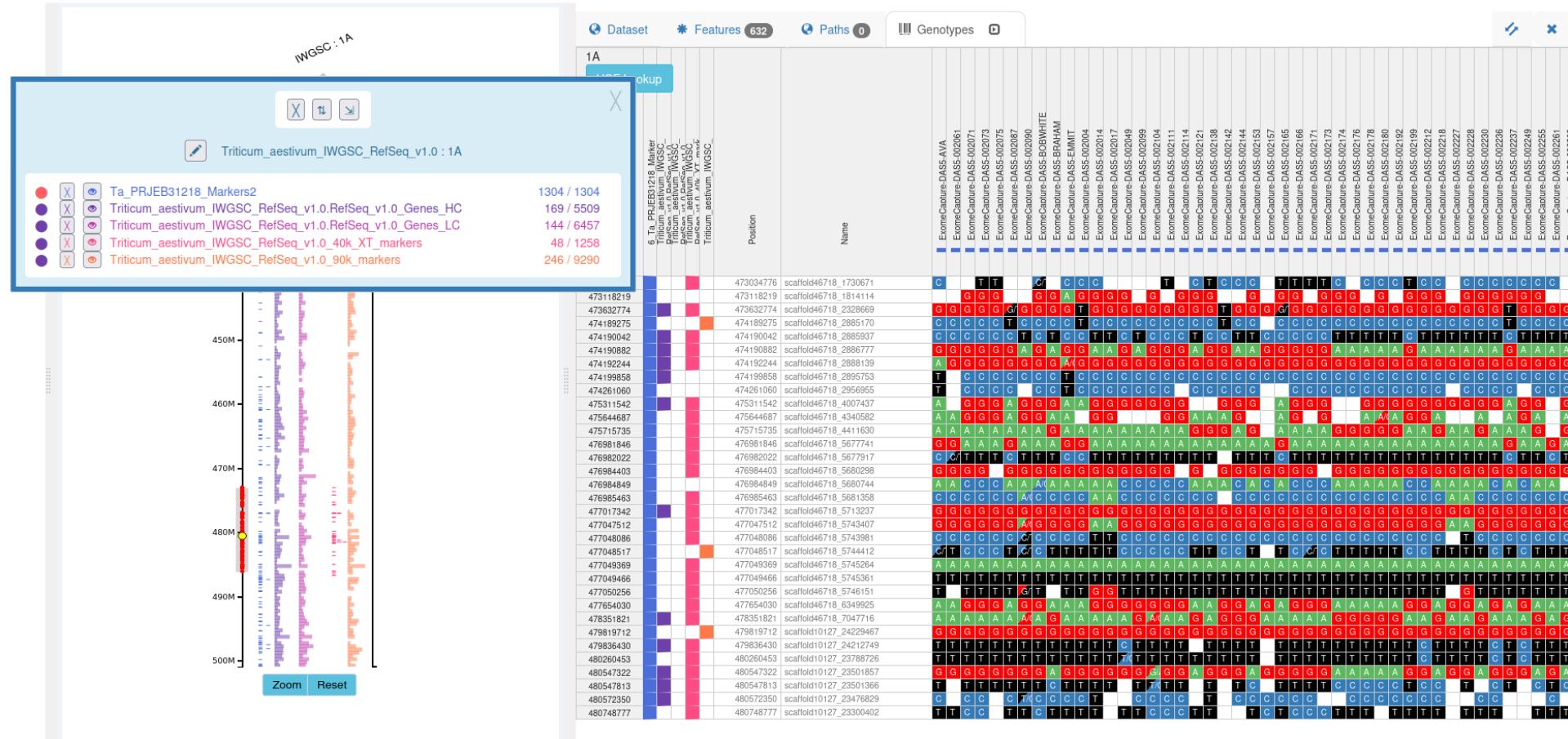
7) Pretzel Requests Genotypes For Each Selected Sample

The screenshot displays the Pretzel web interface. On the left, the 'Map Viewer' section shows a search bar with 'germinate.plantinformatics.io' and a list of datasets including 'Ta_PRJEB31218_Markers' and 'Ta_PRJEB31218_Markers2'. The central map viewer shows a genomic map with a vertical axis labeled 'NWSS : 1A' and a horizontal axis with positions from 458.0M to 480.0M. On the right, the 'Genotypes' panel shows a 'VCF Lookup' table with columns for 'Position' and 'Name', and rows of genotype data for various samples.

Position	Name	ExomeCapture-DAS5-AVA	ExomeCapture-DAS5-BOBWHITE	ExomeCapture-DAS5-BRAHAM	ExomeCapture-DAS5-EMMIT	ExomeCapture-DAS5-002004	ExomeCapture-DAS5-002014	ExomeCapture-DAS5-002017	ExomeCapture-DAS5-002049	ExomeCapture-DAS5-002061
463959335	scaffold35254_31146316	T	C	T	C	C	C	C	C	C
464793850	scaffold35254_31980831	C	T	T	C	C	T	T	T	T
464794906	scaffold35254_31981887	T	C	T	T	T	T	T	T	C
464795439	scaffold35254_31982420	A	G	G	A	A	G	G	G	G
464799624	scaffold35254_31986605	C	C	T	C	C	T	T	T	C
464845095	scaffold35254_32032076	G	G	G	A	G	G	G	A	G
464910978	scaffold35254_32097959	A	G	A	A	A	A	A	A	A
465301215	scaffold35254_32488196	C	C	C	C	C	C	C	C	C
465301429	scaffold35254_32488410	G	G	G	G	A	G	G	G	G
465301522	scaffold35254_32488503	G	A	A	G		A	A	A	A
465499864	scaffold35254_32686845	C	T	T	C	C	T	T	C	
466610497	scaffold35254_33797478	A	C	A	C	A	A	A	A	A
466610658	scaffold35254_33797639	T	C	C	T	C	C	C	C	C
466610782	scaffold35254_33797763	C	C	C	C	T	C	C	C	C

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

Integrating Germinate Genotype Data With *Pretzel* Data



Animation Of The Workflow

The screenshot shows a web application interface with a top navigation bar containing a logo, 'Map Viewer', 'Change Password', and 'Groups'. The version 'v2.17.0+3fc02bd3 released 2023-11-27' and a 'View on GitHub' link are on the right, along with a 'Logout' button.

The left sidebar contains several panels:

- Explorer**: Includes 'View' and 'Search' buttons.
- Upload**: A button for uploading data.
- Datasources**: A panel with a text input containing 'https://dev.plantinformatics.io', an email 'gabriel.keeble-gagnere@agriculture.vic.gov.au', and a long alphanumeric string. It includes a 'Connect new datasource' button with a plus sign.
- Datasets**: A panel with radio buttons for 'All', 'Recent', 'Favourites', and 'Viewed'. It has a 'Viewed Chromosomes' checkbox and a 'Group' dropdown menu.
- Search All Datasets**: A search bar with 'Case insensitive' and 'Match all' options. The search result 'dev.plantinformatics.io' is highlighted with a purple box.
- Filtering**: Buttons for 'Trait', 'Ontology', and 'Genetic Map'. Below are '10 Genomes', 'Genome', and 'linear' options.

The main content area is currently empty. The bottom panel features a toolbar with 'Dataset', 'Features 0', 'Paths 0', and 'Genotypes' buttons. It also includes checkboxes for 'Column Stretch Horizontal', 'Auto Column Width', 'Editable', and 'Save To Database'. Below the toolbar is a table with three columns: 'Block', 'Feature', and 'Position'.

Block	Feature	Position
-------	---------	----------

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.