E

BMT-TWA/Maize/2/5 Add.
ORIGINAL: English
DATE: December 3, 2007

UPOV

**INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS**

GENEVA

# *AD HOC* CROP SUBGROUP ON MOLECULAR TECHNIQUES FOR MAIZE

## Second Session
## Chicago, United States of America, December 3, 2007

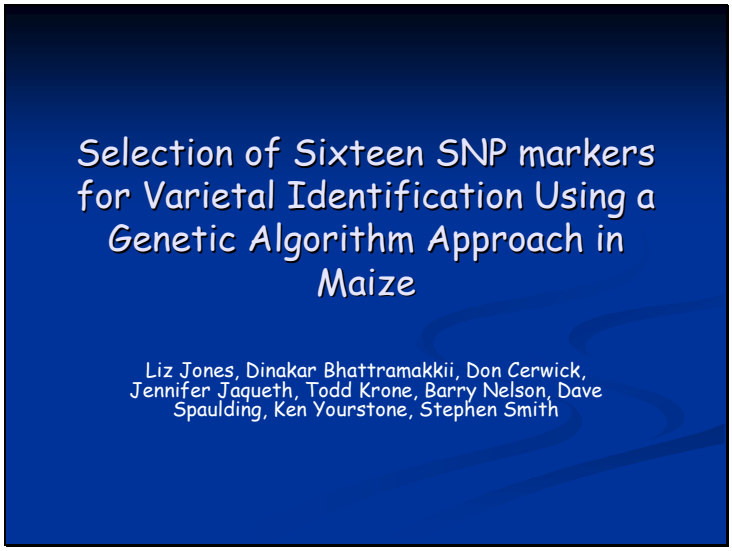
ADDENDUM TO DOCUMENT BMT-TWA/MAIZE/2/5

SELECTION OF SIXTEEN SINGLE NUCLEOTIDE POLYMORPHISM (SNP) MARKERS FOR VARIETAL IDENTIFICATION USING A GENETIC ALGORITHM APPROACH IN MAIZE INBREDS

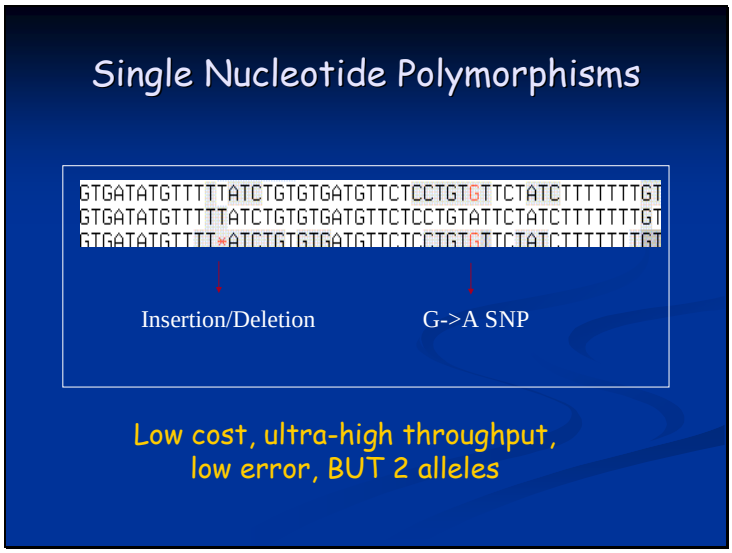
*Document prepared by experts from Pioneer Hi-Bred International*

This document is an addendum to document BMT-TWA/Maize/2/5 “Selection of Sixteen Single Nucleotide Polymorphism (SNP) Markers for Varietal Identification Using a Genetic Algorithm Approach in Maize Inbreds” and contains a copy of the presentation made by experts from Pioneer Hi-Bred International at the second session of the *Ad Hoc* Crop Subgroup on Molecular Techniques for Maize.

Slide 1




Slide 2




Here a segregating population of maize individuals are being interrogated as to whether they are homozygous for the A allele (blue) or homozygous for the C allele (red) or heterozygous (yellow). By the time profiles from 30 or more SNP loci are interrogated each maize inbred essentially has a fingerprint.

Slide 3

## Using SNPs for Varietal Identification

- Can bi-alleleic markers give sufficient resolution to be used in variety identification?
- Need small numbers to be inexpensive enough to be routinely used in variety identification
- How do we select the best set of SNPs that together can most effectively identify maize germplasm?

Slide 4

## Genetic Algorithm Approach

- Genetic algorithm: A search technique used to find exact or approximate solutions to problems
- Uses techniques inspired by evolutionary biology such as inheritance, mutation, selection and recombination
- 'Randomly place an item into a set and then test the result to see if it is better or worse than the original set. Once the replacement strategy settles on a plateau, it randomly replaces within that set in an attempt to find a higher plateau. The process repeats thousands of times and you will get a very good answer rather quickly.'

Slide 5

## Genetic Algorithm Approach to Finding Small Numbers of Markers

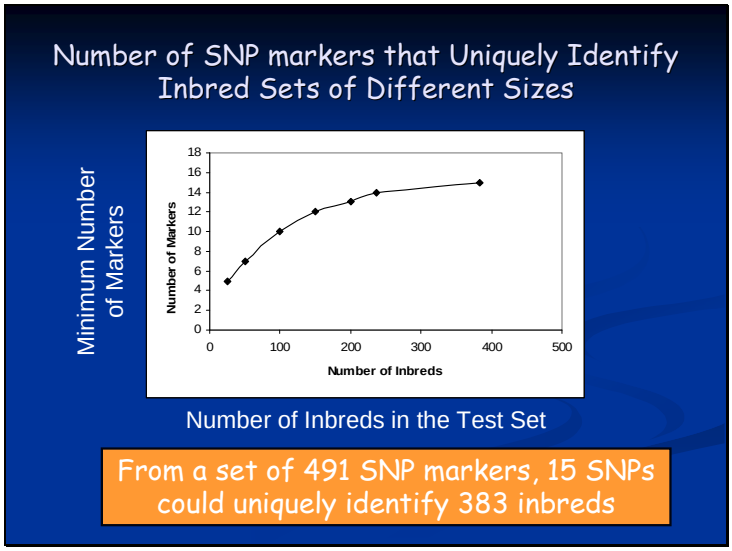
- The 'Travelling Salesman' Problem (TSP)
  - 'I thought selecting markers would be much like selecting the order of cities in the Travelling Salesman Problem'
- Proprietary software, 'Uniqueness', to find:
  - the least number of inbreds that uniquely define a marker set
  - the least number of markers that uniquely define an inbred set

Slide 6

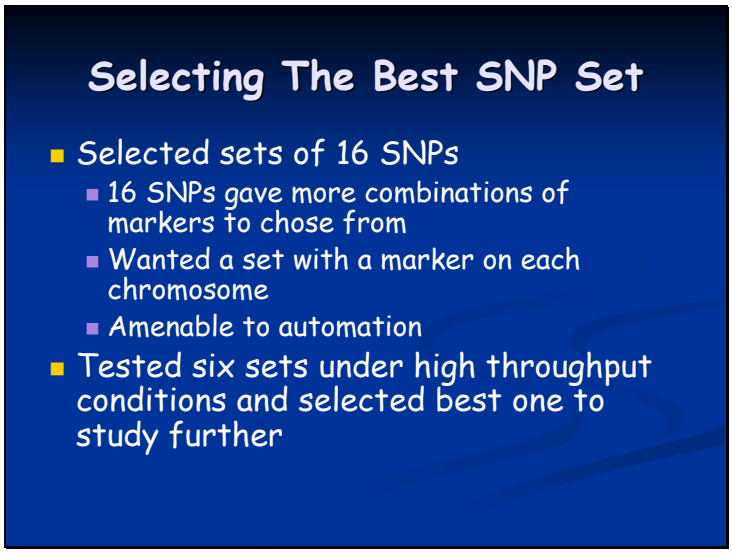
## Available Data for Analysis

- 383 diverse US and EU inbreds
- 491 SNP markers
  - good quality data under high throughput conditions
  - High polymorphism information content (PIC) in US and EU commercial germplasm
- Tested sub-sets of inbreds of different sizes to determine the minimum number of markers that could uniquely identify members of each sub-set

Slide 7




Slide 8



Slide 9

## Direct Comparison of SNPs with Isozymes

- 10 inbreds sampled and compared to data for 212 inbreds, some highly related
  - The same plants were sampled using
    - 15 isozymes (coleoptile tissue)
    - 16 SNPs (DNA extracted from leaves)
  - Replicate samples of between 15 and 143
- The sample profiles (including missing data, heterozgous and wrong calls for that sample) were compared to profiles for 212 inbreds
  - a 'resolution score' was calculated = 1/the number of matching profiles.
  - A score of 1 indicates complete resolution ie the only matching profile is to itself, and decreasing values indicate decreasing resolution power

Slide 10

## Results

- Missing data rates – higher for SNPs
  - 0.8% isozymes
  - 2% SNPs
- SNPs still had much higher resolution compared with isozymes

Slide 11

## Overall Resolution Scores

| Inbred | Number of samples | Overall resolution 15 isozymes | Overall resolution 16 SNPs |
|---|---|---|---|
| A | 145 | 0.05 | 0.94 |
| B | 20 | 0.05 | 0.91 |
| C | 21 | 0.08 | 1 |
| D | 16 | 0.07 | 0.94 |
| E | 23 | 0.05 | 1 |
| F | 20 | 0.07 | 1 |
| G | 16 | 0.03 | 1 |
| H | 15 | 0.03 | 1 |
| I | 48 | 0.17 | 0.98 |
| J | 48 | 0.17 | 0.98 |
| **Overall** | **387** | **0.06** | **0.96** |

A set of 16 SNPs has 16 X resolution of 15 isozymes

Slide 12

## Analysis of PVPd Inbreds with 16 SNPs

- 309 US Pioneer inbreds
  - 47292/47542 (99.9%) pairs could be resolved
- 192 European Pioneer inbreds
  - 18319/18336 (99.9%) pairs could be resolved
- Some missing data – with complete data the resolution could be higher

Slide 13

## Conclusions

- Genetic algorithms provide a powerful method for selecting markers that collectively provide high resolution power for variety identification
- A carefully selected set of SNPs will provide a much greater level of resolution than isozymes and can tolerate missing data due to sufficient redundancy
- 16 SNPs are extremely powerful at distinguishing among US and EU inbreds that are relevant to commercial germplasm today

Slide 14

## Appendix

Slide 15

## Examples of Resolution Scores - SNPs

| Inbred | Sample | Profile | Highest match | Resolution score |
|---|---|---|---|---|
| Inbred A | 1 | GCCTACCGGGATGGCG | Inbred A | 1 |
| | 2 | [A/G]CCTACCGGGATGG C[A/G] | Inbred A and 1 other inbred (sib) | 0.5 |
| | 3 | GCC[C/T]ACCGG[G/T]AT GG[C/T]G | Inbred A | 1 |
| | 4 | GCCTACCGGGANGGCG | Inbred A | 1 |
| | 5 | GCNNACNNGGANNNNG | Inbred A and 4 other related inbreds | 0.2 |

Slide 16

## Examples of Resolution Scores - Isozymes

| Inbred | Sample | Profile | Highest match | Resolution score |
|---|---|---|---|---|
| Inbred A | 1 | 2,9,4,6,6,16,12,12,4,4,3.8 ,5,4,6,4 | Inbred A and 17 other inbreds | 0.06 |
| | 2 | 2,9,4,6,6,16,12,12,4,4, 2/3.8,5,4,6,4 | Inbred A and 29 other inbreds | 0.03 |
| | 3 | 2,9,4,6,6,16,12,12,4,4,N, 5,4,6,4 | Inbred A and 30 other inbreds | 0.03 |

[End of document]