

**Working Group on Biochemical and Molecular Techniques  
and DNA-Profiling in Particular**

**BMT/18/11 Add.**

**Eighteenth Session  
Hangzhou, China, October 16 to 18, 2019**

**Original:** English  
**Date:** November 7, 2019

---

**ADDENDUM TO  
REPORT ON DEVELOPMENTS OF A SOFTWARE TOOL FOR MARKER SELECTION USING THE  
TRAVELING SALESMAN ALGORITHM**

*Document prepared by an expert from the Seed Association of the Americas (SAA)*

*Disclaimer: this document does not represent UPOV policies or guidance*

The annex to this document contains a copy of a presentation on “Uniqueness – A software program to aid in selecting informative molecular marker sub-sets”, made at the eighteenth session of the BMT.

[Annex follows]

# UNIQUENESS – A software program to aid in selecting informative molecular marker subsets

Barry Nelson  
Corteva Agriscience  
Seed Association of the Americas

## Uniqueness - Purpose

### *My uses:*

- Tool to select small, yet powerful SNP sets for variety ID applications
  - Intellectual property issues
  - Seed quality assurance

# Uniqueness - Background

## Algorithm

- Similar to The Traveling Salesman Problem
  - ‘Selected the most efficient route while visiting each city one time’
  - Applied to markers: ‘Select the fewest number of markers to differentiate the most varieties’.



# Uniqueness – Input File Matrix Example

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

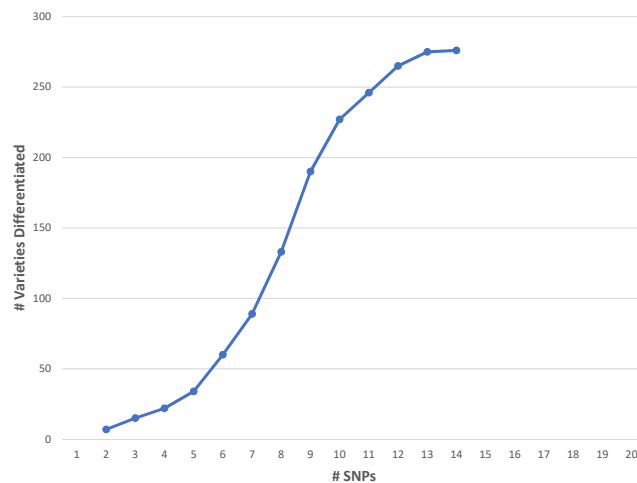
1=AA  
2=CC  
3=GG  
4=TT

## Using Uniqueness – A SNP Marker Example

The screenshot shows the 'Uniqueness' software interface. On the left, a 'Search Input' panel contains several parameters: 'Number of search items (rows): 5346', 'Number of data points (columns): 322', 'Search items filtered: 5243', and 'Percent missing allowed: 5'. Below this is the 'Search Criteria' section with 'Requested target items: 15', 'Number of worker processes: 8', 'Target result: 322', and 'Search method: Diversity'. On the right, the 'Search Results' panel shows 'Best result: 0'. A 'View Heatmap' button and a 'Go' button are at the bottom right. Red callout boxes point to various elements: 'Total number of SNPs' points to the search input area; 'Total number of GEs' points to the search criteria; 'SNPs remaining after missing data filter' points to the search items filtered; '%missing data allowed' points to the percent missing allowed; 'Target # of SNPs to select' points to the requested target items; '# of processors used; read automatically from users pc' points to the number of worker processes; '# of GEs to differentiate' points to the target result; 'Diversity for differentiating SNPs; Similarity for redundant SNPs' points to the search method; 'Highest # of GEs differentiated by iteration' points to the best result; and 'Identity of SNPs selected with details' points to the search results area.

## Uniqueness – Results

Uniqueness performed very well in selecting 14 SNPs to differentiate a database of 276 soybean varieties.



## Things to consider using Uniqueness

- Your results are only as good as your input data set.
- Heterozygotes in inbred or variety profiles are a pain.
  - Remove if they don't significantly impact your data set.
- Inbred data will allow differentiation of inbred parents and hybrids in hybrid crops.
- Selecting redundant marker sets that each uniquely identify germplasm can add durability and confidence in results.
- Uniqueness is a 'tool' among others in the tool box.

## Uniqueness – Other Applications

- Core germplasm selection
- Evaluate morphology traits
- Many others I'm sure I haven't thought of!

## Uniqueness – Open Source

Uniqueness is written and positioned to be open source.

- The program can be downloaded at: <https://github.com/corteva>

Thank you!  
Questions?