E

BMT-TWA/Maize/2/7-c
ORIGINAL: English
DATE: November 26, 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

EDV IN CORN:

IDENTIFYING ESSENTIALLY DERIVED VARIETIES WITH MOLECULAR MARKERS

Document prepared by experts from the International Seed Federation (ISF)

Slide 1

Identifying Essentially Derived Varieties with Molecular Markers

M. Bohn, M. Frisch,
M. Heckenberger, D. Klein,
H.P. Maurer, A.E. Melchinger

Slide 2

Outline

- Theory
- Computer Simulation
- Conclusions for EDV identification
- Verification Studies

Heckenberger et al. 2005. TAG 111:598

Slide 3

Questions – Simulation Study

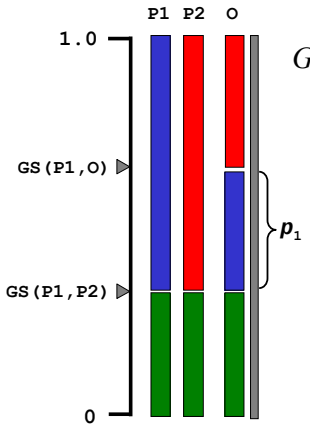
- How do …
  - chromosome number and length,
  - marker genome coverage and distribution,
  - degree of polymorphism between parental inbreds

  influence the distribution of GS estimates between parental inbreds and their offspring derived from …
  - F2 populations,
  - BC populations?
- How do these factors influence the power of molecular markers to discriminate …
  - F2 vs. BC1, BC1 vs. BC2 derived lines?

GS = genetic similarity BC = back cross

Slide 4

Theory 1



$$GS_{(P1,O)} = GS_{(P1,P2)} + p_1(1 - GS_{(P1,P2)}) \quad \text{Eq.[1]}$$

$$p_1 = \frac{GS_{(P1,O)} - GS_{(P1,P2)}}{(1 - GS_{(P1,P2)})} \quad \text{Eq.[2]}$$

If GS(P1,P2) = 0

then GS(P1,O) = p_1

p_1 proportion of the genome transmitted from P1 to O

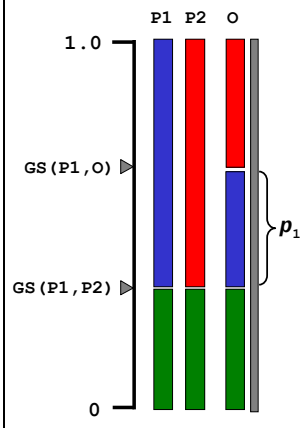
$$\sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(GS - GS_{true})^2} \quad \text{Eq.[3]}$$

Accuracy of estimated GS values

N = sample size.

Slide 5

Theory 2



Effective Number of Markers

$$m_e = m \times (1 - GS_{(P1,P2)})$$ Eq. [4]

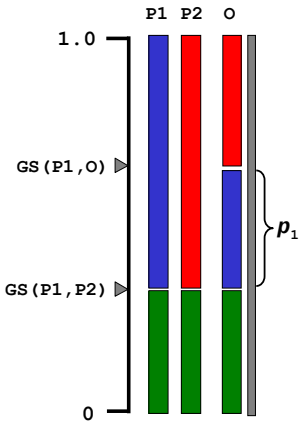
Effective Marker Distance

$$d_e = \bar{d} \times \frac{1}{1 - GS_{(P1,P2)}}$$ Eq. [5]

$$m_e = \frac{l_t}{d_e}$$ Eq. [6]

Slide 6

Theory 3: Mean and Variance of GS(P1,O)



$$\mu_{GS_{(P1,O)}} = \mu_{GS_{(P1,P2)}} + \mu_{p_1}(1 - \mu_{GS_{(P1,P2)}})$$ Eq. [7]

$$\sigma^2_{GS_{(P1,O)}} = (1-\mu_{p_1})^2 \sigma^2_{GS_{(P1,P2)}} + (1-\mu_{GS_{(P1,P2)}})^2 \sigma^2_{p_1} + \sigma^2_{GS_{(P1,P2)}} \sigma^2_{p_1}$$ Eq. [8]

- Distribution of GS(P1,O) depends on
  - Pedigree
  - Number and length of chr.

Knowledge of the distribution of GS(P1,O) is a key prerequisite to develop a statistical test for identifying EDVs. However, an analytical description of this distribution is not available.

Slide 7

Computer Simulations - Scenarios

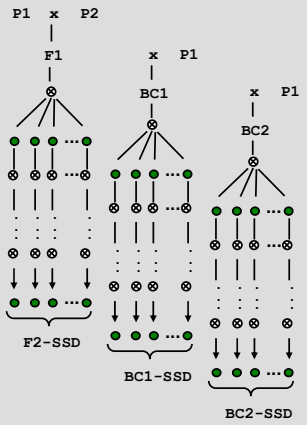
- Chromosome Number

  5 – 10 – 17 – 19 – 20 – 21
- Chromosome Length (cM)

  20 – 40 – 80 – 160 – 320
- Average Marker Distance (cM)

  4 – 8 – 16 – 32 – 64 – 128 – 256
- GS(P1,P2) = 0, 0.25, 0.50, 0.75
- Marker Distribution

  Random vs. uniform

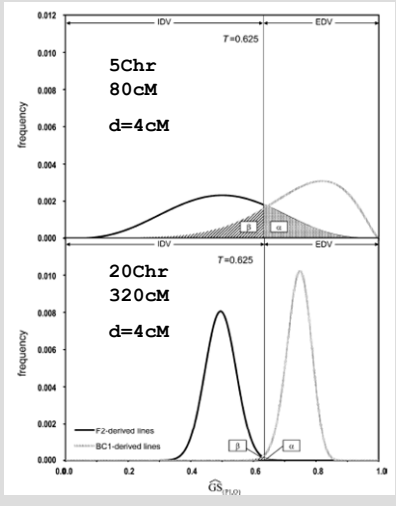


Slide 8

Var[GS(P1,O)] and Statistical test

H_0: Line O is an F2-derived lines.

H_A: Line O is more closely related to P1 than expected for F2-derived lines.



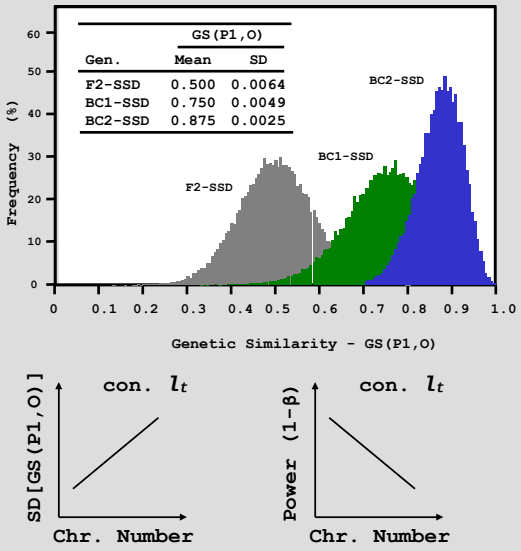
Slide 9

Results – "true" GS Across the Whole Genome

Assumptions:

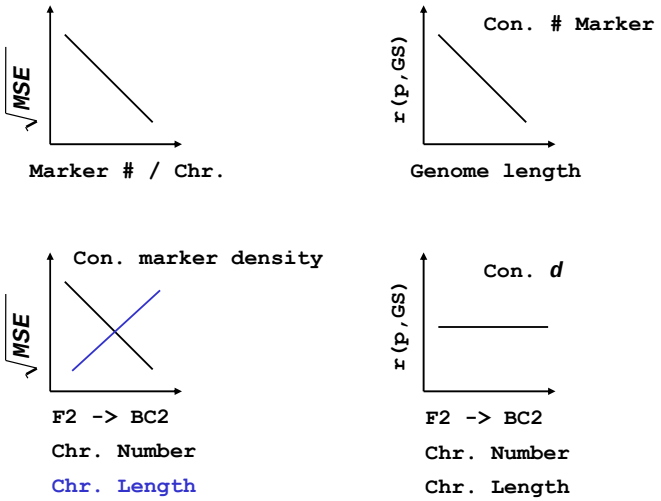
1. Many markers
2. Uniform distribution
3. GS(P1,P2) = 0

$$\sigma^2_{GS_{(P1,O)}} = \sigma^2_{p_1}$$



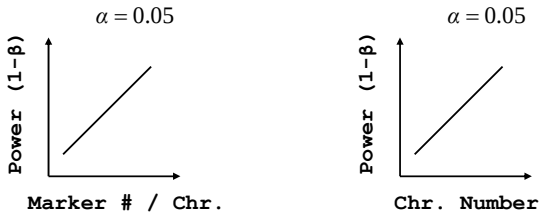
Slide 10

Results – Marker Estimates of GS



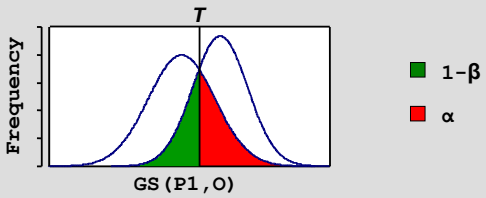
Slide 11

Results – Thresholds *T* for Estimates of GS



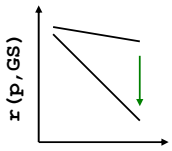
Decrease for d < 16cM marginal for all simulated genomes.

Crop specific thresholds!



Slide 12

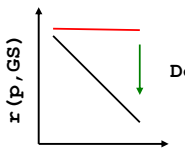
Results – Marker Coverage and Polymorphism



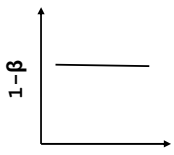
Decreasing marker density
Decreasing chr. length

Uniform - random

1-β: uniform > random



Decr. marker dens.

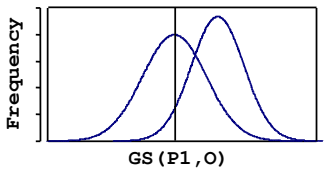


Adequate m_e and d_e

d_e < 16 cM

Slide 13

Summary



- SD and distribution overlaps are smaller with:
  - Increasing chromosome # and length.
  - Increasing marker density.
  - Uniform vs. random distribution of marker.
- GS(P1,P2) influences power of EDV identification only if *me* is small.

Slide 14

Guidelines for EDV Identification Procedures – Parental Inbreds of putative EDV are known

Step 1

Estimate GS(P1,P2)and GS(P1,O)with a sufficient number of markers!

Step 2

Estimate p!

Step 3

Generate a large number of virtual F2 and BC1 derived lines to determine a critical threshold T based on a given α or 1-β.

Use for simulations polymorphic marker used for estimating GS(P1,O).

Step 4

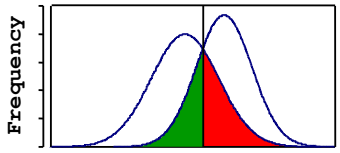
Perform statistical test for EDV identification

Slide 15

Guidelines for EDV Identification Procedures

H_0: Line O is an F2-derived lines.

H_A: Line O is more closely related to P1 than expected for F2-derived lines.



GS(P1,O)

"Maize" > 200 polym. Marker		
	T = 0.75	
α	0.001	0.05
1-β	0.500	0.91
		T = 0.64

Slide 16

Guidelines for EDV Identification Procedures – One Parental Inbred of putative EDV is unknown

Adjustment according to GS(P1,P2) is in general not possible, because P1 & P2 are often not available.

Solution:

Instead of GS(P1,P2) use the mean GS of unrelated pairs.

Consequence:

- Too conservative for parental lines more similar than expected for unrelated lines
- Too liberal for parental lines more distant than expected for unrelated lines.

Slide 17

Summary and Conclusions – "Maize"

- Type I and II error rates were substantially different for material groups for fixed EDV thresholds.
- Joint threshold for intra-pool and inter-pool crosses increases risk to produce EDV from intra-pool cross.
- Thresholds must be pool specific!
- Different thresholds for intra-pool and inter-pool crosses are necessary!
- EDV thresholds must account for lab errors and intra-varietal variation.

Slide 18

Summary and Conclusions – "Maize"

- Standard errors of GS estimates were of considerable size.
- 100 SSRs seem to be the lower limit for EDV identification, because high SE for GS increase Type I and Type II error rates.
- We propose a two step procedure:

  1. Step: Use of 100 evenly distributed SSR markers.

  2. step: If relationship between initial variety (IV) and potential EDV is in doubt a second set of SSRs is used.

- Alternatives:

  Use of new high throughput marker systems (SNPs), haplotypes, and microarrays.

Slide 19

Triplet Study: Materials and Methods

Lines: 220

89 European Flints
74 European Dents
14 U.S. Dents
43 Introgression

Triplets: 163

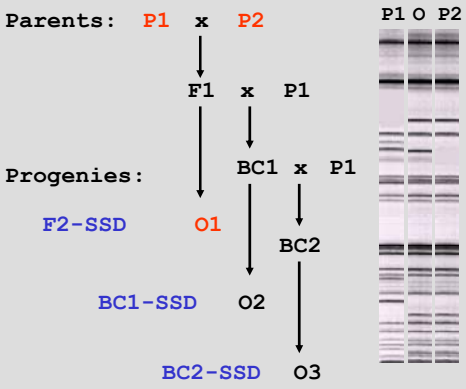
118 Intra-pool
45 Inter-pool

83% F2 derived
17% BC1, BC2 derived

SSR: 100

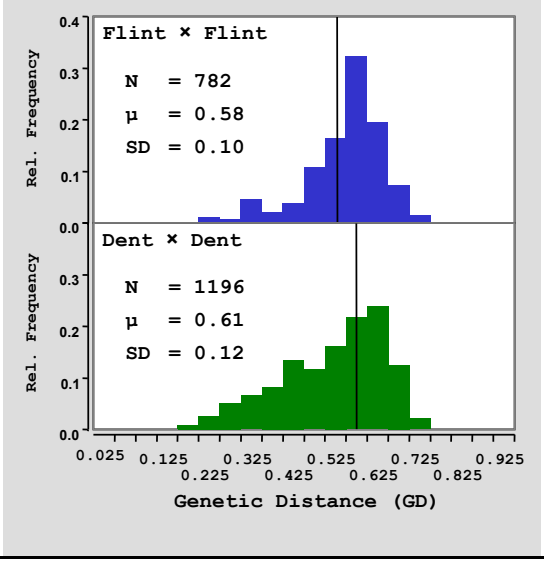
GD = Rogers' Distance

Triplet structure



Slide 20

Triplet Study: GD of Unrelated Lines (*f* = 0)



Slide 21

Estimation: Variance of GD(P1,O)

	F_2		BC_1	
Par.	Obs.	Sim.	Obs.	Sim.
μ_p	0.4893	0.5000	0.6567	0.7500
σ^2_p	0.0107	0.0102	0.0088	0.0076
$\mu_{GD_{(P1,P2)}}$	0.6314		0.7277	
$\sigma^2_{GD_{(P1,P2)}}$	0.0024		0.0034	
$\hat{\mu}_{GD_{(P1,O)}}$	0.3095	0.3157	0.2465	0.1819
$\sigma^2_{GD_{(P1,O)}}$	0.0051	0.0063	0.0034	0.0043

$$\sigma^2_{GD_{(P1,O)}} = (1-\mu_{p_1})^2 \sigma^2_{GD_{(P1,P2)}} + \mu^2_{GD_{(P1,P2)}} \sigma^2_{p_1} + \sigma^2_{GD_{(P1,P2)}} \sigma^2_{p_1}$$

Obs. = observation Sim. = simulation

Slide 22

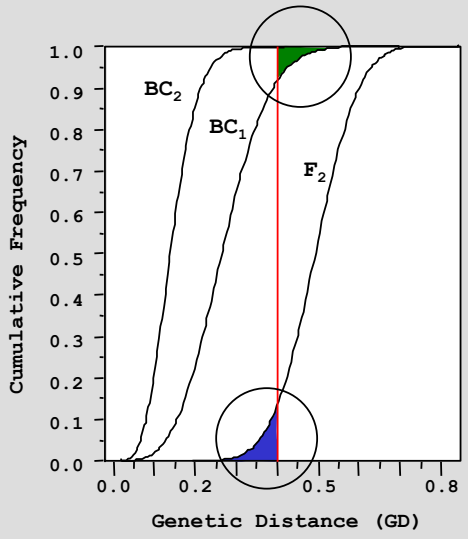
Estimation: Variance of GD(P1,O)

Percent of $\sigma^2_{GD_{(P1,O)}}$ explained by:

Par.	F_2	BC_1
	---- % ----	
$\sigma^2_{p_1}$	65	94
$\sigma^2_{GD_{(P1,P2)}}$	34	5
$\sigma^2_{GD_{(P1,P2)}} \sigma^2_{p_1}$	< 1	< 1

Slide 23

EDV Threshold Scenarios



Criteria for choosing EDV threshold (T) are:

- Type I Error (α)
- Type II Error (β)

Slide 24

EDV Threshold Scenarios: F_2 *vs.* BC_1

	Flint		Dent	
	obs.	sim.	obs.	sim.
α = 0.05				
T	0.21	0.17	0.24	0.18
(1-β)	77%	72%	63%	71%
α = β				
T	0.24	0.20	0.28	0.22
(1-β)	12%	12%	17%	14%

Slide 25

EDV Threshold Scenarios: F_2 *vs.* BC_1

	Flint		Dent	
	obs.	sim.	obs.	sim.
T = 0.25				
α	0.18	0.30	0.07	0.26
(1-β)	92%	98%	68%	95%
T = 0.20				
α	0.03	0.11	0.01	0.09
(1-β)	72%	87%	39%	81%

Slide 26

Summary and Conclusions

- ASSINSEL proposed EDV threshold (T = 0.20) results in acceptable Type I and Type II error rates for the Flint pool, but power of EDV identification is low in the Dent pool.
- The use of crop- and gene pool-specific EDV thresholds based on a fixed α-level or α = β is recommended.

Slide 27

Acknowledgements

Gesellschaft zur Foerderung der privaten Pflanzenzuechtung (GFP), Germany

European Union

Iowa State University

Commercial Breeding Companies

[End of document]