Global genomic population structure of wild and cultivated oat reveals signatures of chromosome rearrangements.

(Bekele et al, under review, contact: wubishet.bekele@agr.gc.ca)

Tutorial 1: Using the Hexaploid Oat Diversity Project Apps

Welcome to the tutorial for the Hexaploid Oat Diversity Project apps. This guide will help you navigate and utilize the features of these R-Shiny apps effectively.

App for complete data set: https://graingenes.shinyapps.io/Avena_diversity/

App for cultivated data set: https://graingenes.shinyapps.io/Oat_diversity/

1. Overview

The project data is formatted using two R-Shiny apps:

- **Complete Project App**: Covers the entire project.
- **Cultivated Avena sativa App**: Focuses on populations primarily composed of cultivated Avena sativa.

Both apps function similarly.

2. Opening the App

When you open the app, you will see a default view of the Avena diversity data. This view is based on a rotatable Multi-Dimension Scaling (MDS) projection of the first three axes, displayed as a 3D scatter plot. The initial display includes 7 *Avena* species.

- **Change Axes**: Use the menu in the top left to select different axes.
- **Two-Axis or Three-Axis Scatter Plots**: Select 2D or 3D near the top right.
- **Select Axis to Show**: Choose which axes to display.
- **Select Coloring Choice**: The left side offers options for Species, Projects, Countries, and Populations.
- **Projects**: Highlight projects shown in Table 1, including "PanRef" project.
- **Country-Based Display**: Use the drop-down menu to change colours for accessions from different countries.
- **Populations**: Derived through sNMF software, default colors for K=21, arbitrary colors for K=12 or K=16.

3. Navigating the Plots

- **Rotate**: Hold the left mouse button and drag to rotate the projection.
- **Move**: To move the projection, hold the right mouse button and drag, or hold the control key and left mouse button while dragging.
- **Change Mouse Behaviors**: Use the icons in the top right to adjust mouse behaviours.

4. Interacting with Data Points

- **Highlight Points**: Hover over a point to see its coordinates and unique identifier. Zoom in if necessary to highlight specific accessions.
- **Add to Table**: Double-click on a point to add its information to the table at the bottom of the screen. Zoom in to avoid adding multiple accessions at once.
- **Download Table**: Click the icon at the top right to download the table. Note that it is formatted as a graphic and cannot be saved as text.

5. Display Options

- **Toggle Items**: Click legend items in the top right to turn display items on or off. Be cautious, as this changes the context of the projection.
- **Isolate an Item**: Double-clicking legend items in the top right corner will show only the selected group. For example, double-clicking the PanRef item in the Project view will display the positions of the assembled genomes in the diversity space.
- **Add Deselected Items Back**: Shift + click legend items to toggle them on or off.

Important Consideration

Some accessions may appear to be misclassified (and some probably are). For example, suppose you colour the *A. byzantina* accessions red. In that case, you will see that some of these accessions are located among the A. sativa clusters rather than the primary *A. byzantina* cluster. Our manuscript discusses why we have not used our data to reclassify accessions systematically. This is partly because there is a gradient in how certain we can be about reclassifying an accession (some might be easy, but many would be contentious). Additionally, formal reclassification would require coordinated changes in gene banks or other collections. If accessions are reclassified in the future based on these or other data, we recommend that this be done systematically in consultation with the appropriate custodians or experts. This may limit formal reclassification to situations where a given accession becomes significant enough to warrant it.

Tutorial 2: Interacting with the Oat Collection Sites Map

Welcome to the tutorial for using the Oat Collection Sites Map. This guide will help you navigate and interact with the map effectively.

1. Accessing the Map

- **Link to App**: https://www.google.com/maps/d/edit?mid=1kYjxF-c5K-VHeIth0oSCwz5zt6-P5Yo&usp=sharing

2. Overview

This map allows you to explore the collection sites of wild Avena accessions and landraces of *A. byzantina* and *A. sativa* from Spain. The accessions are grouped into populations based on their membership in one of K=21 populations. Only populations 1 through 6 are shown, as the collection sites of cultivated oats are less relevant.

3. Understanding the Icons

- **Icons**: Each icon represents the collection site of a single accession.
- **Numbers on Icons**: Indicate the population to which the accession belongs.
- **Colours of Icons**: Pre-selected to match the colours used in the R-Shiny app and manuscript figures for K=21 populations.

4. Navigating the Map

- **Zoom In/Out**: Use the mouse scroll wheel or the "+/-" icons in the bottom right corner.
- **Toggle Collections**: Use the menu on the left to turn the display of collections on or off.

5. Interacting with Icons

- **Jittered Coordinates**: Icons are slightly offset to distinguish points collected from the same site. The exact coordinates are slightly different from the true collection sites.
- **True Coordinates**: Double-click an icon to display the exact collection coordinates in the "true-lat-long" field.

Tutorial 3: Using the Sang-Based SNPs interactive data browser

Data browser: https://divbrowse.triticeaetoolbox.org/index.html

This application lets you interact with the raw data for Sang-based SNP calls from the "Matrix50" set. Here's how you can use its features:

1. Viewing Raw Data

- **Raw Data Access**: The application provides access to the raw SNP data from the "Matrix50" set.

2. Selecting genomic regions

- **Subset Selection**: You can select specific subsets of data based on a particular chromosome region.

3. Performing PCA-Based Analysis

- **PCA Analysis**: The application allows you to perform simple PCA-based data analysis for the selected regions.

4. Highlighting SNPs

- **Default Highlighting**: By default, SNPs are highlighted when they differ from the reference SNP (the copy in the Sang genome).
- **Custom Highlighting**: You can change the settings to display nucleotides in four colours, with magenta used for heterozygotes.

5. Sorting Taxa

- **Taxa Sorting**: Currently, the application does not support searching or displaying specific taxa other than sorting them alphabetically.