

Family-Based Association Analyses in Plants

Clay Sneller, Ohio State University

Kevin Smith, Jon Massman, University of Minnesota

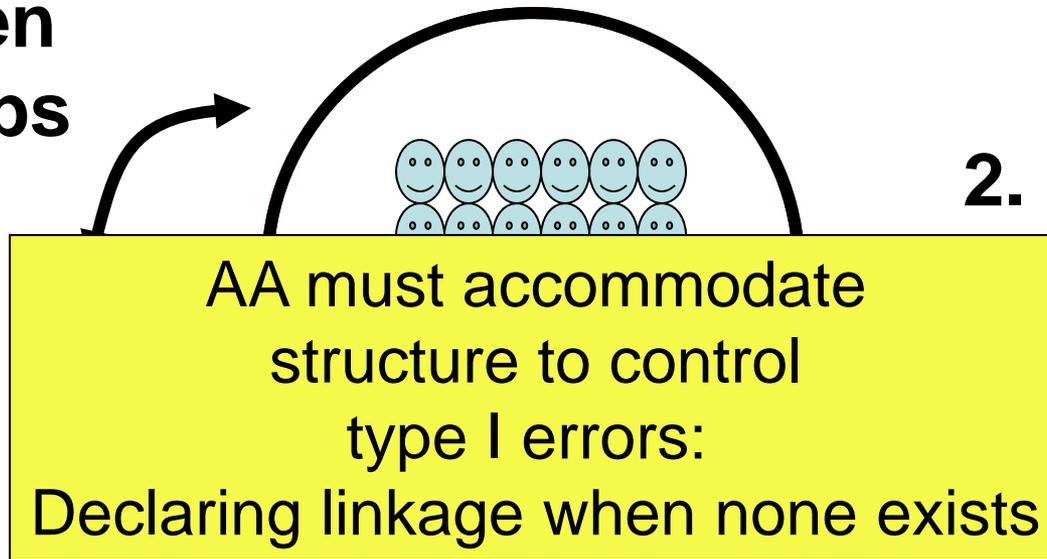


Association Analyses (AA)

- Associate – *to connect in the mind or imagination*
- Statistically associate marker and phenotypic data
- Detect a physical linkage of marker and trait loci (QTL)
- Normally used in complex populations: many parents
- AA must deal with population structure

Population Structure: Unequal relationship between individuals

1. Between Subgroups



2. Within Subgroups

Population- vs Family-Based AA

	Population	Family
Estimation association parameter	Over entire population	Within lineages, between relatives then compiled
Population structure	Estimated & modeled	Negated by sampling
Inference of linkage	Implied by significance	Required for significance

Population-Based AA

- Commonly used in plants
- Applicable to many population types
- Common statistics
 - Main effect of marker: means comparison
 - Covariance for effect of subgroups
 - TASSLE+STRUCTURE, unified mixed-model of Yu et al. 2006

Mean Freq "1" Freq "0"

0 1 1 0 1 0 1 0 1 0

~~50~~

~~0.5~~

~~0.5~~

1 0 1 0 1 0 1 0 1 0

"75"

0.1

0.9

100
"75"

0.9
0.9

0.1
0.1

$Y_i = u + g_i + \text{other effects}$

$Y_i = u + \text{Cov} + g_i + \dots$

" $\overline{X}_1 > \overline{X}_0$ "
" $\overline{X}_1 = \overline{X}_0$ "

Mean Freq "1" Freq "0"

1Q
0q

0 1 0 1 0
1 0 1 0 1

75 0.5 0.5

$$\bar{X}_1 > \bar{X}_0$$

1q
0Q

1 0 1 0 1
0 1 0 1 0

75 0.5 0.5

$$\bar{X}_1 < \bar{X}_0$$

$$Y_i = u + \text{Cov} + g_i + \dots$$

$$\bar{X}_1 = \bar{X}_2$$

Family-Based AA

- As individuals become more related, they become more similar
- Estimate association parameter within lineages
- Compile and test for significance



Mean Freq "1" Freq "0"

1Q

0 1 0 1 0

75

0.5

0.5

0q

1 0 1 0 1

$$\bar{X}_1 > \bar{X}_0$$

1q

1 0 1 0 1

75

0.5

0.5

0Q

0 1 0 1 0

$$\bar{X}_1 < \bar{X}_0$$

“Sib” Pair Regression

Haseman
& Elston, 1972



Behavior

Hair
Pigment

Marker

Sweet

2

A



Sassy

2

A



Steady

7

B

Regress Phenotypic Difference² on Proportion of IBD alleles at Marker

Pair

$$P^D = (X_i - X_j)^2$$

Mark 1 IBD

1



0

1

Shared allele

2



25

0

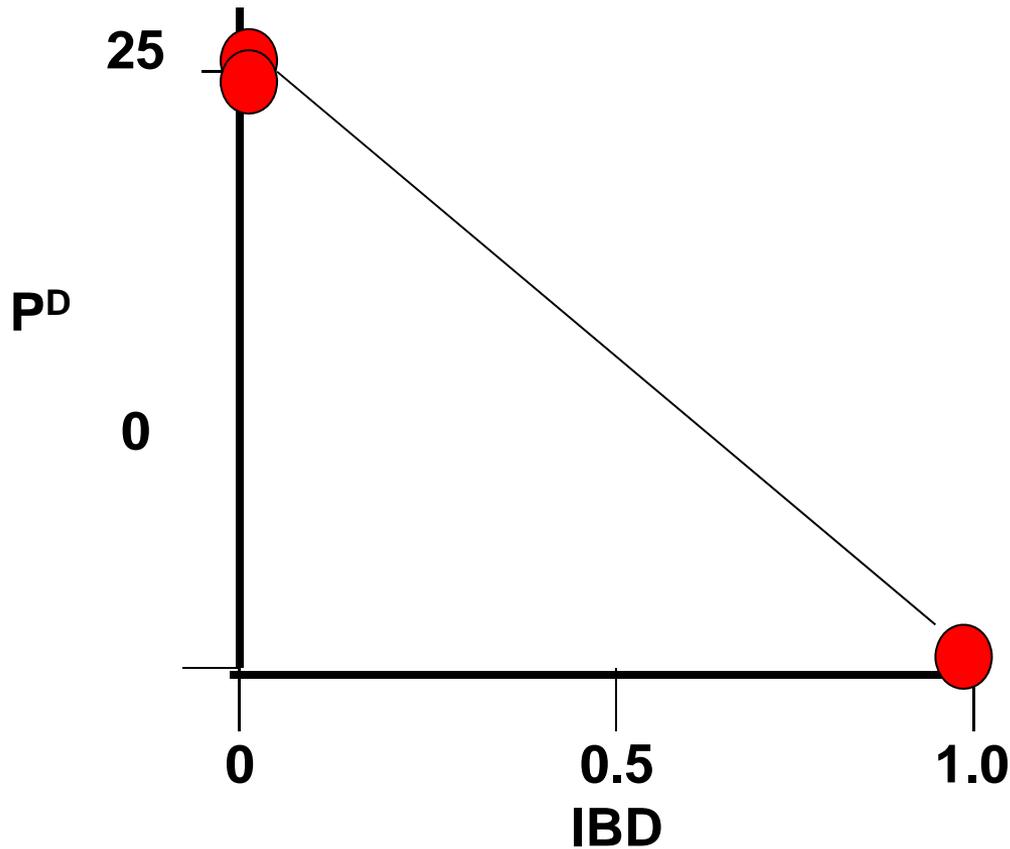
No shared allele

3



25

Regress P^D on IBD



$$B = -25$$

$$B = -2(1-2c)^2 \sigma_a^2$$

$$\sigma_a^2 = 12.5 \quad \text{if } c=0$$

Multiple Families: Lineages

Family	n	No. Pairs	Freq 0	Freq 1	Freq 2
Snellers	3	3	0.66	0.33	0
Vassilyev	69	2346	0.50	0.50	0
Daad	86	3655	0.35	0.55	0.10
Hatfields	35	595	0.90	0.10	0
McCoys	35	595	0.90	0.10	0

7194

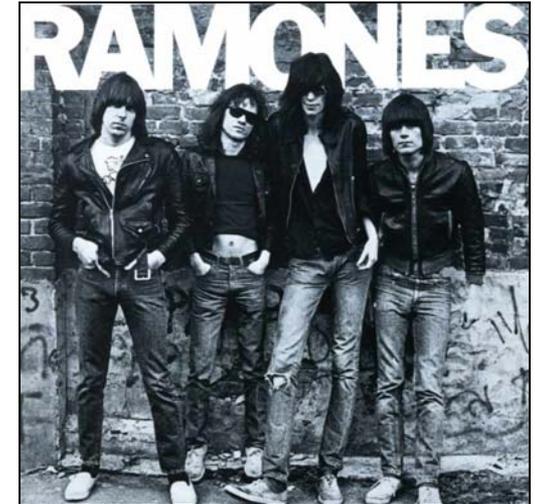
Human Genetics

FBAA



PBAA

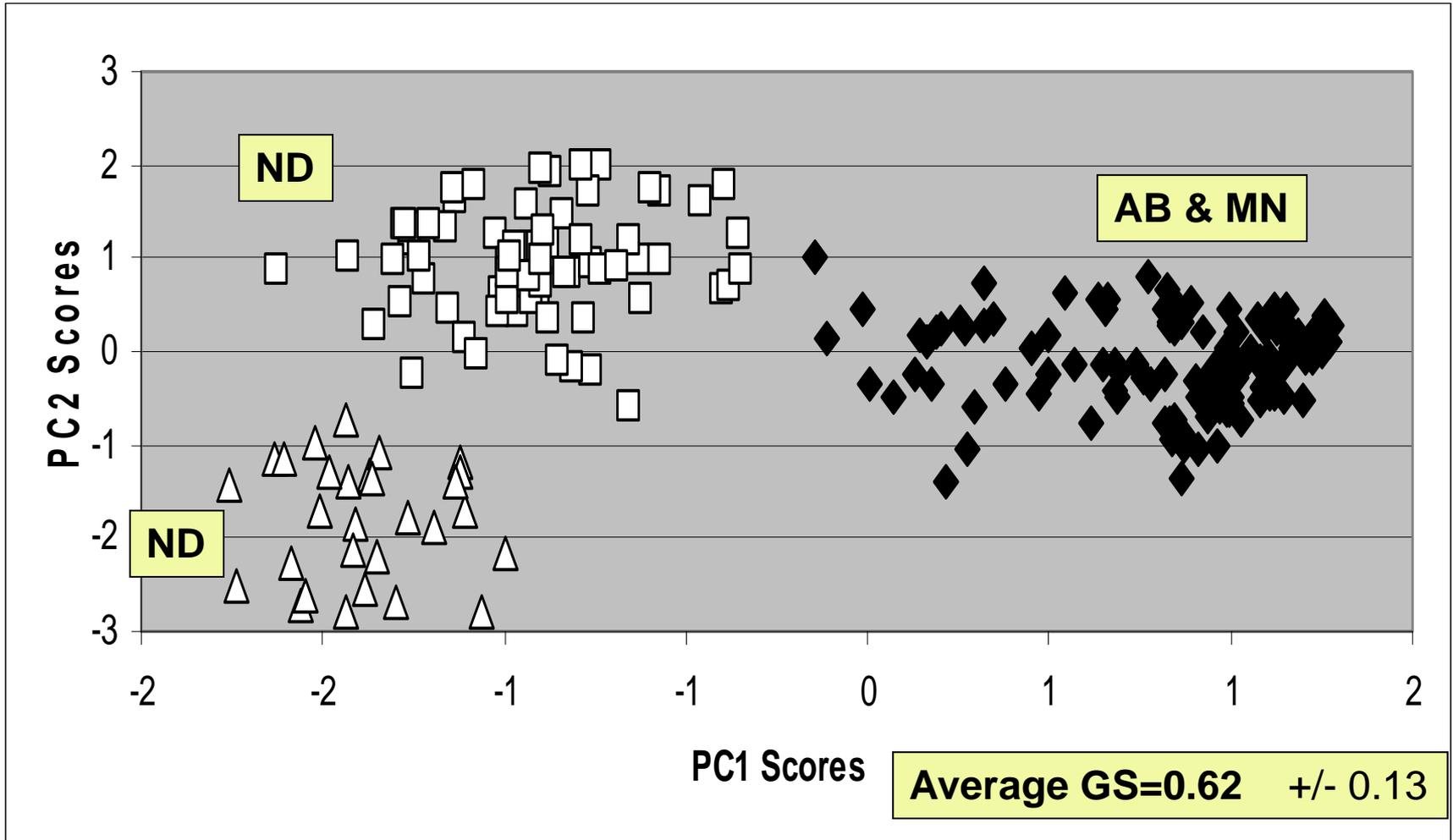
- Family data ~~X~~ is hard to collect, ~~X~~ verify parentage
- Studied ~~X~~ populations are not highly ~~X~~ structured - random
- Careful ~~X~~ *apriori* sampling to minimize ~~X~~ effect of structure
- VERY large population size



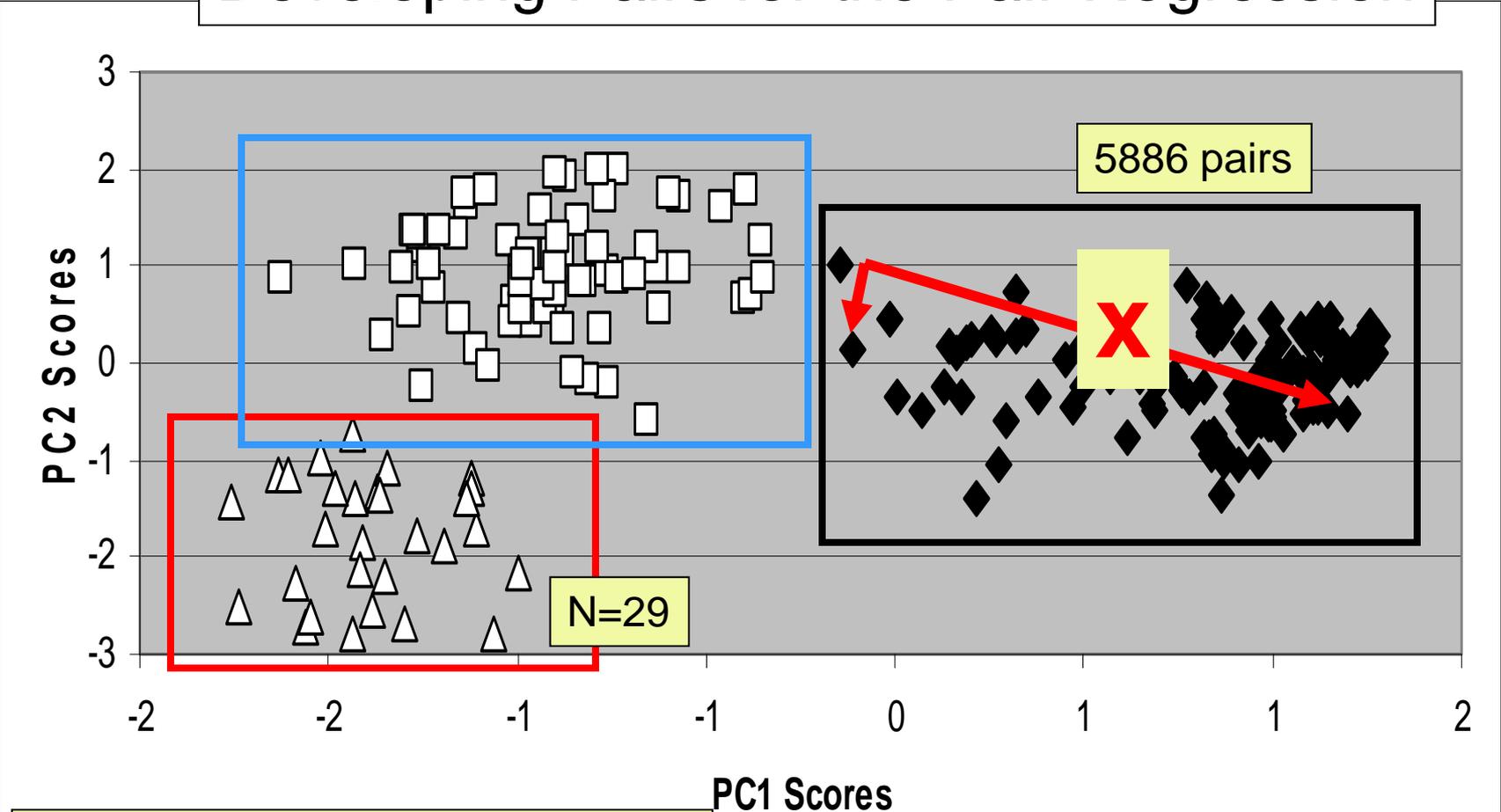
FBAA Example: 206 Barley Lines, Barley CAP

- Derived from 65 biparental crosses
- Average 3.1 progeny per cross
- DON data from three environments
 - $h^2 = 0.52$
- Genotyped with 2924 SNP markers
BOPA_C(1)
- Analysis used 676 SNPs ($PIC > 0.18$)

PCA of Genetic Similarity Matrix



Developing Pairs for the Pair-Regression



3 Lineages
Used Lineages with PIC >0.18
Used pairs with GS >0.75

Average GS=0.62 +/- 0.13

Models

TASSLE
(Q+K)

$$Y_i = u + \text{Cov} + g_i + \text{polygene}$$

STRUCTURE

Intercept

Genetic similarity

IBD Proportion

Pair Regression

$$P^D_i = u + B_1 S_i + B_2 I_i$$

Covariance of individuals
within a lineage

10 22
Mark 27

Pair Reg		Tassle	
(VAR)	PR	T (LOD)	
	*****	*** 7.0	

50	*****		

	*****	*	
	*****	* 2.8	
	*****	*	
	*****	*	

Prob < .00001

9 46
Mark 49

4 55
Mark 56

(VAR)	PR	T (LOD)	
	***	**	

	***	*	
105	*****	***** 10.2	
	*	**	
		**	
	*	**	
		**	
		**	

(VAR)	PR	T (LOD)	
	*	**	
	*****	**	
190	*****	***** 10.2	
	*****	**	

Chromosome 4H

Tassle vs Pair-Regression

of QTL

Tassle & Pair-Regression	16
Tassle Only	1
Pair Regression Only	4

Population well suited for both
Clear lineages
3 lineages

Xsm	cM	Var	PR	T (LOD)
-----	----	-----	----	---------

7H	161	43	*****	* 2.6
----	-----	----	-------	-------

6H	13	58	****	* 2.7
	13		*****	
	17			

5H	87	26	*****	
	89		*****	
	94			
	94		*****	
	95		*	

5H	173			** 4.0
----	-----	--	--	--------

Xsm	cM	Var	PR	T (LOD)
-----	----	-----	----	---------

3H	145	46	*****	** 3.1
	148		*****	
	150		*****	
	155		*****	

1H	51	47	***	
	53			
	56		*****	

FBAA is Well Suited for Plant Breeding Populations

- Populations are EXTREMELY relevant
- Many lines are phenotyped annually
- Multiple large lineages are present
 - Full Sibs
 - Half Sibs
 - Other degrees of relationship, lineages

2009 YR1 Phenotyping: FHB Index

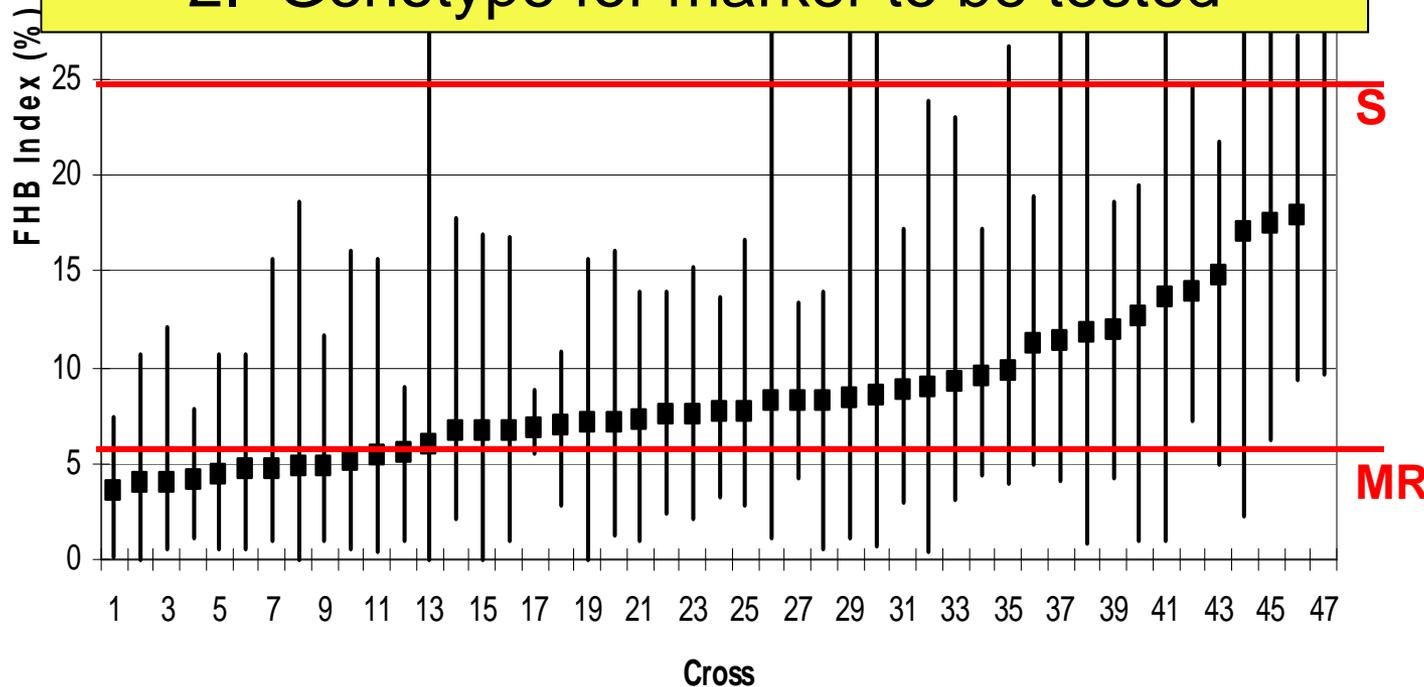
FBAA to evaluate a marker
in a breeding population:

1. Build lineages based on pedigree: FS, HS
2. Genotype for marker to be tested

570 Lines
47 crosses
12 lines/cross

Many Xs seg

4597 Full-Sib
pairs



Other Types of FBAA

- Quantitative Inbred Pedigree Disequilibrium Test
- Two-level Haseman-Elston Regression

Quick Takes on FBAA

- 1 study, much more needed to see applications: simulations
- Well suited for breeding populations
- May circumvent some issues inherent to population-based AA
- Can handle rare alleles
- QTL validation & evaluation in breeding populations
- Stability of QTL effects over lineages

Thanks

- Kevin Smith, Jon Massman
- Barley CAP folks
- Dr Elston
- Diane Mather



Types of Plant Populations and Association Analyses

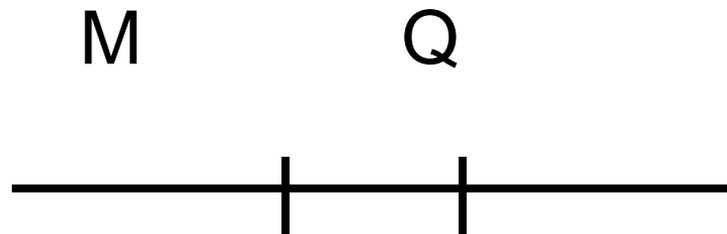
	Diverse	Breeding	Biparental
Number of Parents	Many Ancestors	Many Elite	2 Selected
Amount of Structure	Lots Evolution	Lots Breeding	V Little
Relevance to Breeding	Some	Lots	Variable
Type of Analysis	Population-Based AA		CIM
		Family-Based AA	

Association Analysis:

Associate: to connect in the mind or imagination

Link: to connect, to tie or bind

- Associate variation of marker genotypes with variation of phenotypes
- Imply linkage of marker locus and QTL



$$P1 = \frac{1 Q}{1 Q}$$

$$P2 = \frac{0 q}{0 q}$$



1. Pop0 and Pop1 likely equivalent
If large

2. Two alleles

3. High LD

4. Significance requires linkage

0	1	0	1	0	1	0
1	0	1	0	1	0	1
0	1	0	1	0	1	0

$$\overline{X}_0$$

$$\overline{X}_1$$

$$Y_i = u + g_i + \text{other effects}$$

0	1	0	1	0	1	0
1	0	1	0	1	0	1
0	1	0	1	0	1	0
0	1	0	1	0	1	0
1	0	1	0	1	0	1
0	1	0	1	0	1	0

Population,
Genotyped 1 marker
Phenotyped

Test Association:
Parameters are means

$$H_0: \overline{X}_0 = \overline{X}_1$$

$$Y_i = u + g_i + \text{other effects}$$

