

# UNIQUENESS

## Table of Contents

Use of this application .....	1
Open source code .....	1
Installation .....	2
File Layout .....	4
File types .....	4
Design .....	4
Import File .....	4
Search Input .....	5
Search Criteria .....	6
Running the Application .....	6
Results Grid .....	7
Heat Map .....	7
Export Results File .....	8

## Use of this application

Uniqueness is primarily a molecular marker selection tool that allows users to select a minimum number of markers (from a database of any size) to *uniquely* identify genotypes for use in Intellectual Property / Seed Testing related activities. Uses are not limited to molecular marker selection only, rather any application where selection of small yet robust data variables are needed.

- Select small, but powerful marker sets for cost effective high throughput assays

Alternatively, Uniqueness can perform a similarity analysis to select the most similar rows among columns.

- Can be used to select similar markers or other data variables that are common among columns.

## Open source code

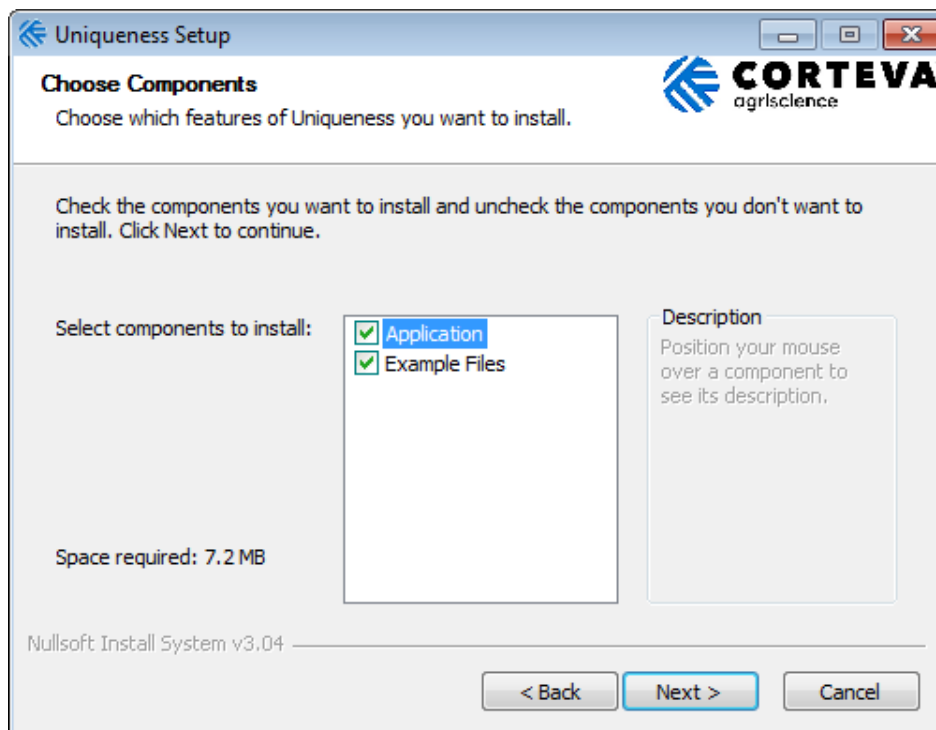
Uniqueness is written and positioned to be open source. The program can be downloaded at:  
<https://github.com/corteva>

Select 'I Agree' only if you agree to the terms of use for the application.

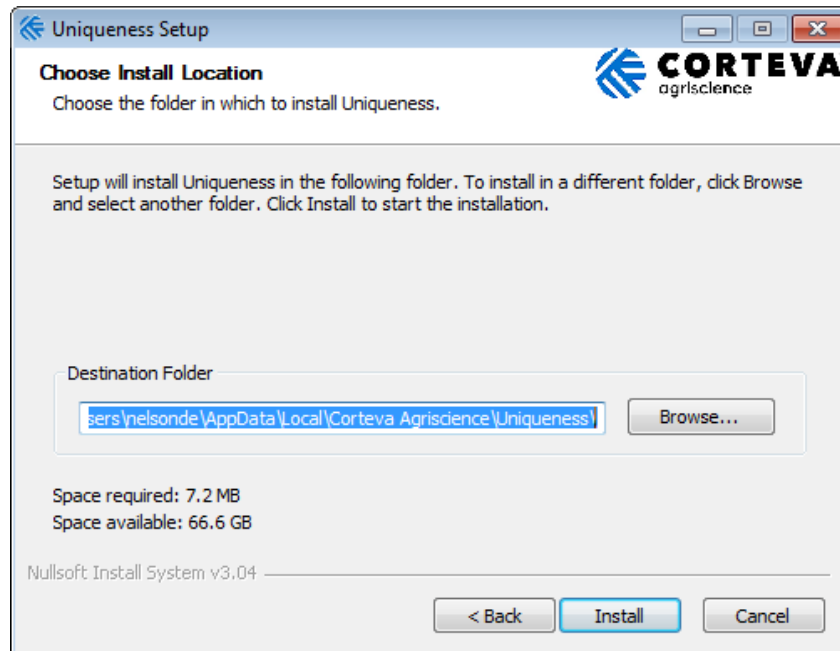


## Installation

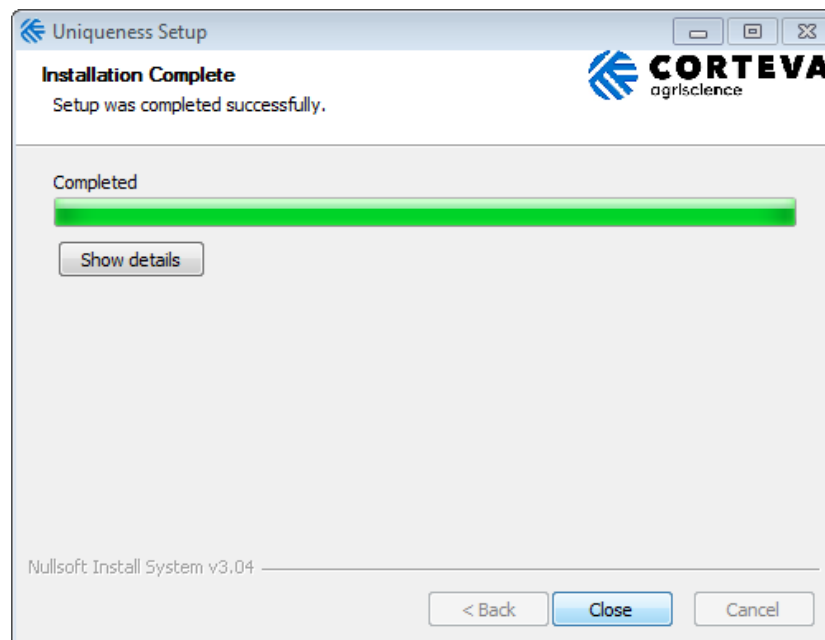
Keep 'Application' checked to install Uniqueness. The example file 'example input.csv' is loaded when the program is installed. Uncheck the example file if you do not wish to have this file.



User has the option to use the default application folder or select a different folder using the browse option.



Close the installation window. Open the application in your start up option or select from the folder destination selected.



## File Layout

### File types

The file type best suited for this application is .txt. The following file types are also acceptable; tab delimited, .csv

### Design

The first row of the file, regardless of the format must contain headers. Note: When the output file is created the column headers will be returned the same as the input file.

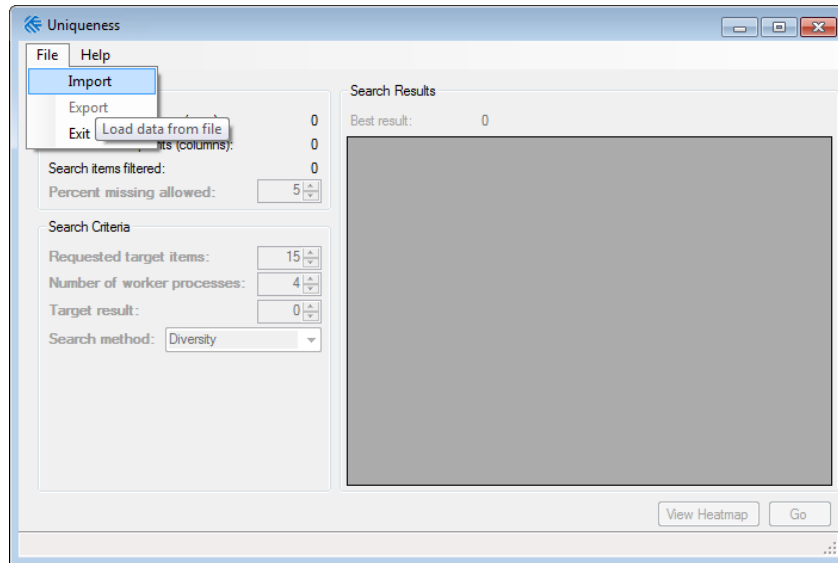
Unique Id	Name	Required	Sprite	Probst	KS4694	MN1302	Iroquois	Simpson
1	BARC_1.01_Gm_01_1013695_A_G	0	1	1	1	1	1	1
2	BARC_1.01_Gm_01_1059407_T_C	0	4	4	4	4	2	4
3	BARC_1.01_Gm_01_10768239_C_T	0	2	2	2	2		2
4	BARC_1.01_Gm_01_11835461_T_C	0	4	2	2	2		4
5	BARC_1.01_Gm_01_11904297_T_C	0	4	4	4	4	4	4

- First Column must contain a '**Unique Id**' for each row
- Second column should be the '**Name**', i.e. marker name
- Third Column signifies if the row is '**Required**' in the final results
  - 0 = Not Required
  - 1 = Required
- Remaining columns contains the Data Set.

There is no limit in place with regards to the number of rows or columns, however please keep in mind the speed of the returned results is dependent on your computer memory.

## Import File

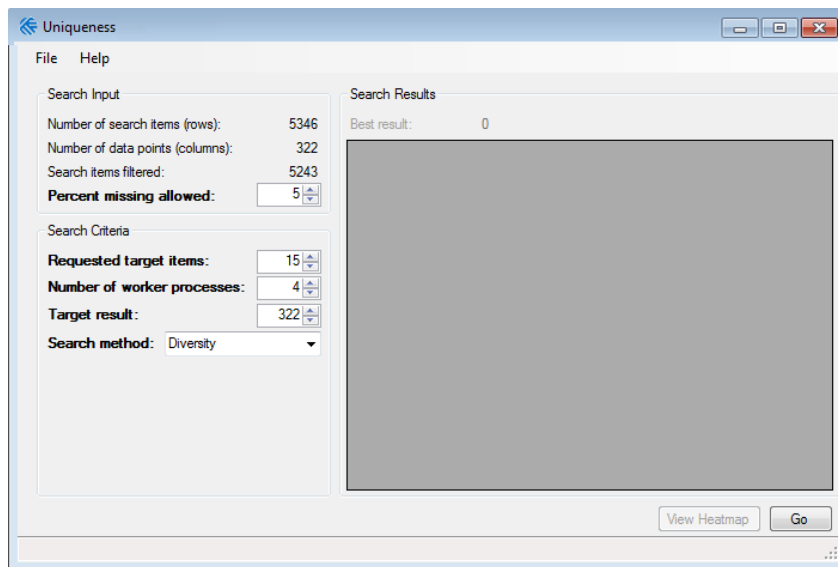
User must import file from the desired file path. If it is not in the correct format, as indicated above, the file will not import.



The User Interface will remain disabled until the file is imported. On import, the program automatically computes percent missing data for each row.

## Search Input

Results from the data input file.



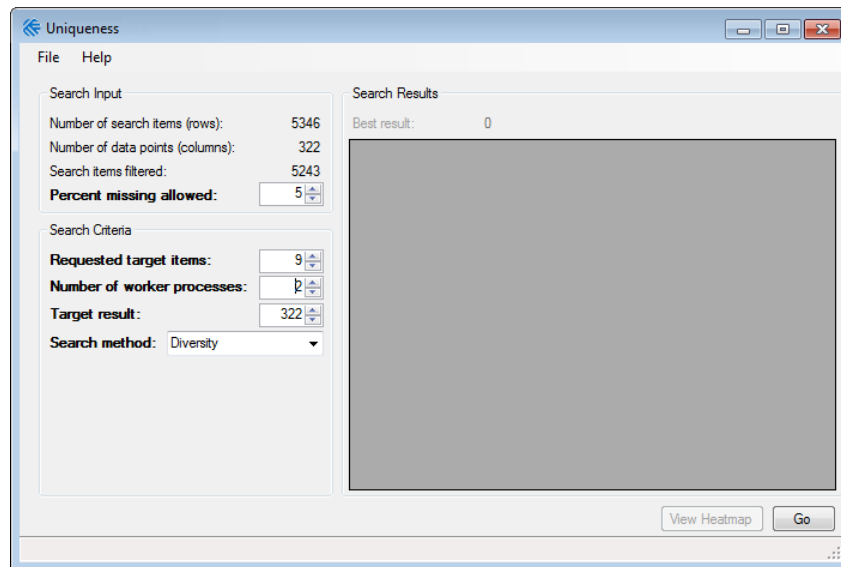
\* Bold variables indicate where the user can adjust default values.

- 'Number of search items' (rows): This field indicates the original number of rows included in the data set.
- 'Search items filtered' (rows): When the user adjusts the percent missing allowed, the data set is auto-filtered. This field tells the user the number of rows available for evaluation.
- 'Number of data points' (columns): This field informs the number of columns being evaluated.

- 'Percent missing allowed': This number can manually be adjusted to review the percent missing in the data set.
  - The minimum number is 0.
  - The maximum number is 100.

## Search Criteria

Allows user to adjust parameters used for searching.



\* Bold indicates user can adjust default value.

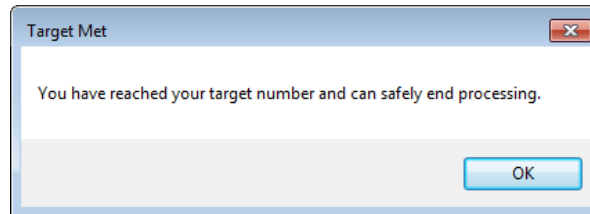
- 'Requested target items': This value corresponds to the number of rows desired by the user.
  - Default set to 15
  - For marker selection, this would be the fewest number of markers to achieve differentiation of fields (columns).
- 'Number of worker process': Automatically read from your pc. This can be cut to allow your computer to do additional tasks or increased to review data set faster.
  - Default: varies by computer
- 'Target Result': Allow the user to manually adjust the number of results to find
  - Defaults to the number of columns in data set.
- 'Search Method' Diversity or Similarity: Allows the user to adjust the application to search the data set for similarity or differences of target items.
  - Defaults to 'Diversity' to select 'Requested target items' (rows) to uniquely differentiate 'Target result' (columns).
  - 'Similarity' performs the opposite analysis selecting the most similar rows among columns.

## Running the Application

Select 'Go' to begin evaluation.

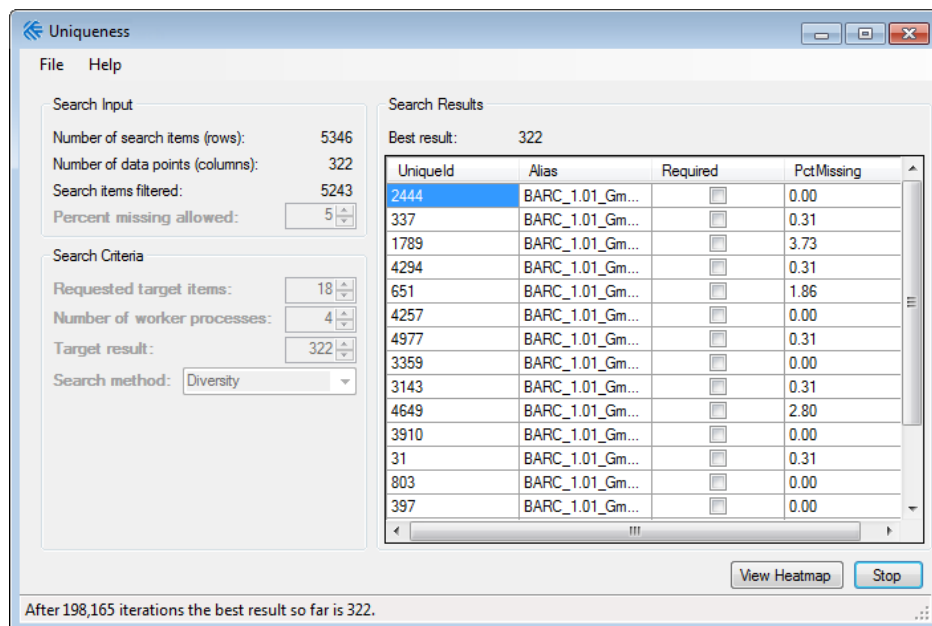
If you wish to stop the application the 'Stop' button will (formerly the 'Go' button) will cease processing.

When the target results have met the number of data points a notification will display indicating you can get your results. Even after this notification displays, the program continues processing until you click the 'Stop' button.



## Results Grid

The results grid will display the records found in the data set.



Best Result: Total number of differentiated columns based on search criteria

- Uniquelid: First column in the data set used to identify the record
- Alias: Name supplied in the input data set, i.e. marker name.
- Required: Checked if the row was required in the input data set.
- PctMissing data: Computed percent missing data for each row (record) in the data set.

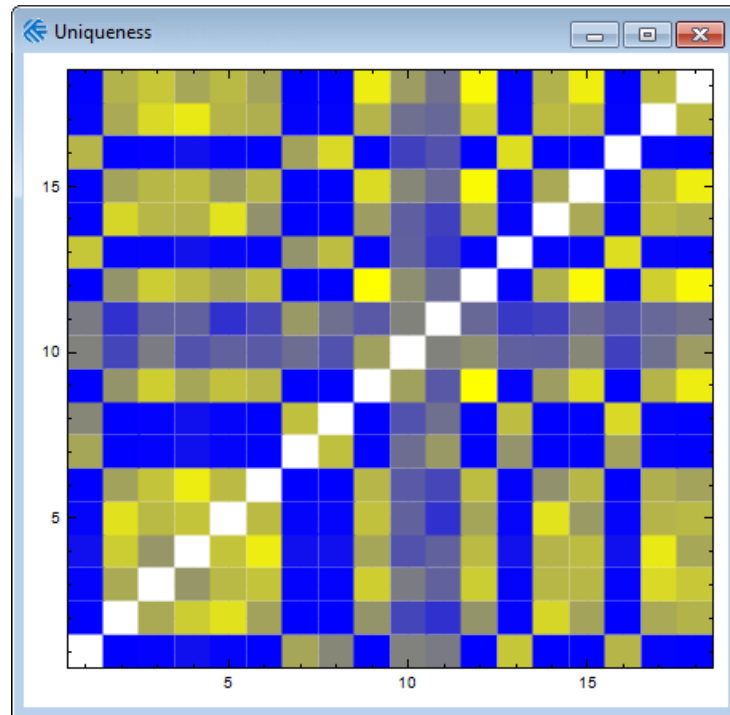
## Heat Map

The heat map will indicate the differences in the results grid.

After 'Go' is selected View Heatmap will be enabled. The heatmap will open in a new window.

When the user hovers over the cell in the heatmap a difference score will display between two row's aliases.

The color key displays from yellow to blue with yellow displaying less diversity and blue showing an increase.



## Export Results File

Original file headers will be included in the output file. The application will default back to original file path. The file name will default to the original file name with the date, time (24 hr with seconds) underscore "result" as the file name.

If the user has not stopped the application the export functionality will not work.