USDA NIFA United States Department of Agriculture National Institute of Food and Agriculture

Solanaceae Coordinated Agricultural Project

THE OHIO STATE UNIVERSITY OARDC

Genomic Selection in Tomato Breeding

SolCAP Workshop, Tomato Breeders Round-Table , Ithaca, NY

David M. Francis, S.C. Sim, Heather Merk The Ohio State University
Allen Van Deynze, U.C. Davis
C. Robin Buell, J. Hamilton, D. Douches, Michigan State University
Walter De Jong, Lucas Mueller, Cornell University

illumina

INRA **Mathilde Causse**

Martin Ganal TraitGenetics

Acknowledgments:

Discussion
Dr. Jean-Luc Jannink, USDA/ARS, Ithaca, NY
Dr. Clay Sneller, Dept. Hort. and Crop Science, OSU

Resources developed by
Dr. Ben Hayes, Dept. of Primary Industries, Victoria, Australia
http://www.ans.iastate.edu/section/abg/shortcourse/notes.pdf

Dr. Ed Buckler and colleagues,
http://www.maizegenetics.net/

Dr. Gustavo de los Campos and colleagues,
http://genomics.cimmyt.org/

Overview

(1) Association Analysis

Identifying significant marker-trait linkages in complex populations

(2) Genomic Selection

Predicting breeding value of an individual based on kinship and genotype

(3) Preparing Data

(4) Resources

(5) Practical Examples

At the end of this module you will be able to:

Describe Association Analysis and Genome Wide Selection (GWS)

Define and estimate a Breeding Value

Define a multiple trait index

Prepare data for AA and GWS

Know how to access demonstrations and practical exercises.

Definitions

Association Analysis

- Mapping in unstructured populations

Marker Assisted Selection (MAS)

- Selection based on Marker-QTL linkage
  - Direct selection
  - Selection for coupling-phase recombination
  - Background genome selection for accelerated BC
  - Selection for multiple QTL, etc…

Genomic Selection (GWS)

- Selection based on breeding value
  - Random effects models and BLUPs
  - Estimate breeding value for markers and individuals

Association Analysis

Proposed as a way to overcome limitations of working with bi-parental populations for QTL-based discovery and subsequent MAS

In complex populations the magnitude of QTL effects tend to be small

Relevance of the complex population to applied goals remains an issue (e.g. inbred lines vs hybrids)

Association Analysis

Data

Vector of trait values from phenotypic evaluation of a large complex population (best if these are BLUPs)

Matrix of Markers

Matrix of population structure (STRUCTURE or PCA)

Kinship matrix

Association Analysis

```
%macro Mol(mark);
proc mixed covtest data = three;
class &mark gen;
model T1 = Q1 Q2 Q3 &mark / solution;
random gen / type = lin(1),data=KIN;
%mend;

        %Mol(SL144);
        %Mol(CT10737I);
        %Mol(CT20244I);
        %Mol(SL10525);
        %Mol(SL10526);
         etc…
```

Trait BLUP

Q Matrix (Q-1) from STRUCTURE

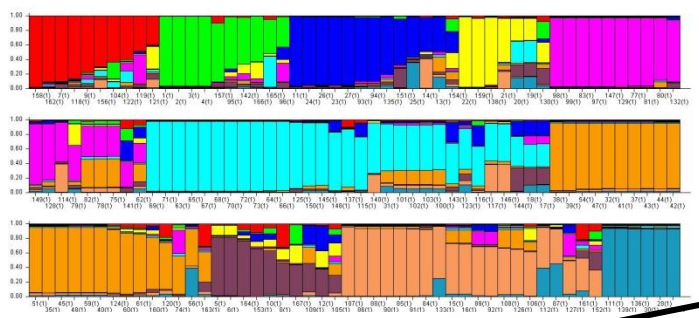
Marker

Kinship matrix (diagonal)

Models can estimate the contribution of STRUCTURE and Kinship to the trait and Marker-Trait linkage…

F-test for significance = $N(1-2r)^2g^2$

N = population size; r = recombination distance (marker to QTL; g^2 = proportion of variance explained by QTL



Chrom 10

Chrom 2

	L K	1/2	a l c	O v	Shelf -life	shape index
	✓		X	✓		
Q =1					180	0.85
					165	1.26
					160	1.27
					175	1.10
					155	0.87
					170	1.24
					135	1.26
					175	0.91
Q = 2					28	0.88
					21	1.18
					14	1.2
					21	0.85
					165	1.22
					175	1.24
					175	0.91
					180	0.87

Within the context of breeding programs, success with association analysis and therefore subsequent MAS will depend on:

(1) population structure;
(2) segregation of the trait within sub-populations;
(3) allelic diversity for the trait of interest;
(4) size of the population;
(5) size of the sub-populations within the larger population;
(6) the magnitude of the QTL (proportion of variance explained).

Recommendation:
(1) Skip association mapping in germplasm collections and focus on Nested Associated Mapping (NAM) style populations A x B; A x C; A x D; individual breeding programs, etc…
(2) Use large populations

Selecting based on molecular markers

Marker Assisted Selection - a subset of statistically significant marker–trait associations are discovered, validated and used for selection

Single markers linked to QTL

Haplotypes linked to QTL

MAS – based on marker –trait linkage

Genome Wide Selection - prediction of performance without evidence of statistically significant association.

Single markers

Haplotypes

GWS – based on sum of breeding values estimated for all markers

Genomic selection (GS)

Selection decisions based on genomic breeding values estimated as the sum of the effects of markers across the genome (Contrast to MAS in which only markers positively associated with trait are used).

Breeding values are derived from Best Linear Unbiased Predictors (BLUPs) as the sum of BLUPs for all markers.

Can estimate the breeding value of an individual, even when there are no observations (e.g. Dairy Sire example).

Genomic Selection

Breeding Value: The part of an individual's phenotypic value that is due to additive genetic effects. The value of an individual as a parent.

Assign a breeding value to each marker, regardless of significance…

GEBV = $\sum_{i}^{n} \mathbf{X}_i \, \mathbf{g}_i$ Genomic Estimated Breeding Value is the sum of all marker effects for an individual

Conceptual change:

Think of the value of a line based on its potential contribution to the next cycle of breeding vs its performance (Breeding vs Seeds/Commercial)

Animal Agriculture: Dairy farms purchase sperm based on its breeding value not performance. In contrast seed is purchased based on performance.

Breeding progress is based on gain under selection.

Implications

Significance of Marker-Trait (QTL) association (linkage) is less important than the estimated breeding value

We need to start thinking about Marker-QTL linkages as random effects

effects (markers) > than phenotypic observations

effects are estimated as BLUPs

Estimates of breeding value are strengthened by data from relatives, therefore pedigrees, kinship matrices, etc… improve estimates of breeding values.

Data:

Vector (or matrix) of trait-value (best if phenotypes are BLUPs)
Matrix of kinship (pedigree or marker-based) n x n
Matrix of Markers (n x k)

SNP scoring:

markers are scored 0 or 1; heterozygotes would be 0.5; could also be number of "common" alleles (0 = homozygous for rare allele; 1 = hetero; 2 = homozygous for common allele)

Approaches:

Step-wise regression

$\gamma = \beta_o + \beta_1 X_1 + \in$; $\gamma = \beta_o + \beta_2 X_2 + \in$; ***etc…***

Multiple linear regression

$\gamma = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots \in$

Multiple linear regression with correction/penalty

Ridge Regression
LASSO
Bayesian (various e.g. Bayesian-LASSO, etc…)

Comparing Stepwise with Multiple Regression (statistically naïve thought experiment)

fit1 = lmer(BL_L~(1|M1))
ranef(fit1)

.

.

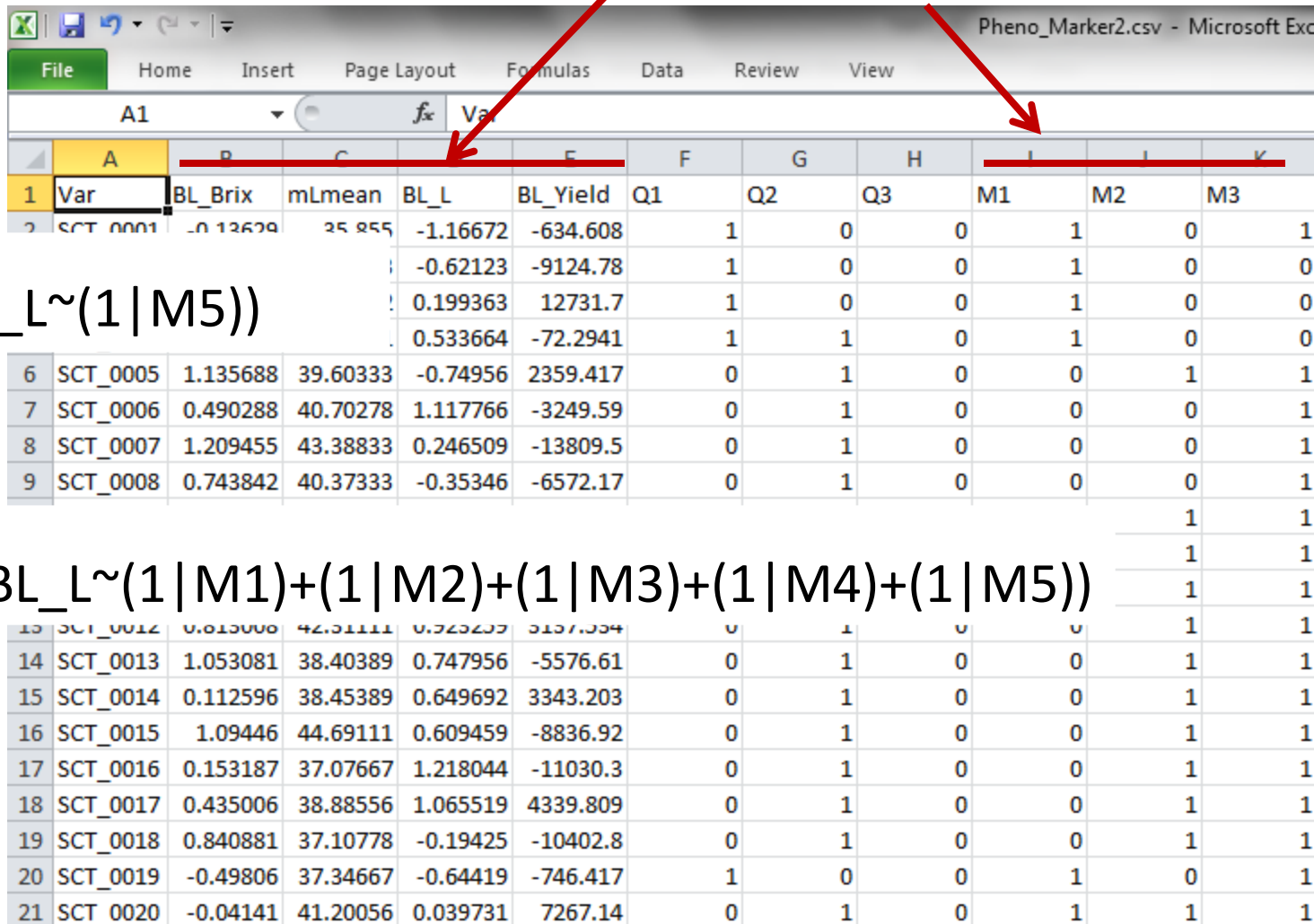
.

fit5 = lmer(BL_L~(1|M5))
ranef(fit5)

fitML = lmer(BL_L~(1|M1)+(1|M2)+(1|M3)+(1|M4)+(1|M5))
ranef(fitML)

Phenotype

Marker



	A	B	C	D	E	F	G	H	I	J	K
1	Var	BL_Brix	mLmean	BL_L	BL_Yield	Q1	Q2	Q3	M1	M2	M3
2	SCT_0001	-0.13629	35.855	-1.16672	-634.608	1	0	0	1	0	1
3				-0.62123	-9124.78	1	0	0	1	0	0
4				0.199363	12731.7	1	0	0	1	0	0
5				0.533664	-72.2941	1	1	0	1	0	0
6	SCT_0005	1.135688	39.60333	-0.74956	2359.417	0	1	0	0	1	1
7	SCT_0006	0.490288	40.70278	1.117766	-3249.59	0	1	0	0	0	1
8	SCT_0007	1.209455	43.38833	0.246509	-13809.5	0	1	0	0	0	1
9	SCT_0008	0.743842	40.37333	-0.35346	-6572.17	0	1	0	0	0	1
										1	1
										1	1
										1	1
										1	1
14	SCT_0013	1.053081	38.40389	0.747956	-5576.61	0	1	0	0	1	1
15	SCT_0014	0.112596	38.45389	0.649692	3343.203	0	1	0	0	1	1
16	SCT_0015	1.09446	44.69111	0.609459	-8836.92	0	1	0	0	1	1
17	SCT_0016	0.153187	37.07667	1.218044	-11030.3	0	1	0	0	1	1
18	SCT_0017	0.435006	38.88556	1.065519	4339.809	0	1	0	0	1	1
19	SCT_0018	0.840881	37.10778	-0.19425	-10402.8	0	1	0	0	1	1
20	SCT_0019	-0.49806	37.34667	-0.64419	-746.417	1	0	0	1	0	1
21	SCT_0020	-0.04141	41.20056	0.039731	7267.14	0	1	0	1	1	1

Table Comparing Stepwise with Multiple Regression (statistically naïve thought experiment)

Step-Wise					
	M1	M2	M3	M4	M5
0	0.2003	0	0	0	0.9449
1	-0.2003	0	0	0	-0.9449
Multiple-Regression					
0	0.1866	0.0001	0.0524	0	0.919
1	-0.1866	-0.0001	-0.0524	0	-0.919

Correction (regularization) involves introducing a penalty that places bounds on the regression

Marker homogeneous or marker-specific corrections

Ridge Regression (Tikhonov regularization) adds a constant λ to the diagonal of the matrix of coefficients

- makes solution unique
- shrinks estimates of marker effects toward 0
- $\lambda = \sigma^2_{\varepsilon} / \sigma^2_{\beta}$

Estimating the correction factor requires sampling the data.

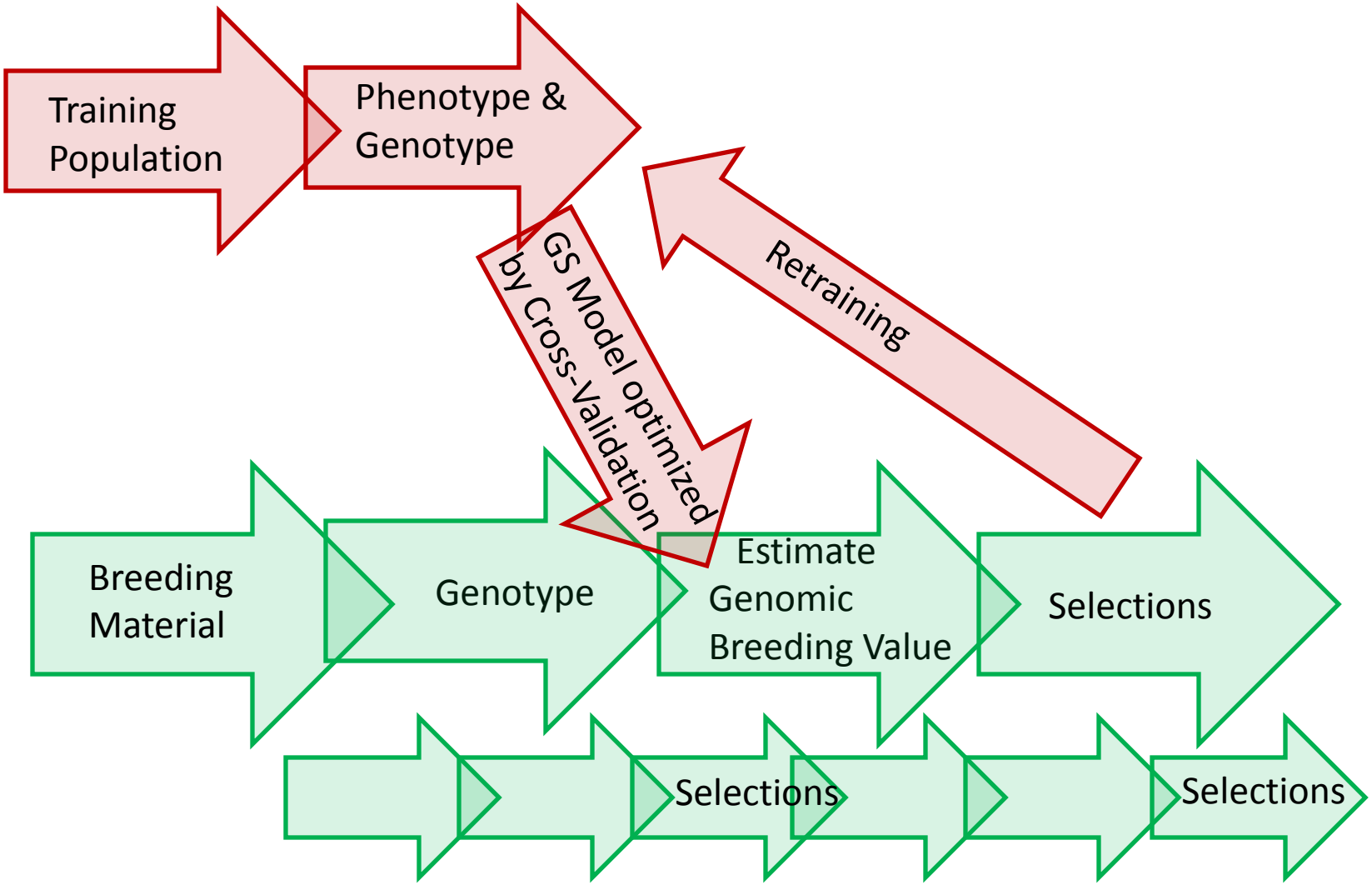
Marker homogeneous correction

RR-BLUP, Estimates of marker effects are penalized to the same extent; may not be appropriate if markers are located in regions of the genome that are not associated with genetic variance

Marker-specific correction

Least Absolute Value Selection & Shrinkage Operator (LASSO-BLUP)

Bayesian Linear Regression



Conceptual Diagram of GWS (adapted from Fig. 1, Heffner, Sorrells, and Jannink, 2009. CROP SCIENCE, VOL. 49)

Genomic selection is based on a prediction of breeding value

Accuracy depends on the size of the training population, number of markers, heritability of the trait, and the number of genes contributing to the trait

We can control the population size (and composition)

The number of markers is no longer limiting (SolCAP infinium Array, Genotyping by Sequencing, etc…)

The process is iterative, with statistical models re-estimated after each cycle of phenotypic evaluation

The relative efficiency of GWS will often be lower than direct phenotypic selection; value is to select during rapid generation turn over such that multiple cycles of selection can occur. The issue of what to select for remains…

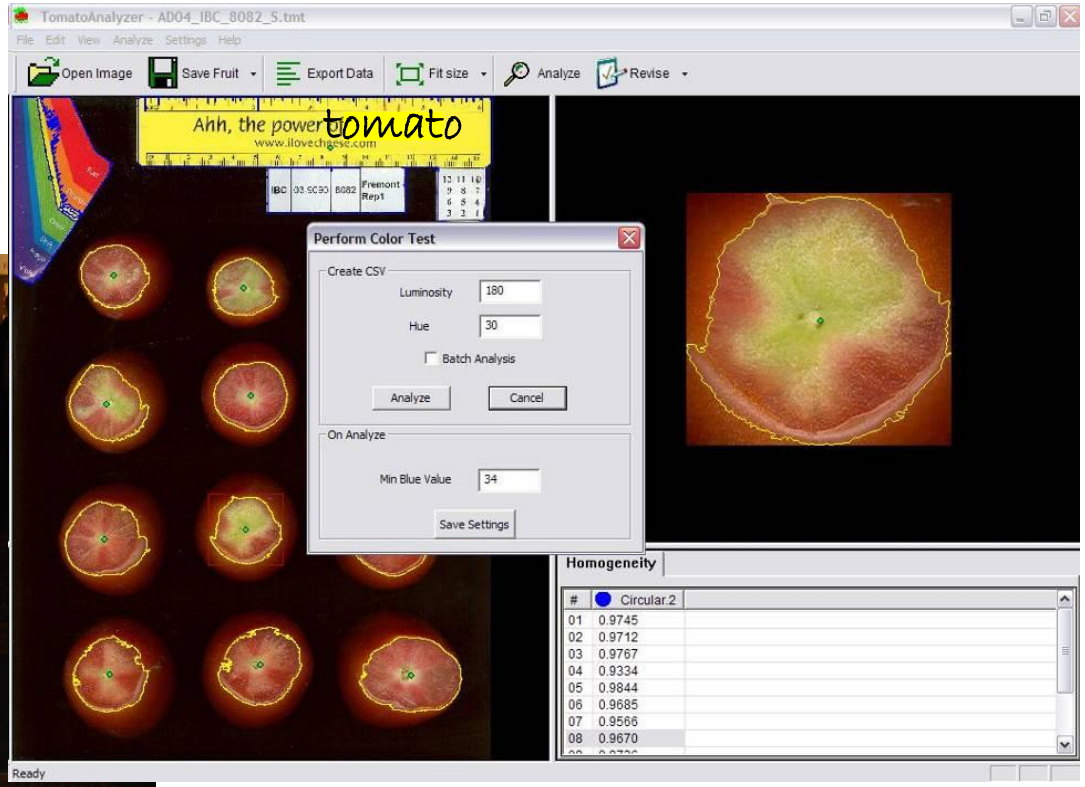
GEBV = $\sum_i^n \mathbf{X}_i \mathbf{g}_i$ is estimated for one trait, but how do we combine traits?

Multi Trait Index (MTI): Linear combination of observations used to compute a criterion for selection

Yield – directly valued
Color – directly valued in contracts in the midwest
Soluble Solids (BRIX) – value?
Disease resistance – value tied to yield loss or insurance adjustment?

Selection criteria are combined into a measure of net merit weighted based on the relative importance of all traits;
will differ between breeding programs due to breeding goals and market demands.

Multi Trait Index



We can measure color as: L, a, b, Hue, chroma, G, R, B, luminosity, % red tissue, % yellow tissue, etc…

Which measurements should we select for?

PCA and Development of Multi Trait Index

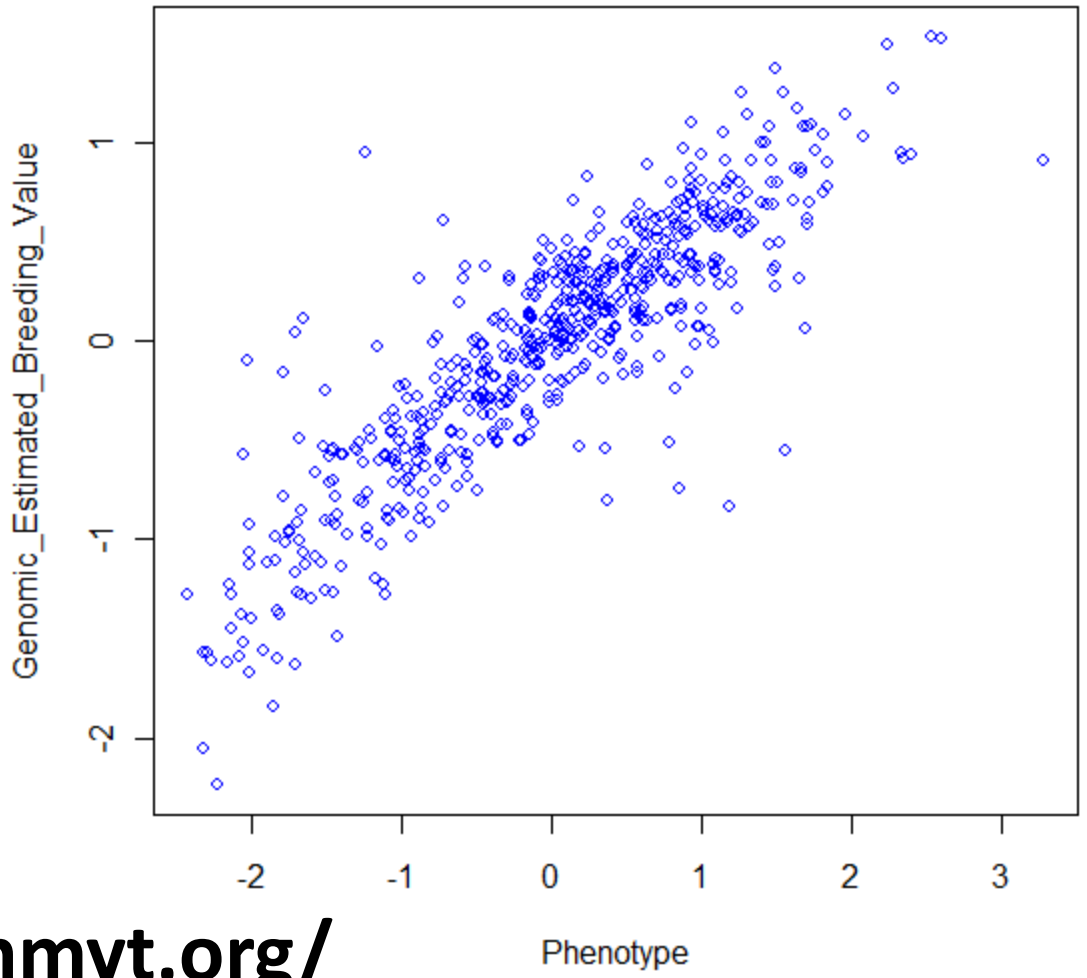
	Principal Component 1					Principal Component 2				
	BC2		BC2S4		TC19F2	BC2		BC2S4		TC19F2
	Fremont	Wooster	Fremont	Wooster	Fremont	Fremont	Wooster	Fremont	Wooster	Fremont
%YSD	0.4819	0.4535	0.4463	0.4157	-0.39377	-0.0267	-0.0911	-0.2147	-0.2896	0.301172
%RED	-0.4171	-0.4487	-0.401	-0.4371	0.444605	0.0485	-0.0451	0.1552	0.0994	-0.28885
L*	0.3731	0.338	0.3658	0.4263	0.110363	0.1677	0.0969	0.2416	0.0798	0.396884
a*	-0.3764	-0.3708	-0.3528	-0.1054	0.506258	0.4461	0.4675	0.4582	0.6575	0.156219
b*	0.2341	0.2807	0.3517	0.3974	0.190713	0.5801	0.5583	0.4777	0.3682	0.588045
Hue	0.5078	0.512	0.4609	0.4471	-0.43707	-0.0179	-0.0161	-0.159	-0.2279	0.347519
Chroma	0.0178	0.0198	0.2138	0.2923	0.389094	0.658	0.6707	0.6388	0.5293	0.421159
Proportion	0.5382	0.5283	0.6178	0.6228	0.517	0.3265	0.3138	0.2726	0.313	0.3333
Cumulative	-	-	-	-	-	0.8647	0.8422	0.8904	0.9358	0.8503

For three separate populations, PCA-1 is strongly weighted toward color uniformity and color while PCA-2 is weighted toward color intensity (Audrey Darrigues)

Predicted Gen. Value relative to BLUP of Phenotype (BLR package)

$GEBV = \sum_i^n \mathbf{X}_i \mathbf{g}_i$

Marker	Value
0	-0.32
1	1.23
0	0.40
1	-0.75
1	0.86
Sum	1.42



http://genomics.cimmyt.org/

Preparing data for GWS

Y - Phenotype of (n) individuals estimated as BLUPs

X – Marker matrix (n x k) with (k) markers scored on proportional scale (e.g. copies of common allele)

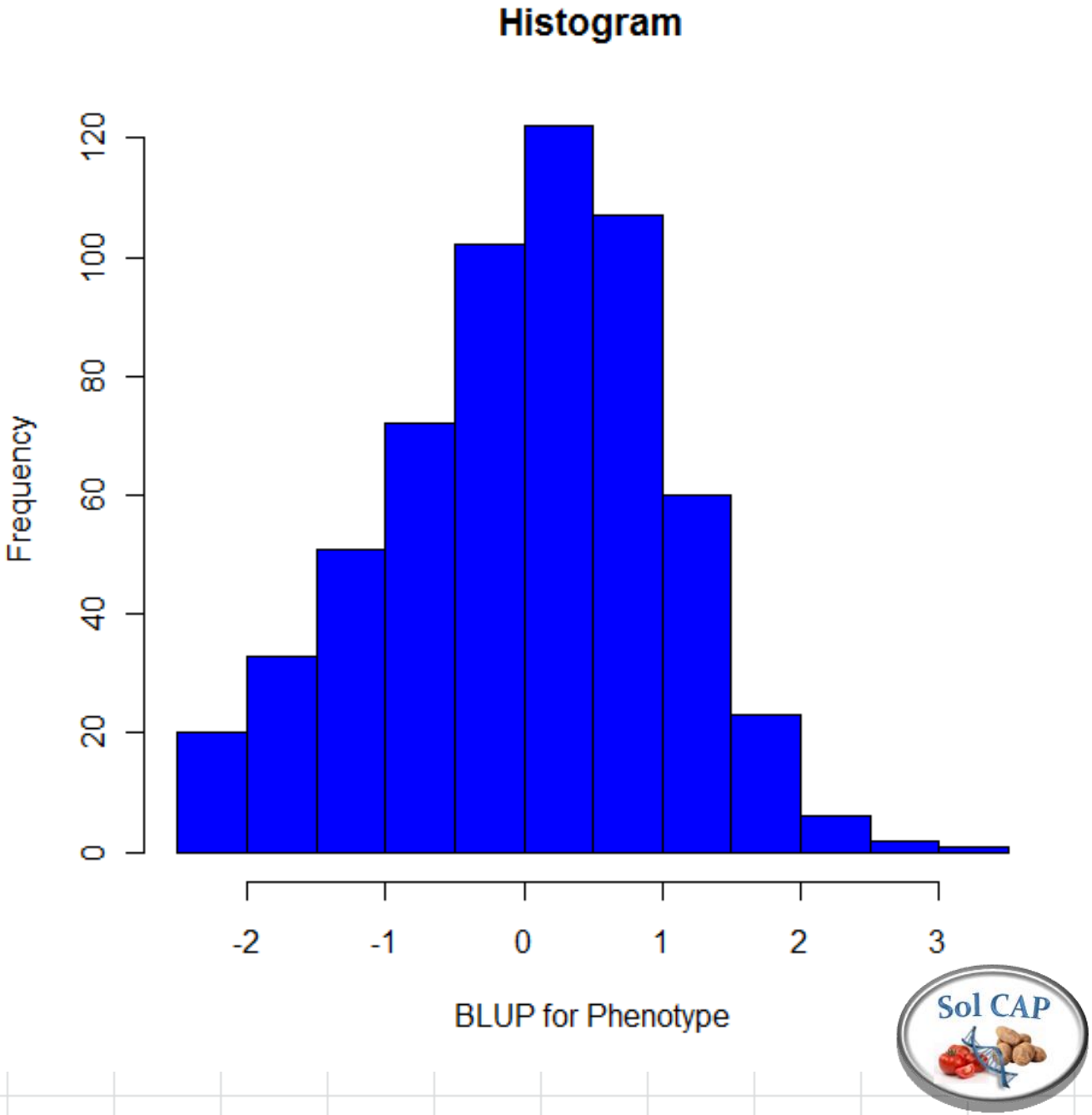
A – Kinship or pedigree matrix (n x n)

Y - phenotypic values

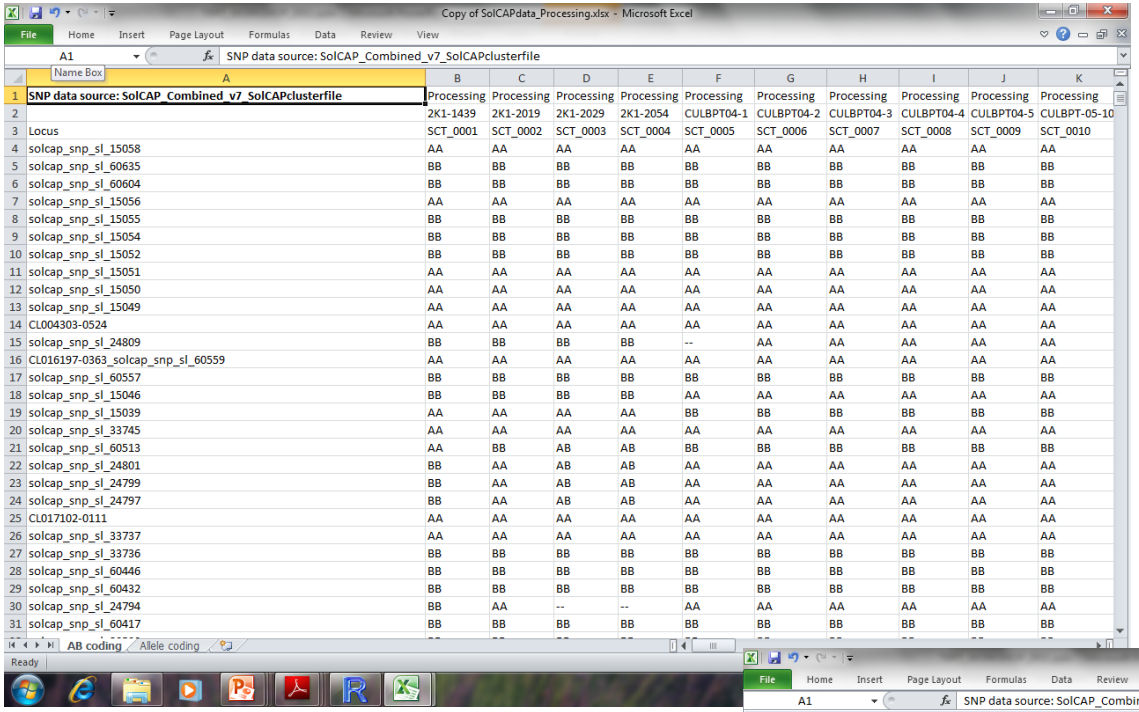
File Home Insert Page Layout

F3

	A	B	C	D
1		Accession	Trait	
2	1	SCT_0001	35.855	
3	2	SCT_0002	37.60833	
4	3	SCT_0003	38.73222	
5	4	SCT_0004	41.99611	
6	5	SCT_0005	39.60333	
7	6	SCT_0006	40.70278	
8	7	SCT_0007	43.38833	
9	8	SCT_0008	40.37333	
10	9	SCT_0009	41.73222	
11	10	SCT_0010	41.66722	
12	11	SCT_0011	43.22222	
13	12	SCT_0012	42.31111	
14	13	SCT_0013	38.40389	
15	14	SCT_0014	38.45389	
16	15	SCT_0015	44.69111	
17	16	SCT_0016	37.07667	
18	17	SCT_0017	38.88556	
19	18	SCT_0018	37.10778	
20	19	SCT_0019	37.34667	
21	20	SCT_0020	41.20056	
22	21	SCT_0021	38.64	



X – Marker Data

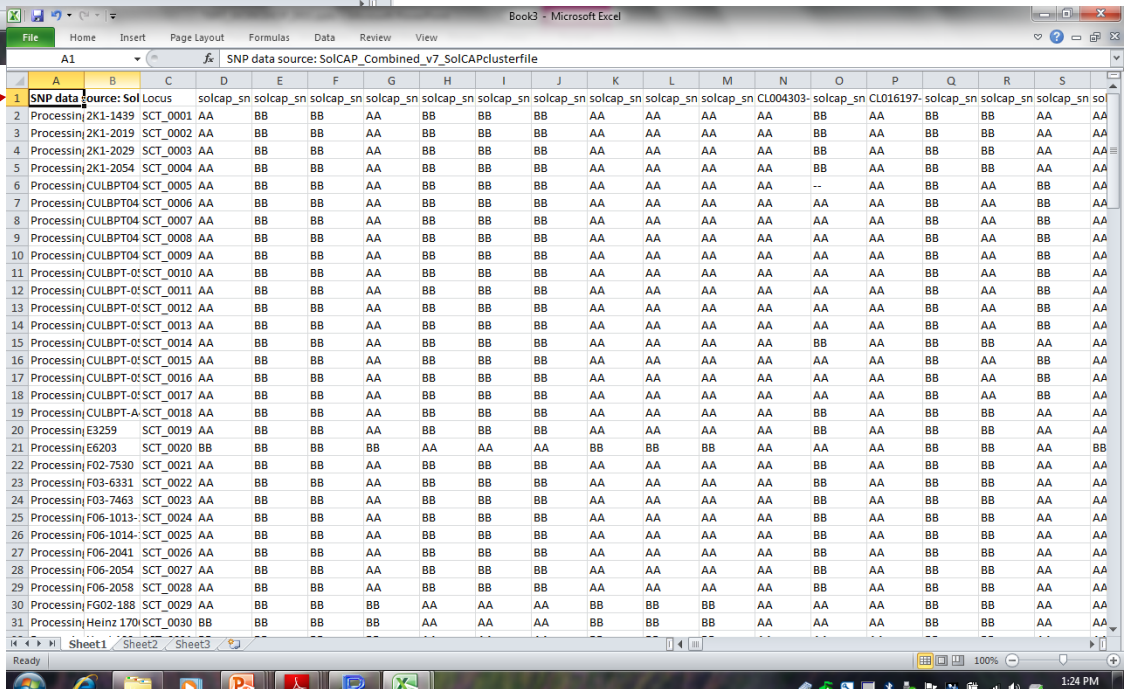


Processing
Variety Name
SCT_No.

Delete rows with redundant or unnecessary information

Marker

Transpose Data



Missing Data

Eliminate markers with >20% missing data
Impute alleles for markers with missing data using data for flanking markers (organize the markers by physical or genetic map)

Tools
PLINK
http://pngu.mgh.harvard.edu/~purcell/plink/pimputation.shtml
MATCH
http://www.sph.umich.edu/csg/abecasis/MACH/tour/imputation.html
IMPUTEv1
http://mathgen.stats.ox.ac.uk/impute/impute_v1.html

Common allele = A, replace AA with 2

Rare allele = B, replace BB with 0

Heterozygotes AB, replace with 1.

Eliminate monomorphic markers

=COUNTIF(B2:B142, “AA”)

117	SCT_0118	AA	BB	BB	AA	BB	BB	BB
118	SCT_0119	AA	BB	BB	AA	BB	BB	BB
119	SCT_0120	AA	BB	BB	AA	BB	BB	BB
120	SCT_0121	AA	BB	BB	AA	BB	BB	BB
121	SCT_0122	AA	BB	BB	AA	BB	BB	BB
122	SCT_0123	AA	BB	BB	AA	BB	BB	BB
123	SCT_0124	AA	BB	BB	AA	BB	BB	BB
124	SCT_0125	AA	BB	BB	AA	BB	BB	BB
125	SCT_0126	AA	BB	BB	AA	BB	BB	BB
126	SCT_0127	AA	BB	BB	AA	BB	BB	BB
127	SCT_0128	AA	BB	BB	AA	BB	BB	BB
128	SCT_0129	AA	BB	BB	AA	BB	BB	BB
129	SCT_0130	AA	BB	BB	AA	BB	BB	BB
130	SCT_0131	BB	BB	BB	BB	AA	AA	AA
131	SCT_0132	AA	BB	BB	AA	BB	BB	BB
132	SCT_0133	AA	BB	BB	AA	BB	BB	BB
133	SCT_0134	AA	BB	BB	AA	BB	BB	BB
134	SCT_0135	AA	BB	BB	AA	BB	BB	BB
135	SCT_0136	AA	BB	BB	AA	BB	BB	BB
136	SCT_0137	AA	BB	BB	AA	BB	BB	BB
137	SCT_0138	BB	BB	BB	BB	AA	AA	AA
138	SCT_0139	AA	BB	BB	AA	BB	BB	BB
139	SCT_0140	AA	BB	BB	AA	BB	BB	BB
140	SCT_0141	AA	BB	BB	AA	BB	BB	BB
141	SCT_0142	AA	BB	BB	AA	BB	BB	BB
142	SCT_0478	AA	BB	BB	AA	BB	BB	BB
143		130	0	0	130	10	10	10
144								

Sheet1 Sheet2 Sheet3

122	SCT_0123	2	2	2	2	2	2	2	2	2	2
123	SCT_0124	2	2	2	2	2	2	2	2	2	2
124	SCT_0125	2	2	2	2	2	2	2	2	2	2
125	SCT_0126	2	2	2	2	2	2	2	2	2	2
126	SCT_0127	2	2	2	2	2	2	2	2	2	2
127	SCT_0128	2	2	2	2	2	2	2	2	2	2
128	SCT_0129	2	2	2	2	2	2	2	2	2	2
129	SCT_0130	2	2	2	2	2	2	2	2	2	2
130	SCT_0131	0	0	0	0	0	0	0	0	2	0
131	SCT_0132	2	2	2	2	2	2	2	2	2	2
132	SCT_0133	2	2	2	2	2	2	2	2	2	2
133	SCT_0134	2	2	2	2	2	2	2	2	2	2
134	SCT_0135	2	2	2	2	2	2	2	2	2	2
135	SCT_0136	2	2	2	2	2	2	2	2	2	2
136	SCT_0137	2	2	2	2	2	2	2	2	0	0
137	SCT_0138	0	0	0	0	0	0	0	0	2	0
138	SCT_0139	2	2	2	2	2	2	2	2	2	2
139	SCT_0140	2	2	2	2	2	2	2	2	0	0
140	SCT_0141	2	2	2	2	2	2	2	2	2	2
141	SCT_0142	2	2	2	2	2	2	2	2	2	2
142	SCT_0478	2	2	2	2	2	2	2	2	2	2
143		130	130	10	10	10	130	130	130	141	23	14
144											
145											
146											

Baldo et al., 2011. AlleleCoder: a PERL script for coding codominant polymorphism data for PCA analysis. Plant Genetic Resources, Available on CJO 2011 doi:10.1017/S1479262111000839

Data

Phenotype matrix (Y)

Marker matrix (X)

Kinship Matrix (A)

MSA

http://i122server.vu-wien.ac.at/MSA/MSA_download.html

See tutorials:

http://www.extension.org/pages/32370/

Software Resources

Structure

PCA

STRUCTURE **http://pritch.bsd.uchicago.edu/structure.html**

see tutorials: **http://www.extension.org/pages/32492/**

Kinship

SPAGeDi

MSA **http://i122server.vu-wien.ac.at/MSA/MSA_download.html**

See tutorials: **http://www.extension.org/pages/32370/**

LD

Tassel **www.maizegenetics.net/tassel/**

GGT 2.0 **www.plantbreeding.wur.nl/UK/software_ggt.html**

GOLD (Graphical Overview of Linkage Disequilibrium)

http://www.sph.umich.edu/csg/abecasis/GOLD/

Software Resources

Haplotypes

PHASE (for short-range haplotypes)

http://www.stat.washington.edu/stephens/software.html

see practical exercises in:

www.ans.iastate.edu/section/abg/shortcourse/notes.pdf

Association Analysis and Genomic Selection

GenABLE -R library for Genome-wide association analysis

http://www.genabel.org/

EMMA (Efficient Mixed Model Association)

http://mouse.cs.ucla.edu/emma/index.html

TASSEL **http://www.maizegenetics.net/**

GAPIT

R-package BLR **http://genomics.cimmyt.org/**

Working Examples

Power of association Studies

R-package ldDesign

http://cran.r-project.org/web/packages/ldDesign/

ldDesign documentation

http://cran.r-project.org/web/packages/ldDesign/ldDesign.pdf

See example script under “Practical Exercises”; *Hayes, 2007. QTL Mapping, MAS, and Genomic Selection, Short Course Sponsored by Dept. of Animal Sciences and Animal Breeding and Genetics Group, Iowa State University*

http://www.ans.iastate.edu/section/abg/shortcourse/notes.pdf

Other functions:

ld.design , ld.power, ld.sim, etc...

Determining the power of association analysis using the R-package ldDesign

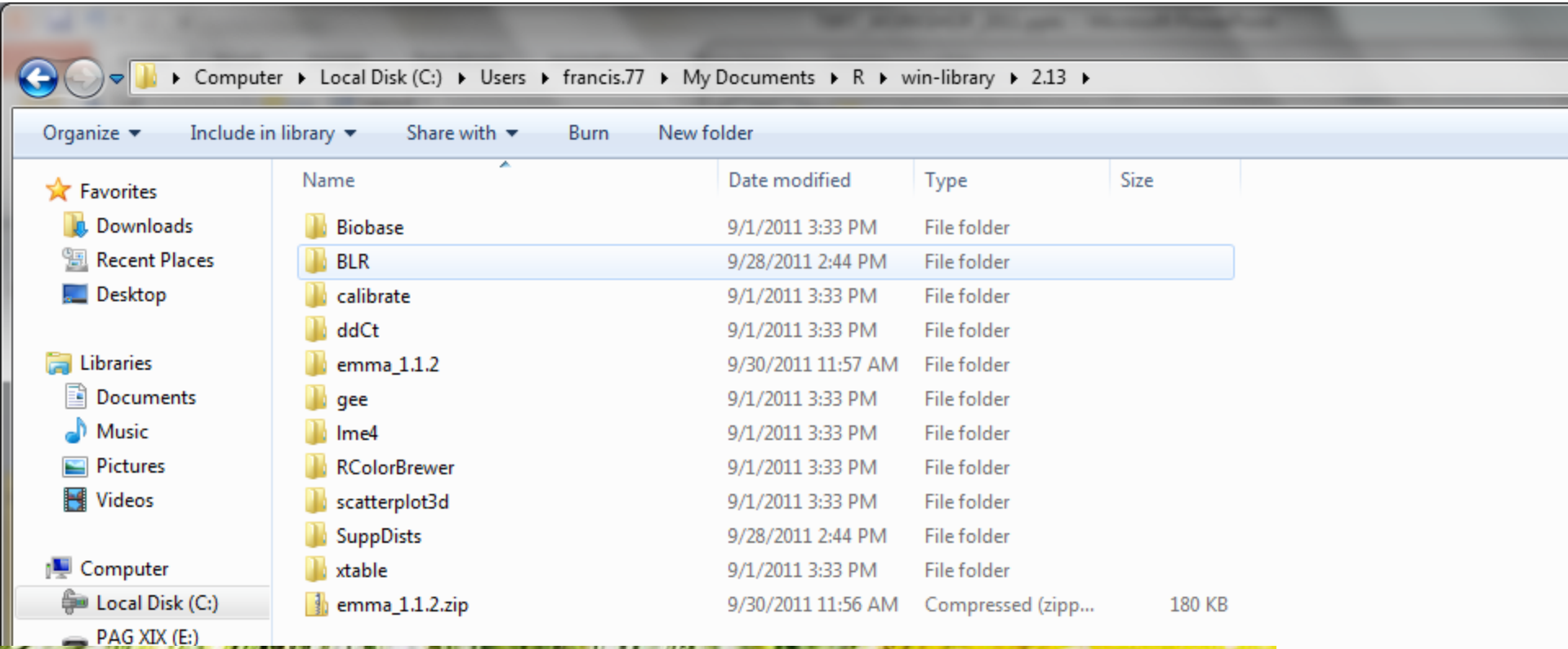
Rod Ball, Scion Research

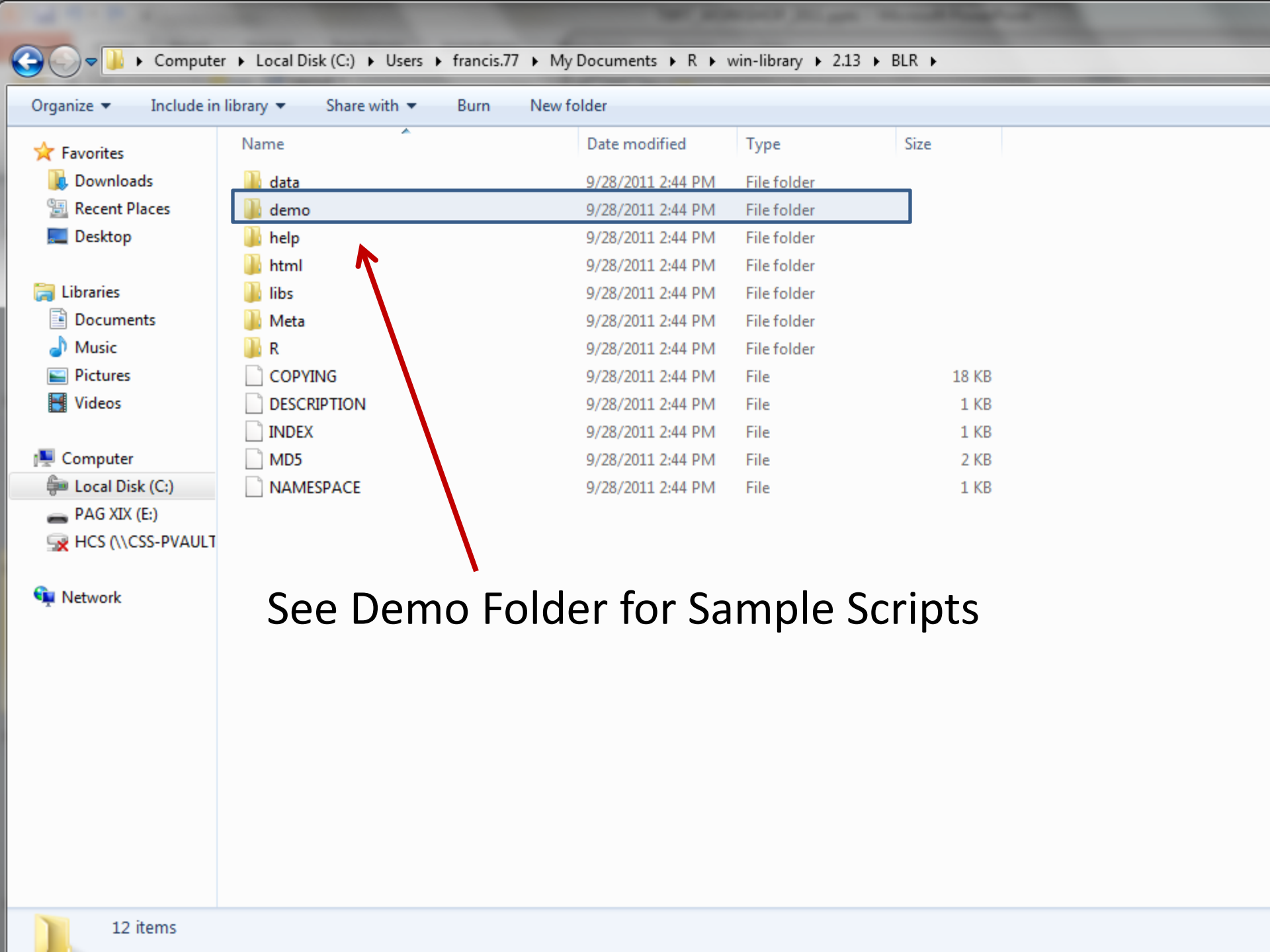
> luo.ld.power(n, p, q, D, h2, phi, Vp , alpha, print.it = TRUE, missclass.rate = 0))

function = luo.ld.power after (Luo, 1998, Heredity 80, 198–208)
n number of individuals genotyped and phenotyped
p frequency of Bi-allelic marker linked to the QTL
q frequency Bi-allelic QTL (gneraly p = q)
D Linkage disequilibrium coefficient
r^2 from LD analysis can be converted to D; $D = [p(1-p)(q(1-q)r^2]^{1/2}$
h2 QTL `heritability'; proportion of variance explained by the QTL (Vm/Vp)
phi Dominance ratio: = 0 for additive, = 1 for dominant allele effects
Vp phenotypic variance; an arbitray number can be used (*Vp* = 100)
alpha Significance level for hypothesis tests

Genomic prediction based on molecular markers and kinship using the BLR package in R

Paulino Pérez, Gustavo de los Campos, José Crossa, and Daniel Gianola **http://genomics.cimmyt.org/**




Computer ▸ Local Disk (C:) ▸ Users ▸ francis.77 ▸ My Documents ▸ R ▸ win-library ▸ 2.13 ▸ BLR ▸
Organize
Include in library
Share with
Burn
New folder
Favorites
Downloads
Recent Places
Desktop
Libraries
Documents
Music
Pictures
Videos
Computer
Local Disk (C:)
PAG XIX (E:)
HCS (\\CSS-PVAULT
Network
Name
Date modified
Type
Size
data 9/28/2011 2:44 PM File folder
demo 9/28/2011 2:44 PM File folder
help 9/28/2011 2:44 PM File folder
html 9/28/2011 2:44 PM File folder
libs 9/28/2011 2:44 PM File folder
Meta 9/28/2011 2:44 PM File folder
R 9/28/2011 2:44 PM File folder
COPYING 9/28/2011 2:44 PM File 18 KB
DESCRIPTION 9/28/2011 2:44 PM File 1 KB
INDEX 9/28/2011 2:44 PM File 1 KB
MD5 9/28/2011 2:44 PM File 2 KB
NAMESPACE 9/28/2011 2:44 PM File 1 KB
See Demo Folder for Sample Scripts
12 items

Loading Sample Data

```
infilepath <- "C:/PATH/wheat.RData"
load(infilepath)
ls()
# BEFORE RUNNING THE DATA FRAME LOOKS LIKE THIS
# [1] "A"          "infilepath" "sets"       "X"          "Y"
```

```
library(BLR)
Load(wheat)

y=Y[,Triat]  # selects a single column, heading Trait (set = 1),
             # from the phenotype matrix

### Creates a testing set with 100 observations
whichNa<-sample(1:length(y),size=100,replace=FALSE)
yNa<-y
yNa[whichNa]<-NA

### Runs the Gibbs sampler and assigns results to the object fm
fm<-BLR(y=yNa,XL=X,GF=list(ID=1:nrow(A),A=A),
                  prior=list(varE=list(df=3,S=0.25),
                  varU=list(df=3,S=0.63),
                  lambda=list(shape=0.52,rate=1e-4,
                  type='random',value=30)),
                  nIter=5500,burnIn=500,thin=1,
                  saveAt="example_")
```

#AFTER RUNNING THE DATA FRAME LOOKS LIKE THIS

> ls() # used to display the files in the dataframe

[1] "A" "COR.trn" "COR.tst" "fm" "infilepath“

[6] "MSE.trn" "MSE.tst" "sets"

[9] "whichNa" "X" "y" "Y" "yNa"

The file fm has the information we want!

>data = fm

>attach(data)

> ls(data)

[1] "bL" "burnIn" "fit" "lambda" "mu" "nIter" "prior"

[8] "SD.bL" "SD.u" "SD.yHat" "tau2" "thin" "u" "varE"

[15] "varU" "weights" "whichNa" "y" "yHat"

Writing output to files

```
> data2 = bL
> write.csv(data2, "C:/PATH/data2.csv")
> data3 = u
> write.csv(data3, "C:/PATH/data3.csv")
> data4 = yHat
> write.csv(data4, "C:/PATH/data4.csv")
```

Concluding remarks:

Accurate and objective phenotypes remain a limiting factor for tomato

Most of the changes in breeding strategy that will improve the power/efficiency of GWS will also improve traditional phenotype-based breeding

Use BLUPs to estimate trait values

Use pedigree/kinship information to strengthen estimates of breeding values of individuals based on trait BLUPs

Use larger populations

Acknowledgments

Collaborators, OSU

Heather Merk
Sung-Chur Sim
Troy Aldrich
Matt Robbins
Audrey Darrigues

Collaborators, MSU

David Douches
C Robin Buell
John Hamilton
Kelly Zarka

Collaborators, Cornell

Walter de Jong
Lucas Mueller
Joyce van Eck
Naama Menda

Collaborators, UCD

Allen Van Deynze
Kevin Stoffel
Alex Kozic

Industry Collaborators

Cindy Lawley
Martin Ganal

Collaborators, UIB

Hipolito Medrano
Pep Cifre
Josefina Bota
Miquel Angel Conesa

Funding

USDA/AFRI

This project is supported by the Agriculture and Food Research Initiative of USDA's National Institute of Food and Agriculture.

PBG Webinar Series

- Putting research into practice
- Suggest topics
- Sign up to present

PBGworks
Home of the Plant Breeding and Genomics Community of Practice

USDA NIFA
United States Department of Agriculture
National Institute of Food and Agriculture

Sol CAP

eXtension

Contact: Heather Merk
merk.9@osu.edu