E

UPOV

BMT/12/15 Add.
ORIGINAL: English
DATE: May 26, 2010

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA PROFILING IN PARTICULAR

Twelfth Session
Ottawa, Canada, May 11 to 13, 2010

ADDENDUM

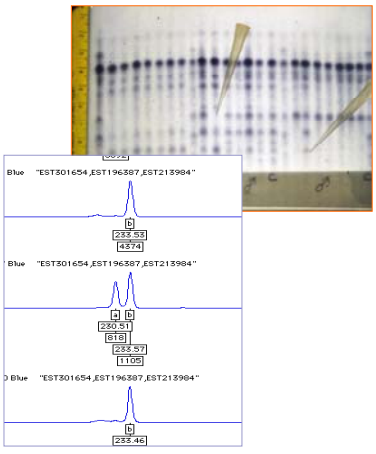
VARIETAL IDENTIFICATION IN MAIZE: ARE SIXTEEN SNP MARKERS SUFFICIENT?

Document prepared by experts from the United States of America

SNP Markers in Varietal Identification and Purity Analysis of Maize

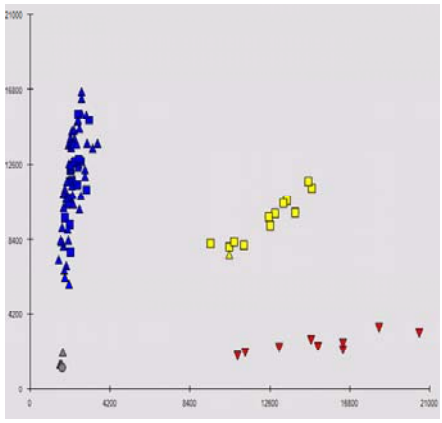
Current marker systems used in maize

- Isozymes
  - Low cost, fairly discriminative
  - Need well trained people
  - Supply issues for reagents
- SSRs
  - Highly discriminative
  - More expensive
  - Can be hard to replicate results
- Other technologies eg isoelectric focusing, AFLPs



Current marker systems used in maize - SNPs

- Allele itself is reported
  - COMMON LANGUAGE
  - Same allele with different systems
  - Compare results directly
- Very high throughput
- Low error rate
- Cost can be comparable with isozymes, and will likely be further reduced
- Many companies have moved to this system



How many markers?

- 10 SSRs - 60 genotypes of strawberry, including among sibling varieties
- 4 SSRs - 66 commercial apple varieties
- 16 SSRs - 548 accessions of rice
- 6 SSRs - 400 potato cultivars
- 22 SSRs used by GEVES in maize




How many SNP markers?

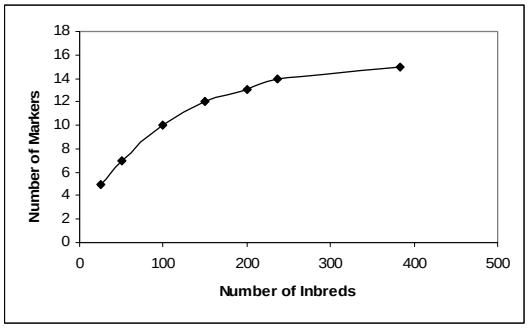
- Biallelic - not as polymorphic as SSRs
- “In theory, as few as 12 such markers can separate up to 4006 (=2^{12}) possible genotypes.” (Gale et al. 2005)
- 23 SNP loci equivalent power to 13 SSR loci in soybean (Yoon et al. 2007)
- 8 SNP loci uniquely identify 43 Japanese rice cultivars (Shirasawa et al. 2006)

Selecting the minimum number of SNPs to uniquely identify a large number of varieties

- Multivariate approach (Song et al. 1999)
- Integer linear approach (Gale et al. 2005)
- Our approach: Genetic algorithm
  - A marker is randomly placed into a set and then it is determined whether discrimination power has been improved, or not, when compared to the previous best combination of markers. The process is repeated thousands of times to find the minimum set

Selecting the best SNP set for variety identification in maize

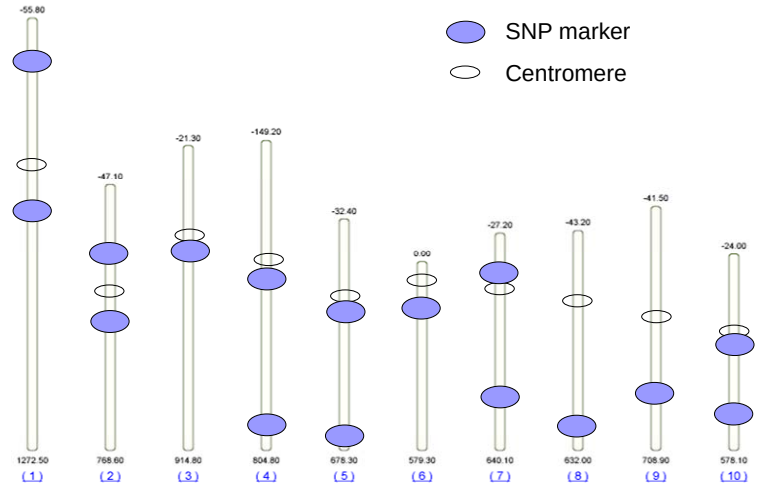
- Starting point:
  - 491 SNPs
  - 383 inbreds
- Selected the minimum number of SNPs that could discriminate EVERY inbred
- 15 SNP markers could discriminate among 383 inbreds



Selecting SNPs

- Selected multiple sets of 16 SNPs: 16 fits plate format and allow some redundancy
- Selected sets with each chromosome being sampled
- Tested sets for ability to perform well under high throughput conditions

Map locations 16 SNPs

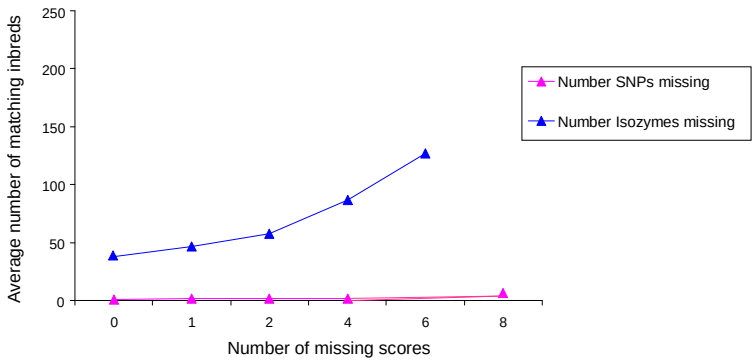


Comparison of isozymes and SNPs

- Used 46-96 seed samples for multiple inbreds and hybrids and assayed each individual seed with both marker systems in blind tests
  - Standard set of 15 isozymes
  - 16 SNPs
- Compared profiles to know profiles for hundreds of inbreds
- Overall, SNPs had 16 times the distinguishing power compared with isozymes
  - Where SNP profiles had on average 2-3 matching inbreds, isozymes had on average 40 matching inbreds
  - Matching SNP profiles had a high degree of pedigree relatedness
- SNPs did have higher missing data (2%) compared with isozymes (0.8%)

Examined Higher Levels Missing Data - simulated

8 inbreds compared to a reference set of 438 inbreds



- SNPs maintain power of discriminatory levels in the face of 50% missing data

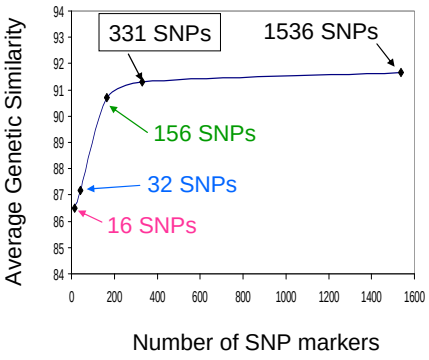
Effectiveness in different sets of inbreds

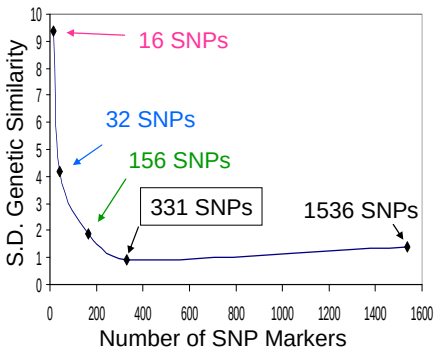
- Among 192 Pioneer European PVPd inbreds
  - 99.9% of pairs could be distinguished
- Among 58 inbreds bred by competitors and now publically available due to expired PVP protection
  - 99.5% pairs could be distinguished
  - The pairs that could not be distinguished all had similar pedigree backgrounds

Using more than 16 markers – what are the gains?

- For 248 inbreds
  - 16 SNP markers, 22 / 61256 pairs of inbreds could not be distinguished
    - 99.96% pairs distinguished.
  - 42 SNP markers, 2 / 61256 pairs could not be distinguished
    - 99.99% pairs distinguished
  - 165 SNP markers, 100% of the inbreds could be distinguished.

Can small numbers of markers accurately determine genetic similarities?





Overview

- 16 well selected SNPs can distinguish among >99% inbreds
- 16 SNPs robust in the face of 50% missing data
- Increasing the number of SNPs may have minimal gains
- 8 SNPs may be sufficient
- Genetic distance analysis needs >150 SNPs

Selecting a common industry set of SNPs for variety identification and purity analysis

- Build on existing ASTA/SEPROMA (UFS) collaboration
  - To find a set of standard SNPs for EDV using 50K chip
- Screen a sub-set (384?) of informative markers on large array inbreds (ASTA, UFS, all public, proprietary, all PVP?)
- Find a sub-set that most efficiently describes all material

[End of document]