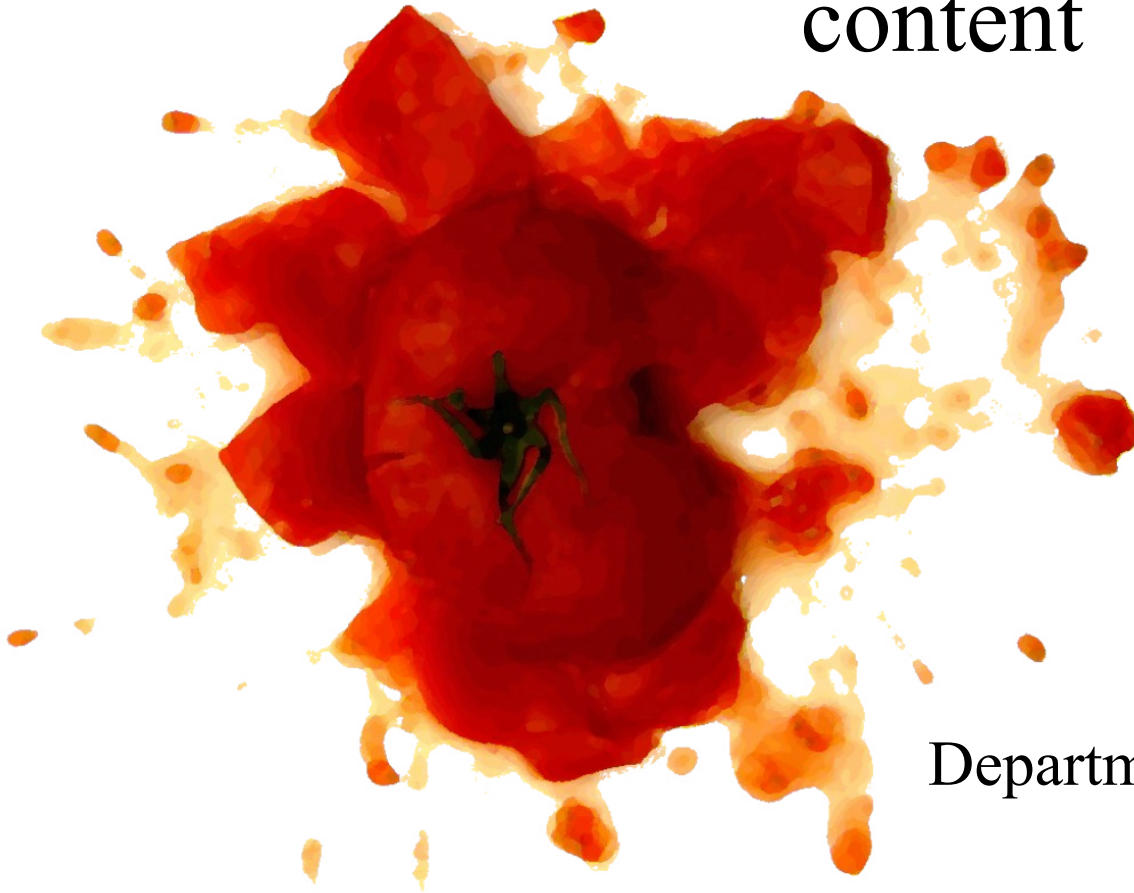


Breeding with molecular markers: a case study for tomato color and lycopene content



Pushpalatha G.

Professor & Head,
Department of Plant Biotechnology,
MSSSoA, CUTM,
Odisha

Email: pushpabhagyalakshmi@gmail.com

Case study

Measuring traits, heritability and gain under selection

Developing markers for “candidate genes”

Estimating the proportion of phenotypic variance explained by a marker (related to concept of heritability)

Demonstrate where to apply markers in a breeding program:
relative efficiency of selection

Select for candidate genes

Select for elite background “background selection”.



Measuring traits: Color

- Color correlates with lycopene and beta carotene content. Beta carotene is a nutrient.
- Color uniformity determines product grade and grower premium.
- Color disorders (e.g. YSD) reduce color uniformity, price and nutritional quality.



Fresh-market



Processed

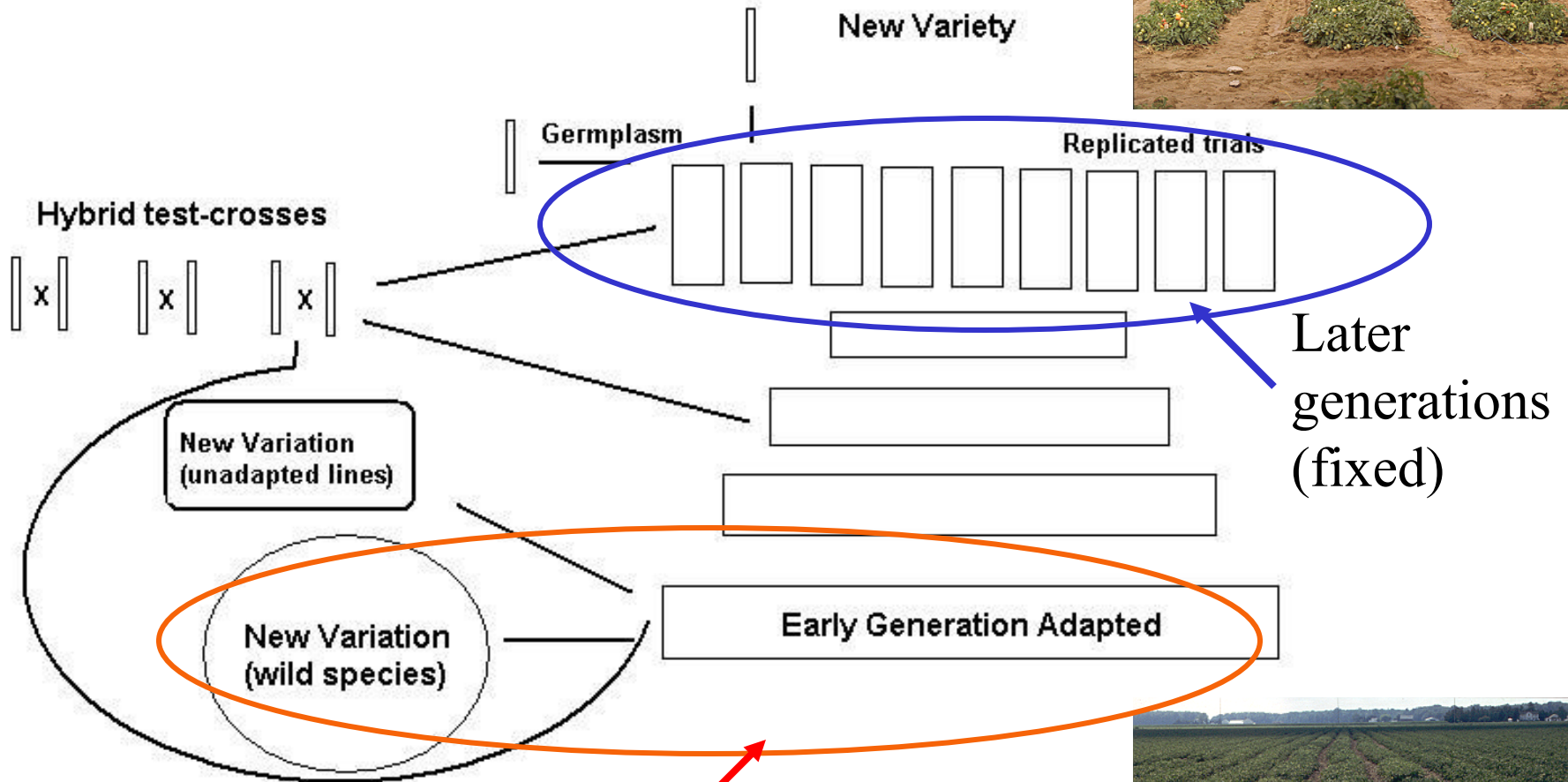
Plant Breeding is predominantly “phenotype” based

We select on the basis of traits

Growers buy varieties on the basis of trait performance

Measuring traits is therefore really important

Breeding programs have populations at all stages

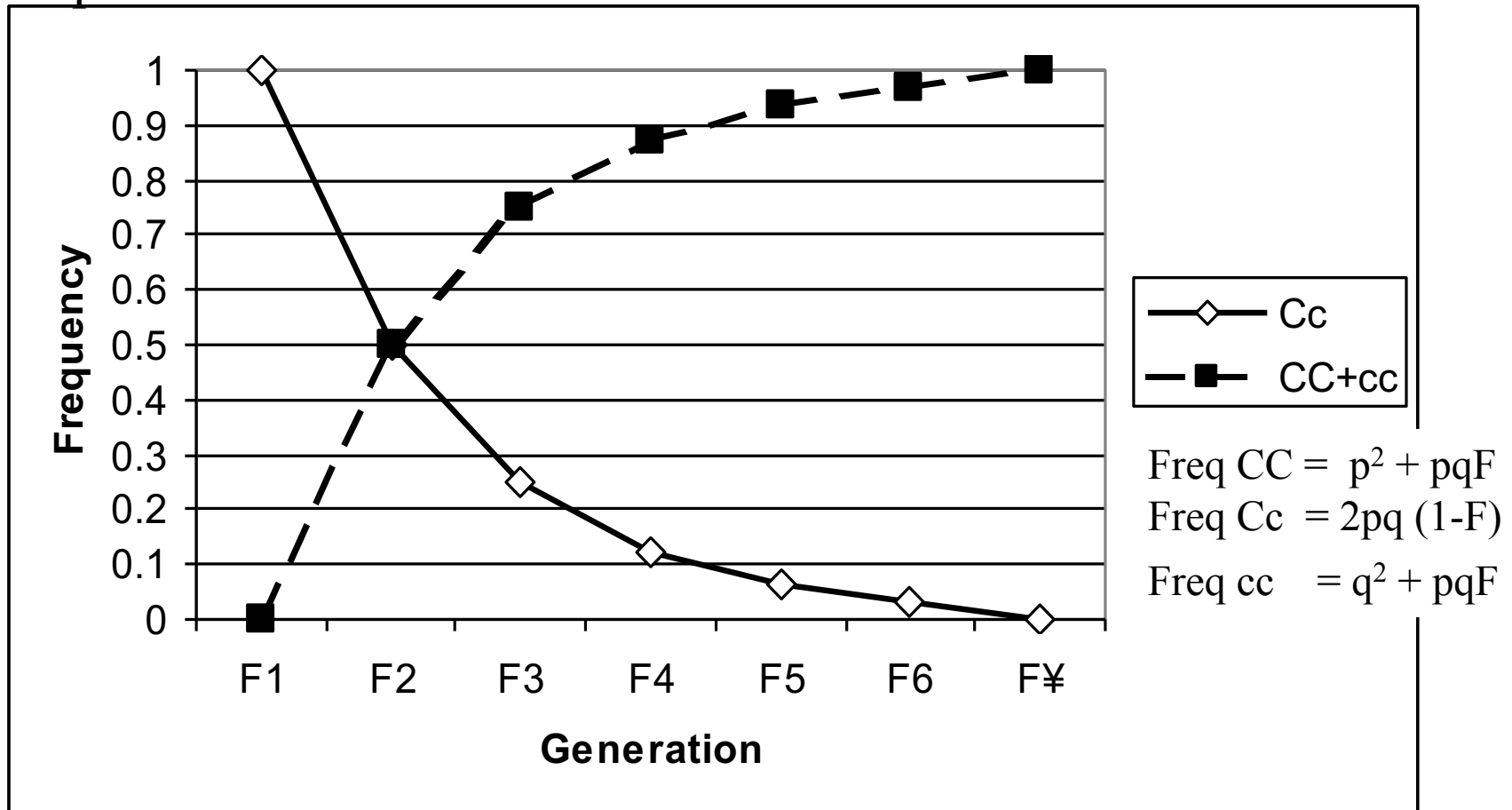


Early generation (not fixed)



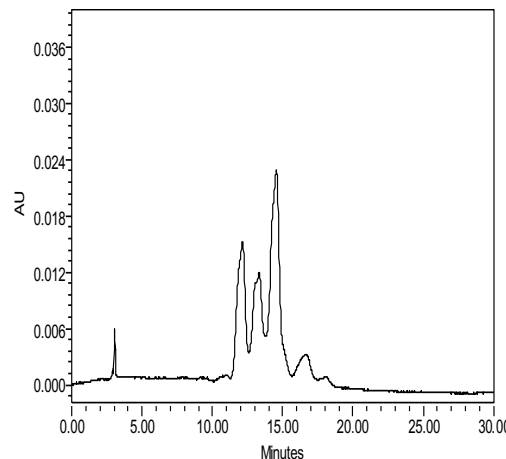
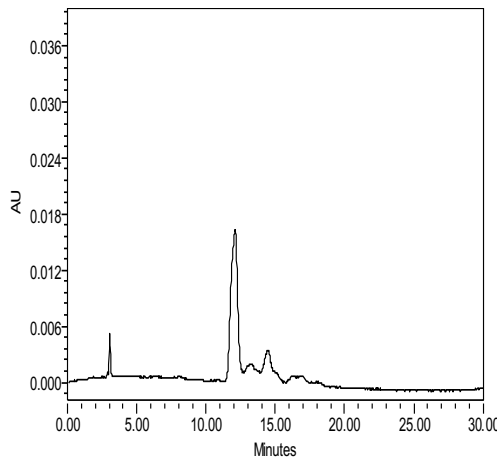
Review: inbreeding drives individuals toward homozygosity.

A population of inbred lines can be replicated because seed from each plant in a line are identical.



Frequency of heterozygotes (Cc) and homozygotes (CC+cc) in each generation of selfing a hybrid (F1).

Measuring Traits



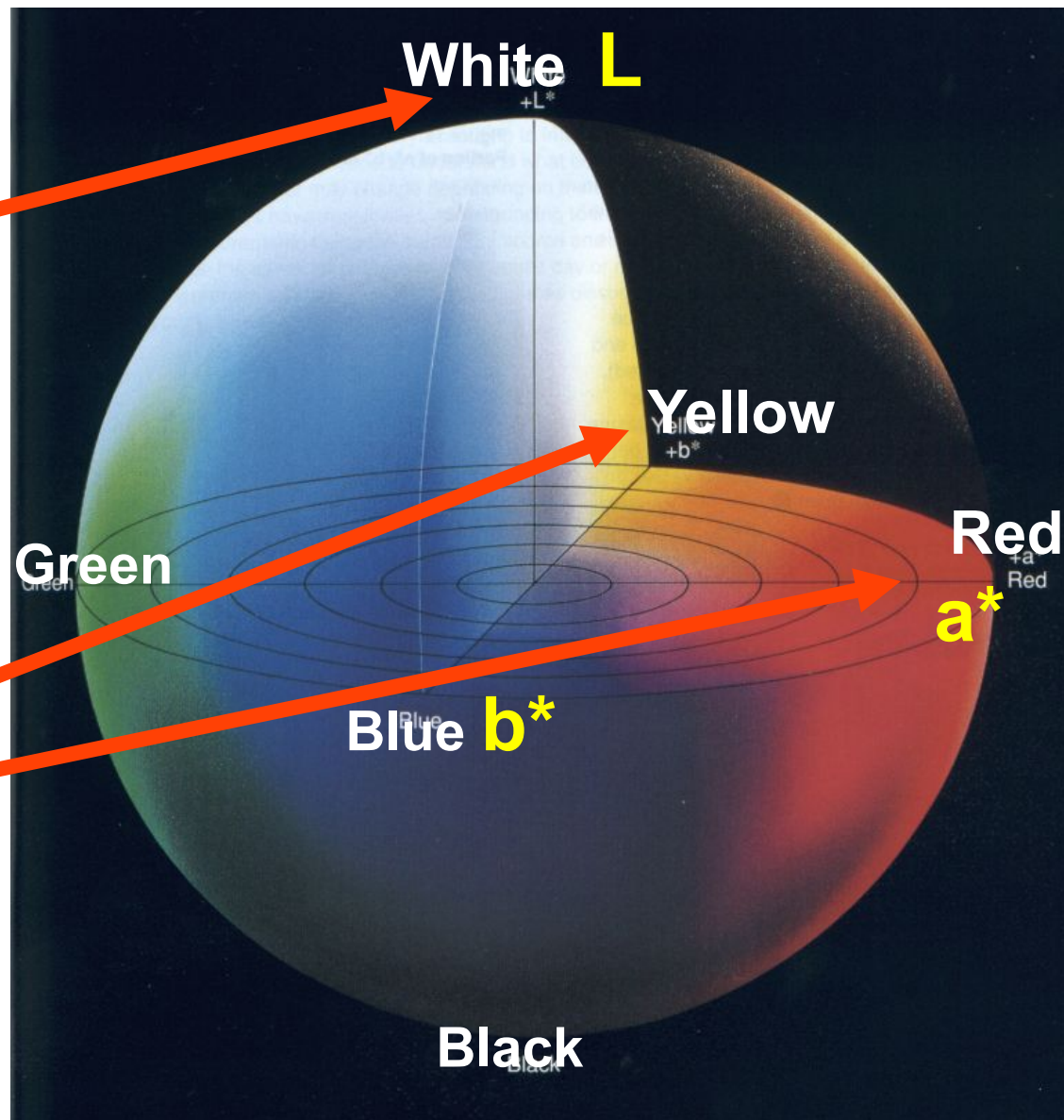
Often a balance between accuracy and efficiency of data collection

Measuring traits: Objective measures of color

Lightness

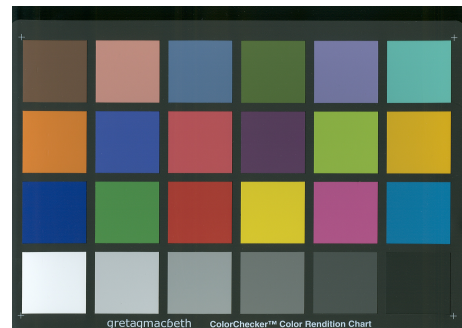
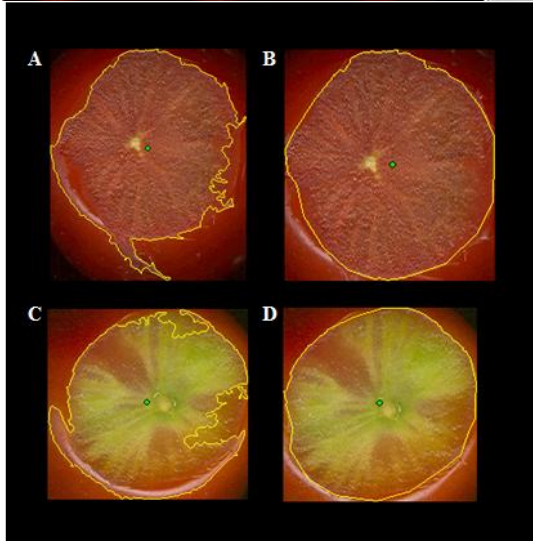
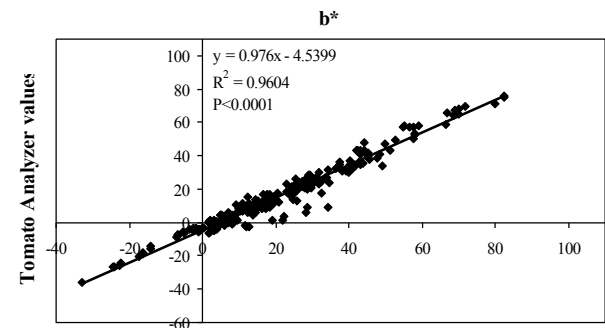
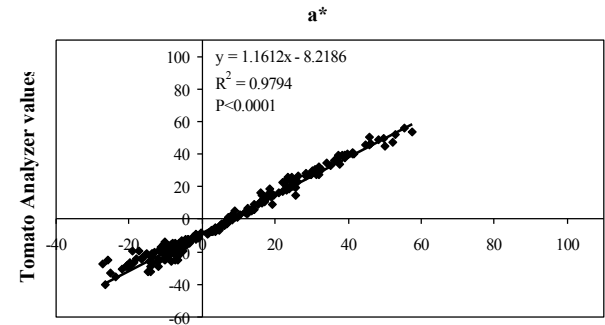
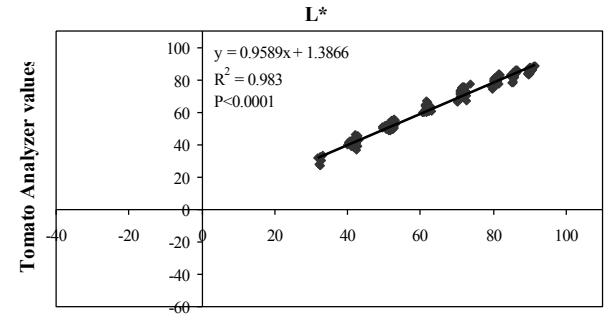
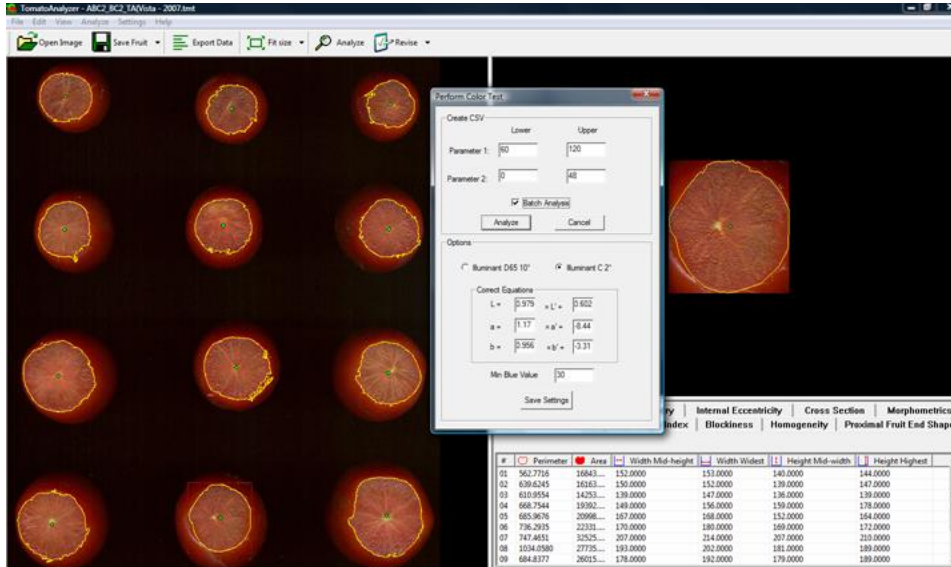
Chromaticity coordinates

Minolta CR-300



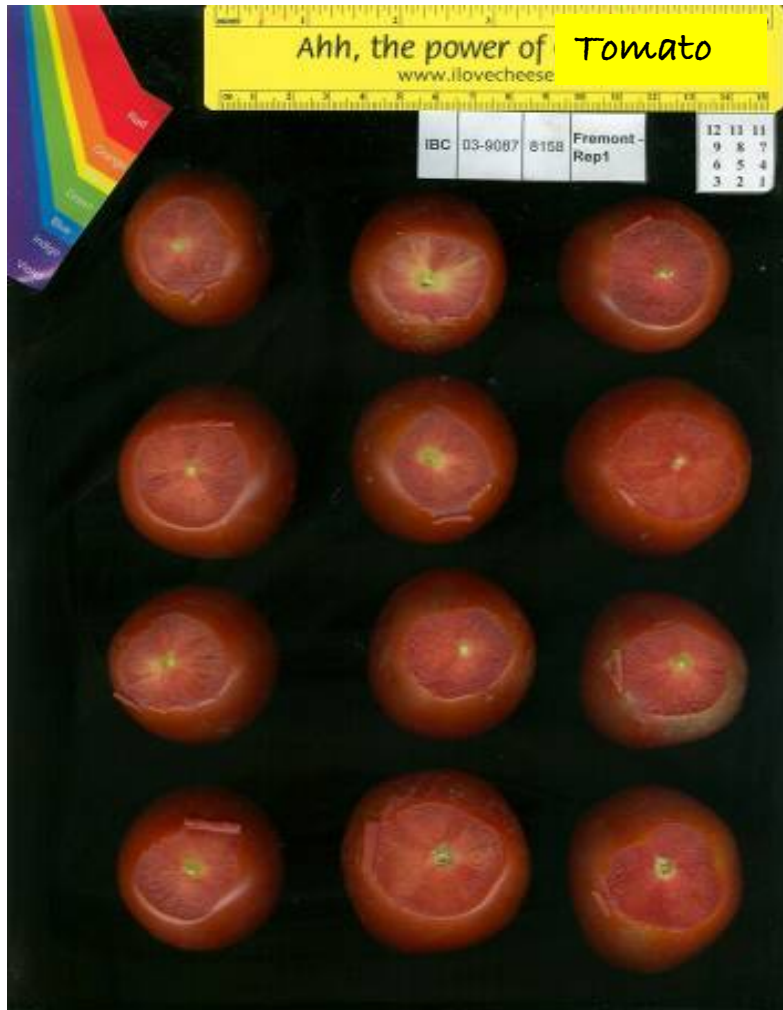
Phenotyping from digital images: fast and objective:

Tomato Analyzer (Brewer et al., *Plant Physiology*, 2006, **141**, 15-25;
Darrigues et al., *JASHS*, 2008, **133**, 579-586)

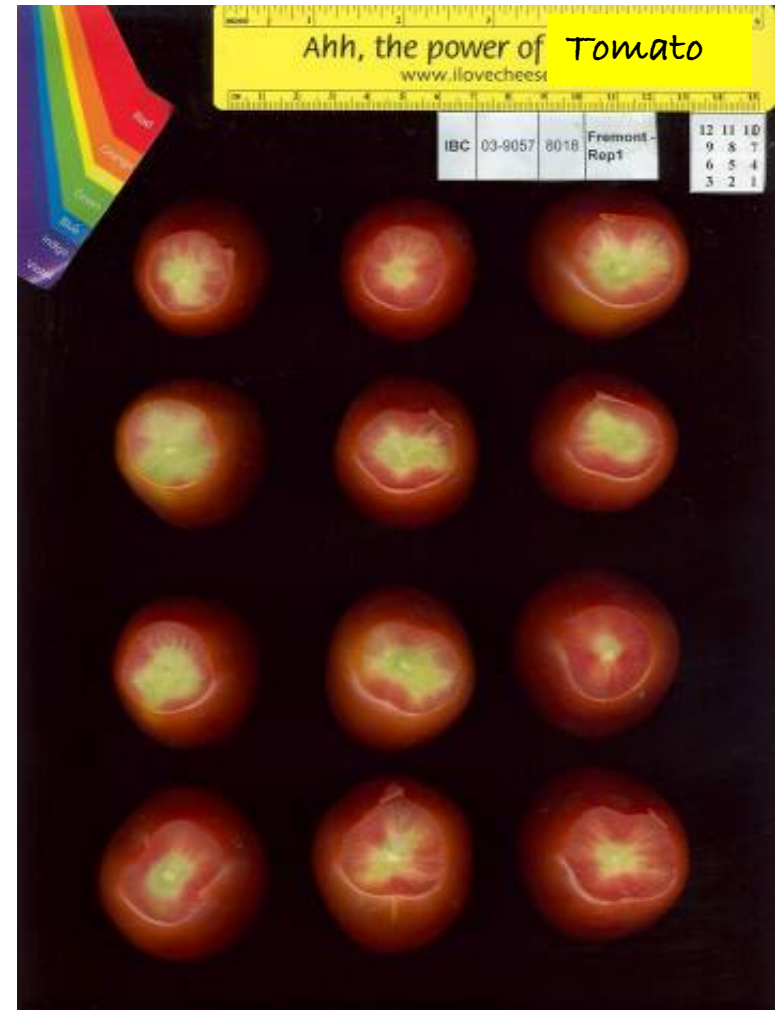


Colorimeter values

Analysis of color from Scanned images is consistent with colorimeter measurements and visual observations

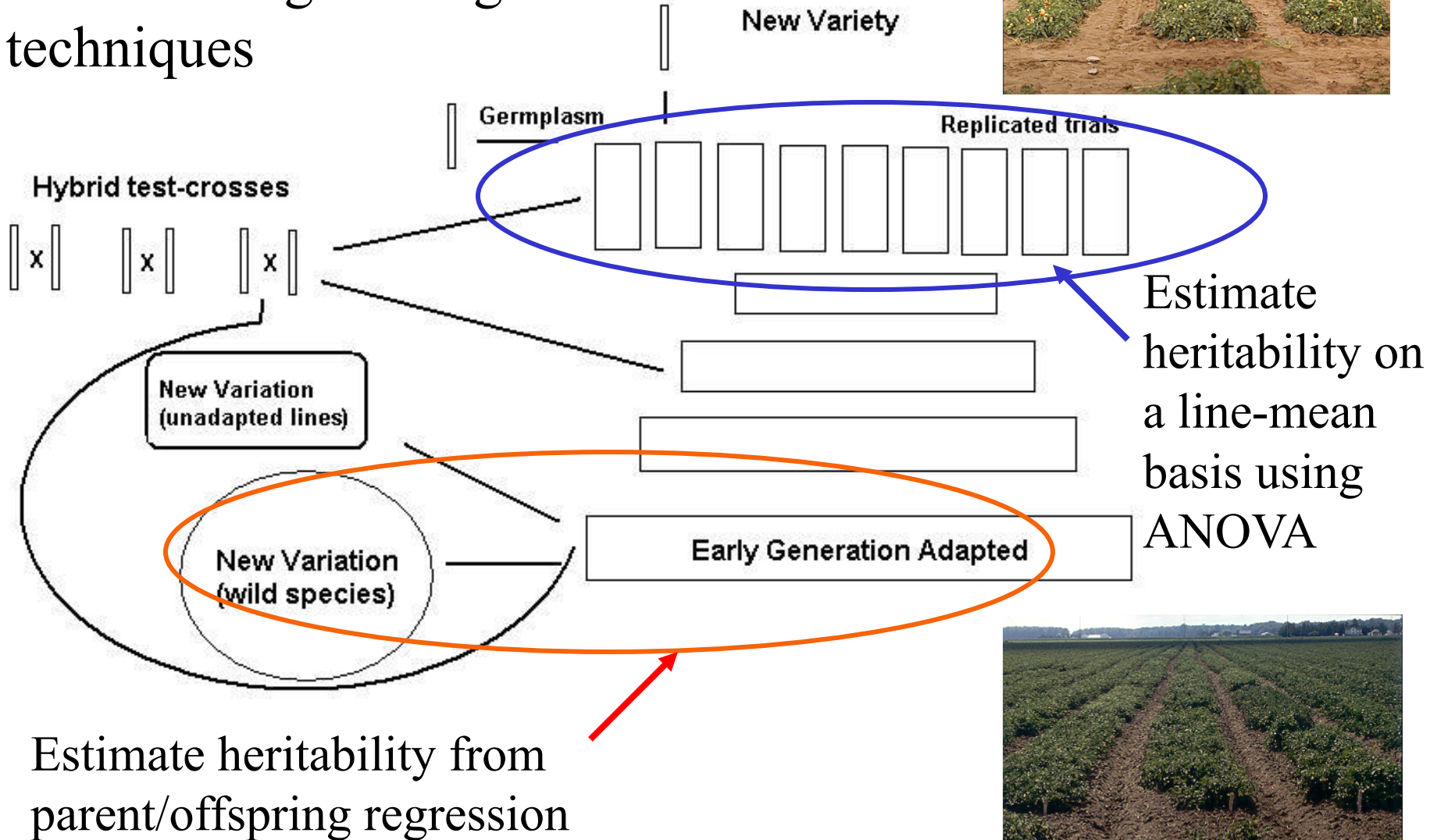


Rank 1:



Rank 216:

We can measure the genetic contribution to a trait (heritability) at different stages using different techniques



Estimating heritability using variance estimates from ANOVA

Source	Expected MS
gen	$\text{Var}(\text{Error}) + p \text{Var}(\text{rep} * \text{gen}) + q \text{Var}(\text{year} * \text{gen}) + (p-1)(q-1) \text{Var}(\text{gen})$
rep	$\text{Var}(\text{Error}) + n \text{Var}(\text{year} * \text{rep}) + p \text{Var}(\text{rep} * \text{gen}) + (n)(p-1) \text{Var}(\text{rep})$
year	$\text{Var}(\text{Error}) + n \text{Var}(\text{year} * \text{rep}) + q \text{Var}(\text{year} * \text{gen}) + (p-1)(q-1) \text{Var}(\text{year})$
year*gen	$\text{Var}(\text{Error}) + q \text{Var}(\text{year} * \text{gen})$
rep*gen	$\text{Var}(\text{Error}) + p \text{Var}(\text{rep} * \text{gen})$
year*rep	$\text{Var}(\text{Error}) + n \text{Var}(\text{year} * \text{rep})$
gen*year*rep	$\text{Var}(\text{Error})$

To estimate $\text{Var}(\text{Gen})$ use MS from ANOVA and solve. e.g. for 2 rep, 2 year trial:

$$13.601 = \text{Var}(\text{Error}) + 2 \text{Var}(\text{rep} * \text{gen}) + 2 \text{Var}(\text{year} * \text{gen}) + 4 \text{Var}(\text{gen})$$

$$2.6926 = \text{Var}(\text{Error}) + 2 \text{Var}(\text{year} * \text{gen})$$

$$10.9084 = 2 \text{Var}(\text{rep} * \text{gen}) + 4 \text{Var}(\text{gen})$$

$$3.4639 = 2 \text{Var}(\text{rep} * \text{gen}) + 2.77724$$

$$7.447 = 4 \text{Var}(\text{gen}) - 2.77724$$

$$10.22174 = 4 \text{var}(\text{gen})$$

$$2.55 = \text{var}(\text{gen}) = \sigma^2(G)$$

$$h^2 = [\sigma^2(G)] / [\sigma^2(G) + \sigma^2_{ge/e} + \sigma^2_{\text{error}/r * e}]$$

Objective data from parents, offspring, and/or replicated trials allows us to estimate the genetic contribution to observed variation in a trait (heritability $h^2 = V_{\text{additive genetic}} / V_{\text{total}}$)

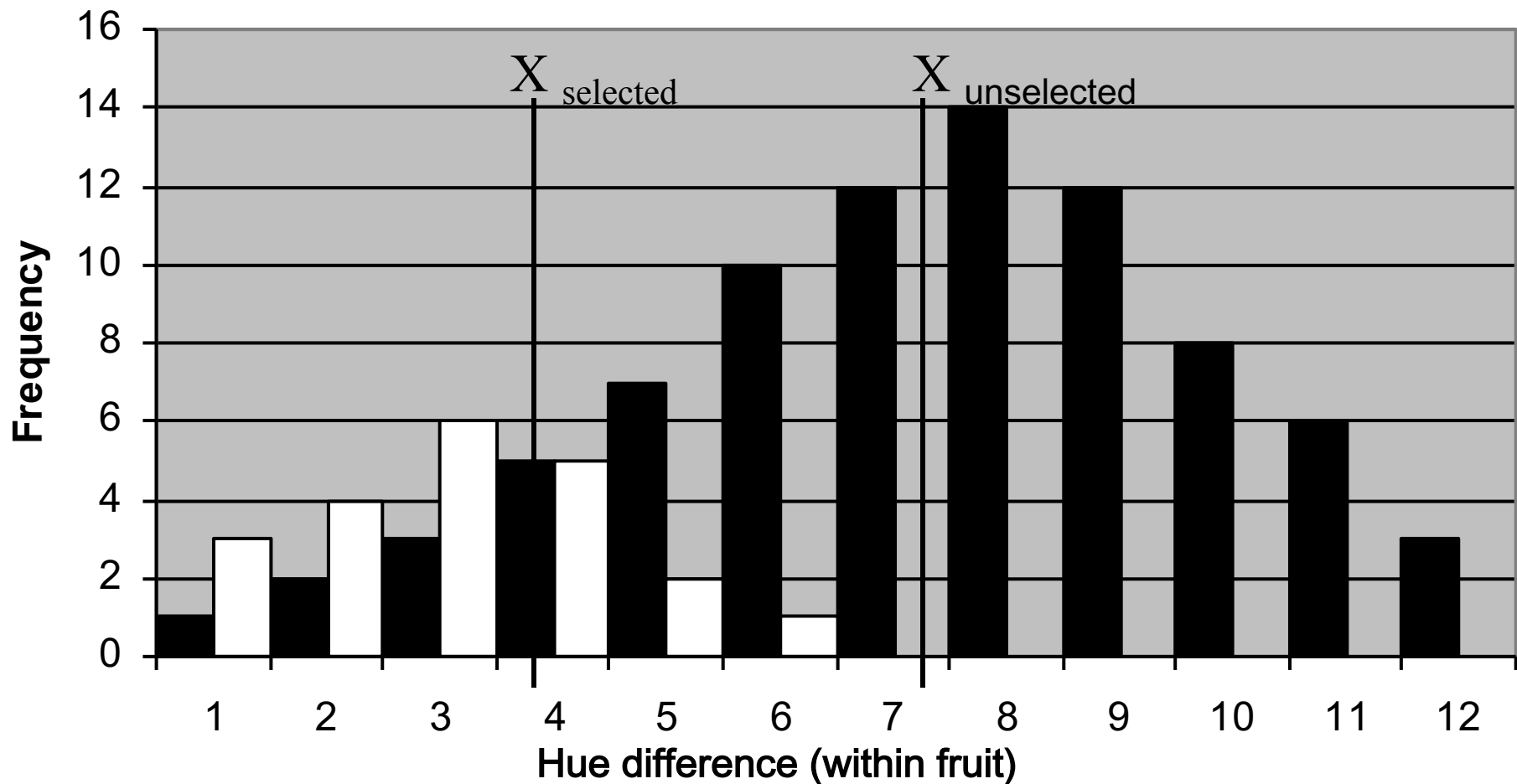
Expected and realized gain under selection is a function of heritability, the phenotypic variance in a population, and the selection differential (K) in standard deviation units.

Heritability can be estimated from parent-offspring regression or from variance partitioning (ANOVA)

Heritability for color increased with objective measurement.

Heritability for color increased when selection is based on a line or family across locations and years

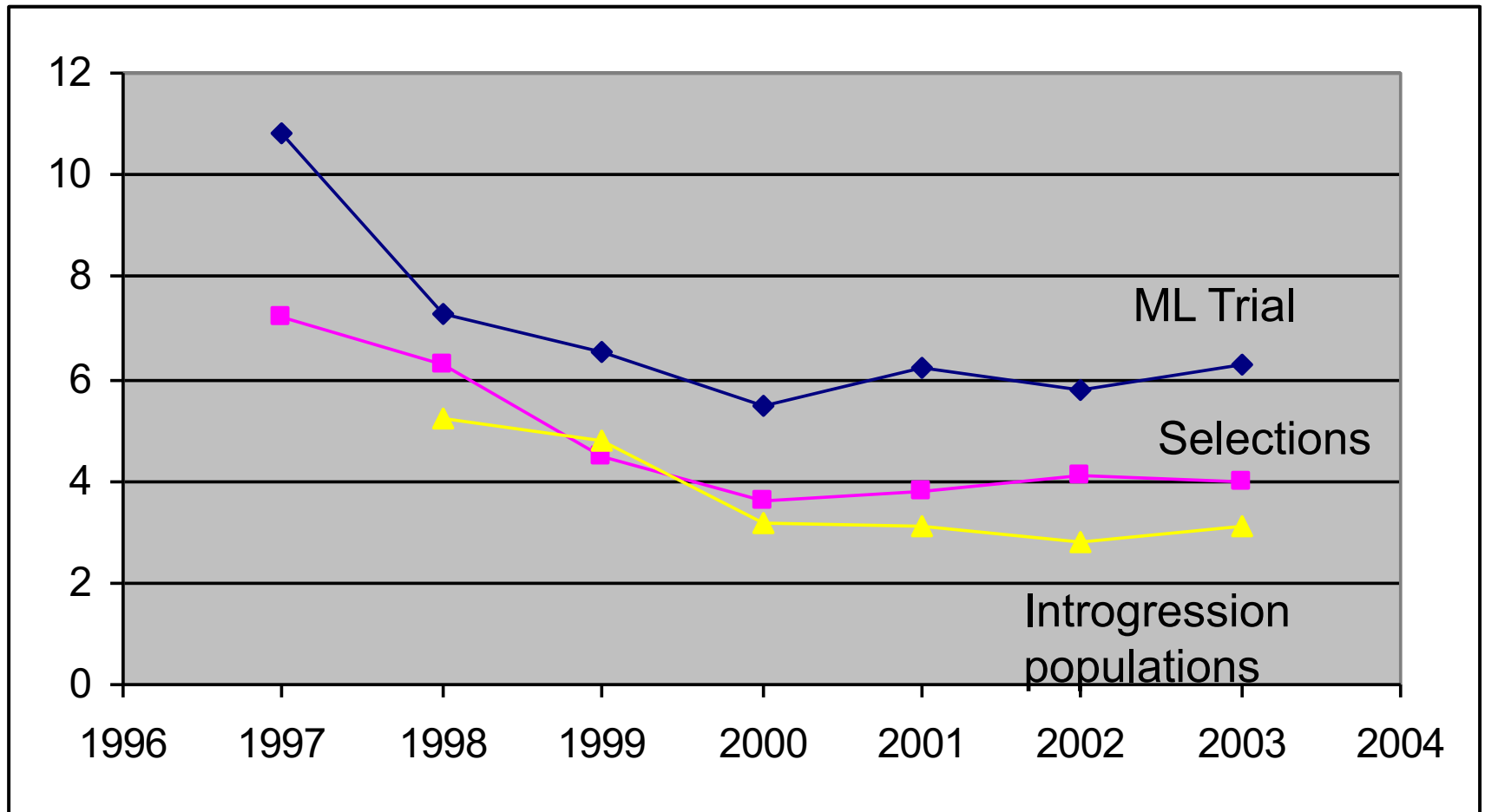
<u>Trait</u>	<u>individual</u>	<u>line or family</u>
L	0.11	0.57
Hue	0.07	0.39



	Unselected	Selected
Average	7.41	3.71
STDEV	1.86	0.63

$$\Delta G = h^2(\sigma_p)K$$

Selection progress: reduction in within fruit variation (hue difference w/in fruit)



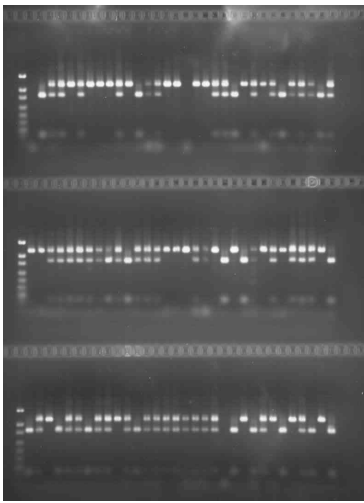
Take home message:

Score traits objectively on a quantitative scale whenever possible (improved measurement often means higher heritability).

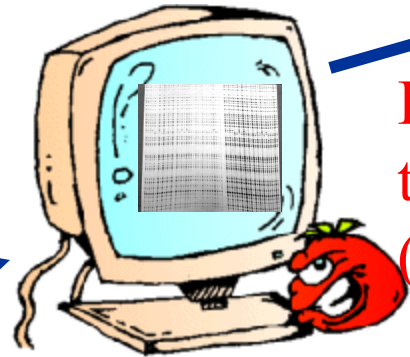
Selection (K) is in standard deviation units relative to the population (σ_p).

Estimate heritability and realized gain at different stages of a breeding program.

Measure progress.

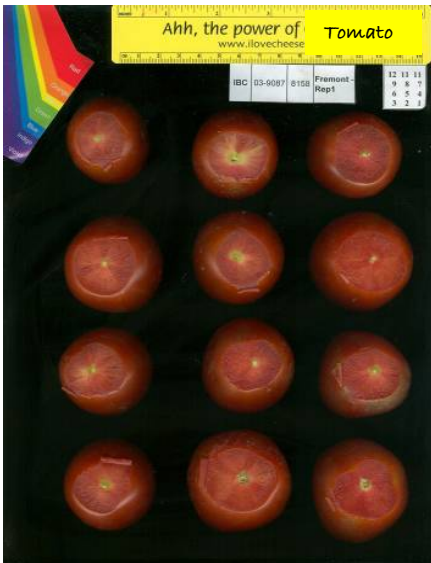


Genetic marker data

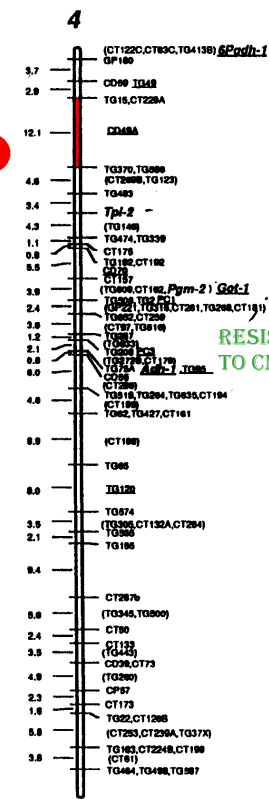


Establish marker-trait association (linkage)

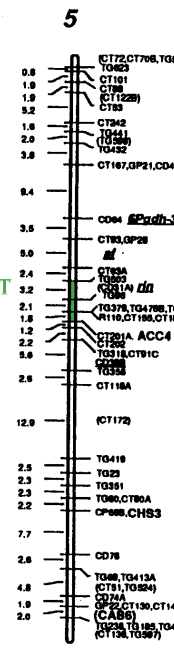
Trait data



RED COLOR (OHIO)



RESISTANT TO CMM



How can markers help make breeding more “efficient”?

- Identifying markers
- Using markers for selection

How do I identify markers linked to a trait?

Random approach (high density mapping).

“Candidate gene” approach.

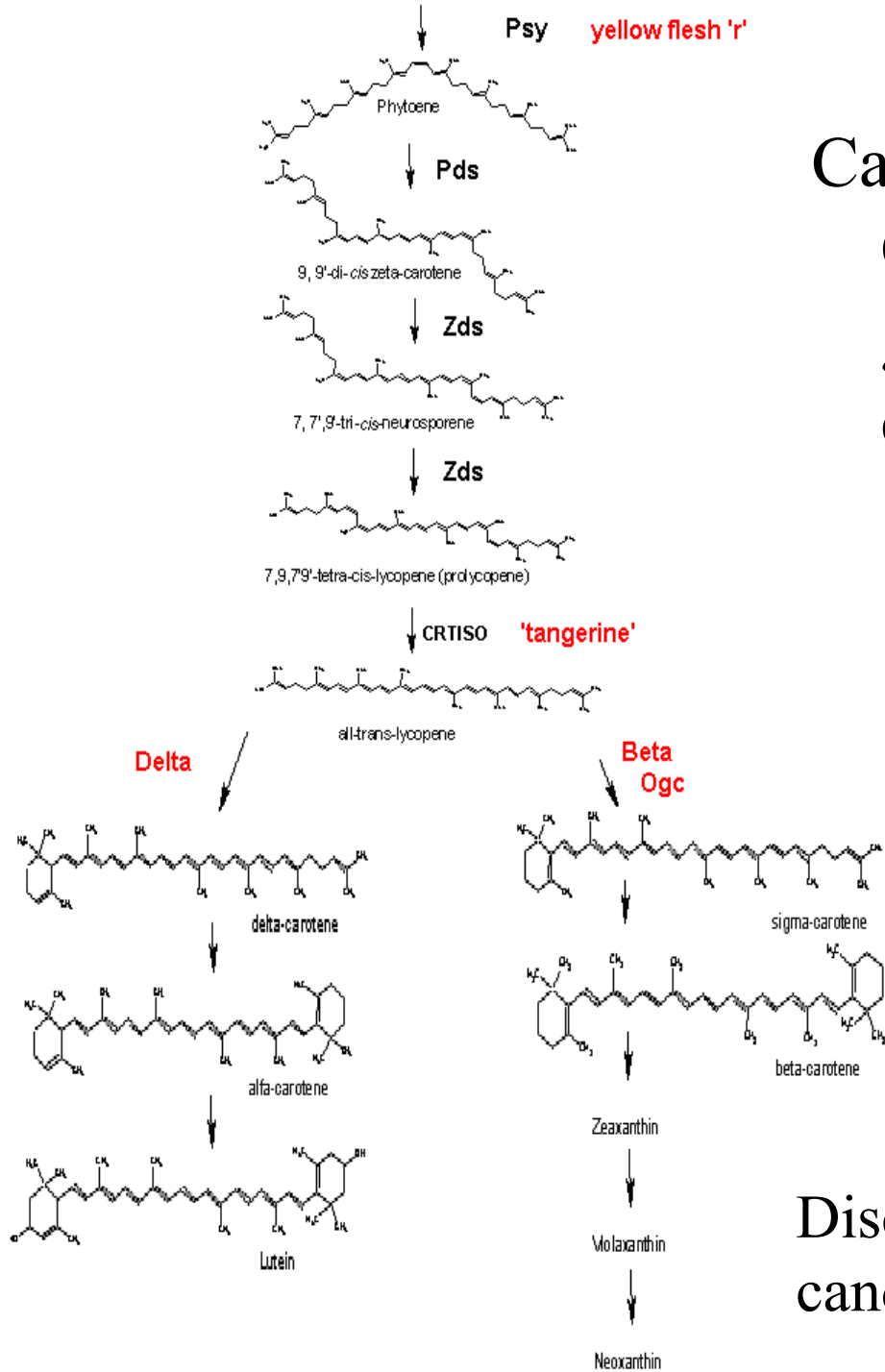
DEFINITION of candidate gene:

A gene suspected of being involved in the expression of a trait.

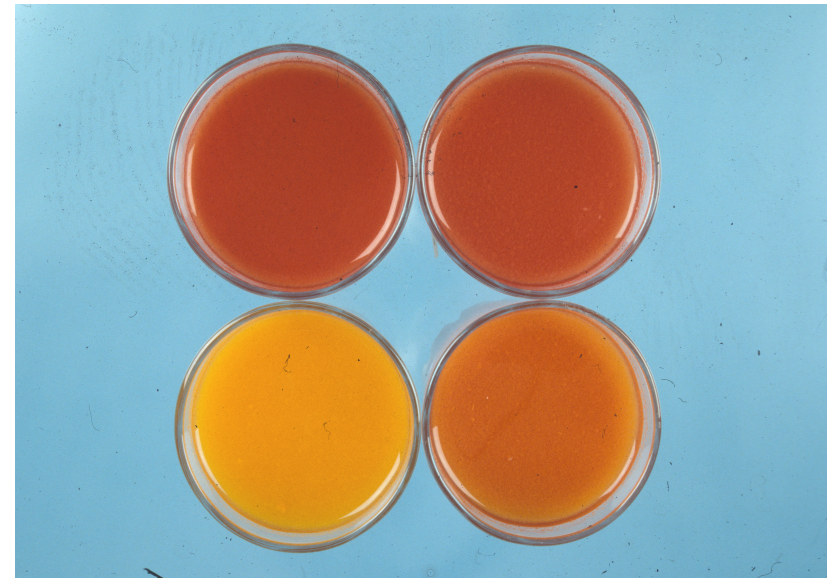
Based on location

Based on presumed function (biochemical pathway, annotation, etc...)

Based on expression patterns



Carotenoid Biosynthesis:
Candidate pathway for
genes that affect color and
color uniformity.



Disclaimer: this is not the only
candidate pathway...

Databases that link pathways to genes

AraCyc Tutorial - Windows Internet Explorer

http://www.arabidopsis.org/help/tutorials/aracyc_intro.jsp

File Edit View Favorites Tools Help

Google aracyc Search

AraCyc Tutorial

Home Help Contact About Us Login/Register

Search Browse Tools Stocks Portals Download Submit News

Home > Help > Tutorials > AraCyc Tutorial

AraCyc Tutorial- Introduction

AraCyc Tutorial

- Searching AraCyc: Pathways, Reactions, Genes and Compounds
- Understanding the AraCyc Detail Pages
- Browsing AraCyc: Pathways, Reactions, Genes and Compounds
- Using the Metabolic Map Overview
- Displaying Gene Expression, Proteomic, Metabolomic and other Data in the Omics Viewer

AraCyc is a metabolic pathway database for *Arabidopsis thaliana* that contains information about both predicted and experimentally determined pathways, reactions, compounds, genes and enzymes. The Omics viewer is a software for displaying large scale data such as microarray gene expression results or proteomic data in the context of biochemical pathways.

This tutorial describes how to query and browse Aracyc. In addition to these pages we have included on the last page, a series of Quicktime movies that demonstrate the features of AraCyc.

See the [AraCyc home page](#) for information about the status of pathways in AraCyc including recently added, or updated pathways that have been manually curated.

AraCyc is a part of the [BioCyc](#) family of metabolic pathway databases and uses the same software.

Internet 100%

Start My Documents Family Desktop 1:29 PM

http://www.arabidopsis.org/help/tutorials/aracyc_intro.jsp

Databases that link pathways to genes

<http://metacyc.org/>

<http://www.plantcyc.org/>

<http://sgn.cornell.edu/tools/solcyc/>

<http://www.arabidopsis.org/biocyc/index.jsp>

http://www.arabidopsis.org/help/tutorials/aracyc_intro.jsp

External Plant Metabolic databases

[CapCyc \(Pepper\) \(C. annuum\)](#)

[CoffeaCyc \(Coffee\) \(C. canephora\)](#)

[SolCyc \(Tomato\) \(S. lycopersicum\)](#)

[NicotianaCyc \(Tobacco\) \(N. tabacum\)](#)

[PetuniaCyc \(Petunia\) \(P. hybrida\)](#)

[PotatoCyc \(Potato\) \(S. tuberosum\)](#)

[SolaCyc \(Eggplant\) \(S. melongena\)](#)


Pathway Tools Query Page - Windows Internet Explorer

http://www.plantcyc.org:1555/

File Edit View Favorites Tools Help

Google aracyc Search Bookmarks Sign In

Pathway Tools Query Page



Plant Metabolic Network

PlantCyc search

About PMN Databases Downloads Tools Useful Sites Submit Data Help Feedba

Pathway Tools Query Page

This form provides several different mechanisms for querying Pathway/Genome Databases.

Select a database: PlantCyc 2 available

• Query Pathway (by name) Submit ?

Error on page. Internet 100%

Start 8 8 E. 5 A. My Documents Family Desktop 1:31 PM

<http://www.plantcyc.org:1555/>

http://www.plantcyc.org:1555/ metacyc

File Edit View Favorites Tools Help

Google aracyc Search + - Bookmarks >> Sign In

Pathway Tools Query Page Page Tools >>

This form provides several different mechanisms for querying Pathway/Genome Databases.

Select a database: PlantCyc 2 available

• **Query** Pathway (by name) carotenoid Submit ?

- To retrieve objects by name, first select the type of object you wish to retrieve, then enter the name of the object and click Submit. All objects containing that name as a substring will be returned. To search for multiple objects in one query, enter multiple names or EC numbers, separated by commas.

• **Browse Ontology:** Pathways Submit ?

- Each database contains classification hierarchies for pathways, for reactions (the enzyme nomenclature system), for compounds, and for genes. Select a classification system to browse.

• **Choose from a list of all** Pathways Submit

• **Cellular Overview [Diagram/Omics Viewer](#)** (not available for MetaCyc)



PlantCyc Pathway: superpathway of carotenoid biosynthesis

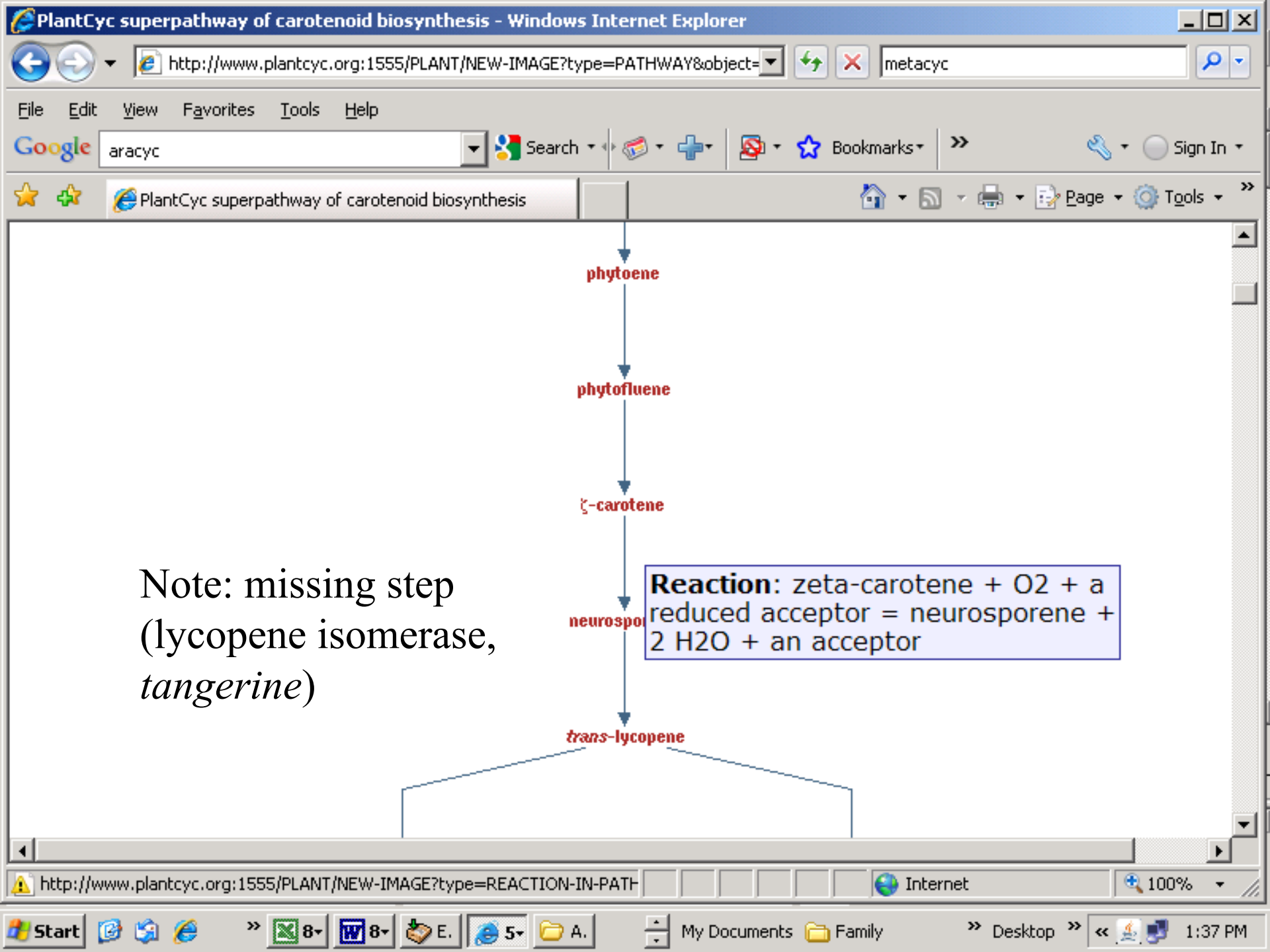


Customize Diagram More Detail Less Detail Cross-Species Comparison
Download Genes BioPAX format

superpathway of geranylgeranyldiphosphate biosynthesis II (plastidic)

2 all-trans-geranyl-geranyl-pp

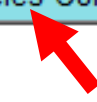
prephytoene diphosphate



PlantCyc Pathway: superpathway of carotenoid biosynthesis



Customize Diagram More Detail Less Detail **Cross-Species Comparison**
Download Genes BioPAX format



superpathway of geranylgeranyldiphosphate biosynthesis II (plastidic)

2 all-*trans*-geranyl-geranyl-pp

prephytoene diphosphate

Browser navigation and search area. Address bar: http://www.plantcyc.org:1555/compare-frame-in-orgs?type=PATHWAY&obje... Search bar: aracyc. Browser menu: File, Edit, View, Favorites, Tools, Help. Search engine: Google. Sign In button.

Object Comparison Organism Selection Page

Note: In addition to reflecting differences in biology of different organisms, these statistics will reflect differences in the levels of curation, data availability, and completeness of the PGDBs for these organisms.

You have requested a cross-species comparison of the following object:

superpathway of carotenoid biosynthesis

To compare the requested object across species, please select from among the following organism databases. The requested object may not be present in all of the organisms listed below. Organisms listed in bold are those in which the object is inferred to be present.

Select two or more organisms:

Arabidopsis thaliana col PlantCyc



Check boxes (Note: MetaCyc has many more choices, but no plants)

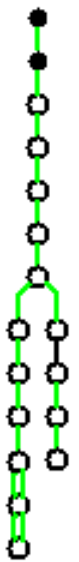
Submit Clear All Select All

Cross-Species Comparison: superpathway of carotenoid biosynthesis

Note: In addition to reflecting differences in biology of different organisms, these statistics will reflect differences in the levels of curation, data availability, and completeness of the PGDBs for these organisms.

This table compares a single pathway across the requested organism databases. The Evidence Glyph is a cartoon of the pathway diagram showing which steps have identified enzymes, which are pathway holes, and which steps are unique to this pathway (a color key is included at the bottom of this page). For each reaction step, we list the identified enzymes and genes. A pathway may not be present in an organism database even if enzymes have been identified for one or more of its reactions, and we indicate if this is the case. Clicking on an organism name will take you to the display page for the pathway in that organism database.

Organism	Evidence Glyph	Enzymes and Genes for superpathway of carotenoid biosynthesis
----------	----------------	---

Organism	Evidence Glyph	Enzymes and Genes for superpathway of carotenoid biosynthesis	
AraCyc col		EC# 2.5.1.32	phytoene synthase: PSY
		EC# 2.5.1.32	phytoene synthase: PSY
		EC# 1.14.99.-	phytoene desaturase: PDS
		EC# 1.14.99.-	phytoene desaturase: PDS
		EC# 1.14.99.30	ζ-carotene desaturase: ZDS
		EC# 1.14.99.30	ζ-carotene desaturase: ZDS
		EC# 5.5.1.-	lycopene ε cyclase: LUT2
		EC# 5.5.1.-	None <input type="button" value="Show page for this object in this organism database"/>
		EC# 1.14.99.-	β-ring hydroxylase: β-OHase 1 carotene beta-ring hydroxylase: LUT5
		EC# 1.14.99.-	ε-ring hydroxylase: LUT1
		EC# 5.5.1.-	lycopene β cyclase: LYC
		EC# 5.5.1.-	lycopene β cyclase: LYC

Cross-Species Comparison - Windows Internet Explorer

http://www.plantcyc.org:1555/compare-frame-in-orgs?orgid=ARA&orgid=PL metacyc

File Edit View Favorites Tools Help

Google aracyc Search Bookmarks Sign In

Cross-Species Comparison

EC# 1.14.99.-	β-ring hydroxylase: β-OHase 1 β-ring hydroxylase: β-OHase 2
EC# 1.14.13.90	antheraxanthin epoxidase / zeaxanthin epoxidase: ABA1 zeaxanthin epoxidase, chloroplast precursor, putative, expressed: LOC_Os04g37619.1 zeaxanthin epoxidase: ABA2 zeaxanthin epoxidase: ABA2
EC# 1.14.13.90	antheraxanthin epoxidase / zeaxanthin epoxidase: ABA1 zeaxanthin epoxidase, chloroplast precursor, putative, expressed: LOC_Os04g37619.1 zeaxanthin epoxidase: ABA2 ←
EC# 1.10.99.3	antheraxanthin deepoxidase / violaxanthin deepoxidase: NPQ1 violaxanthin de-epoxidase family protein, expressed: LOC_Os01g51860.1 violaxanthin de-epoxidase, putative, expressed: LOC_Os04g31040.1 antheraxanthin de-epoxidase / violaxanthin de-epoxidase: VDE1
EC# 1 10 99 3	antheraxanthin deepoxidase / violaxanthin deepoxidase: NPQ1 antheraxanthin de-epoxidase / violaxanthin de-epoxidase:

Internet 100%

Start 7 8 E. 5 A. My Documents Family Desktop 3:41 PM

PlantCyc zeaxanthin epoxidase - Windows Internet Explorer

http://www.plantcyc.org:1555/PLANT/NEW-IMAGE?type=ENZYME&object=N

metacyc

File Edit View Favorites Tools Help

Google aracyc Search Bookmarks Sign In

PlantCyc zeaxanthin epoxidase

PMN Plant Metabolic Network

PlantCyc search

About PMN Databases Downloads Tools Useful Sites Submit Data Help Feedback

PlantCyc Enzyme: zeaxanthin epoxidase

Species: [Nicotiana plumbaginifolia](#)

Summary:

Analysis of the aba2 mutant indicates the zeaxanthin epoxidase is impa of aba2 and expression in E.coli confirmed the zeaxanthin epoxidase e

Citations: [Merin06]

Scroll down page

Capsicum annum sequence retrieved

PlantCyc zeaxanthin epoxidase - Windows Internet Explorer

http://www.plantcyc.org:1555/PLANT/NEW-IMAGE?type=ENZYME&object=N

metacyc

File Edit View Favorites Tools Help

Google aracyc Search Bookmarks Sign In

PlantCyc zeaxanthin epoxidase

Gene-Reaction Schematic: ?

1.14.13.90 Np-ABA2 Nicotiana plumbaginifolia

1.14.13.90 Os-LOC_Os04g37619.1 Oryza sativa

1.14.13.90 Ca-ABA2

Gene: ABA2
Product: zeaxanthin epoxidase
Species: Capsicum annuum

Enzymatic reaction of: zeaxanthin epoxidase

[zeaxanthin + NAD\(P\)H + H⁺ + O₂ <=> antheraxanthin + NAD\(P\)⁺ + H₂O](#)

The reaction direction shown, that is, A + B <=> C + D versus C + D <=> A + B, is in accordance with the Enzyme Commission system.

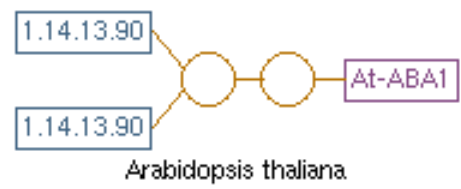
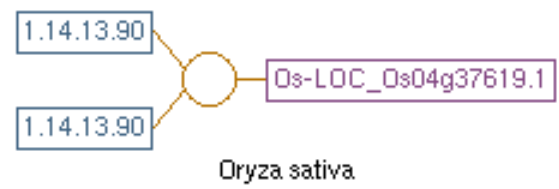
Reversibility of this reaction is unspecified.

In Pathways: [xanthophyll cycle](#)

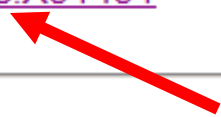
http://www.plantcyc.org:1555/PLANT/NEW-IMAGE?type=GENE&object=ABA2&on

Internet 100%

Start 7- 8- E. A. My Documents Family Desktop 3:44 PM



Unification Links: [Entrez-Nucleotide:X91491](#)



Quick Search

Page generated by SRI International [Pathway Tools version 12.5](#) on Tue Jun 2, 2009.

[Home](#) [Help](#) [Contact](#) [Citing PMN TAIR](#)



Enter Query Sequence

Enter accession number

X91491.1

Or, upload file

Job Title

Align two or more

Choose Search

Database

Organism

Optional

- Genomic plus Transcript**
 - Human genomic plus transcript (Human G+T)
 - Mouse genomic plus transcript (Mouse G+T)
- Other Databases**
 - Nucleotide collection (nr/nt)**
 - Reference mRNA sequences (refseq_rna)
 - Reference genomic sequences (refseq_genomic)
 - NCBI Genomes (chromosome)
 - Expressed sequence tags (est)
 - Non-human, non-mouse ESTs (est_others)
 - Genomic survey sequences (gss)
 - High throughput genomic sequences (HTGS)
 - Patent sequences(pat)
 - Protein Data Bank (pdb)
 - Human ALU repeat elements (alu_repeats)
 - Sequence tagged sites (dbsts)
 - Whole-genome shotgun reads (wgs)
 - Environmental samples (env_nt)

Clear

Query subrange

From

To

Nucleotide collection (nr/nt)



Select database

Enter organism name or id-completions will be suggested

▼ Descriptions

Legend for links to other resources: **U** UniGene **E** GEO **G** Gene **S** Structure **M** Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage
AK327384.1	Solanum lycopersicum cDNA, clone: LEFL2028G11, HTC in fruit	1254	1254	99%
EF581828.2	Solanum lycopersicum cultivar Zhongshu 4 zeaxanthin epoxidase (2	1249	1249	99%
AC215383.2	Solanum lycopersicum chromosome 2 clone C02HBa0104A12, comp	1214	1214	95%
AC215355.2	Solanum lycopersicum chromosome 2 clone C02HBa0012A12, comp	1214	1214	95%
Z83835.1	L.esculentum mRNA for zeaxanthin epoxidase	1037	1037	82%
EU004202.1	Solanum lycopersicum chloroplast zeaxanthin epoxidase precursor,	983	983	78%
DQ206629.1	Solanum tuberosum zeaxanthin epoxidase mRNA, complete cds	887	887	80%
X91491.1	C.annuum mRNA for xanthophyll epoxidase	542	542	70%
X95732.1	N.plumbaginifolia mRNA for zeaxanthin epoxidase	532	532	82%
EU123520.1	Nicotiana tabacum ABA2 mRNA, partial cds	497	497	75%
EF057488.1	Solanum tuberosum zeaxanthin epoxidase mRNA, partial cds	187	187	16%

SGN cvterm GO:0016117 (carotenoid biosynthetic process) - Windows Internet Explorer

http://sgn.cornell.edu/chado/cvterm.pl?cvterm_id=8093

File Edit View Favorites Tools Help

Google Search

http://tomatomap.net/ SGN cvterm GO:0016117...

sol genomics network home | forum | contact | help

search maps sequencing tools sol search

log in

GO:0016117 'carotenoid biosynthetic process'

Genus
specific
database

SGN

Cvterm details

Term id **GO:0016117**

Term name **carotenoid biosynthetic process**

Definition **The chemical reactions and pathways leading to the synthesis of tetraterpenoid compounds in which the isoprenoid units are themselves joined tail-to-tail.**

Comment

Synonyms:
 carotenoid biosynthesis
 carotenoid anabolism
 carotenoid formation
 carotenoid synthesis

SGN cvterm GO:0016117 (carotenoid biosynthetic process) - Windows Internet Explorer

http://sgn.cornell.edu/chado/cvterm.pl?cvterm_id=8093

File Edit View Favorites Tools Help

Google Search

http://tomatomap.net/ SGN cvterm GO:0016117...

- is_a GO:0042440 pigment metabolic process
 - is_a GO:0046148 pigment biosynthetic process
 - is_a GO:0016117 carotenoid biosynthetic process
- is_a GO:0006721 terpenoid metabolic process
 - is_a GO:0016114 terpenoid biosynthetic process
 - is_a GO:0016109 tetraterpenoid biosynthetic process
 - is_a GO:0016117 carotenoid biosynthetic process
- is_a GO:0016108 tetraterpenoid metabolic process
 - is_a GO:0016116 carotenoid metabolic process
 - is_a GO:0016117 carotenoid biosynthetic process
- is_a GO:0016109 tetraterpenoid biosynthetic process
 - is_a GO:0016117 carotenoid biosynthetic process

Phenotype data/QTLs (0)	None
<input checked="" type="checkbox"/> Annotated loci (51)	
Annotated accessions (0)	None
User comments	None [Add comment]

Organized
by
'ontology'

51 annotated loci



Tomato	crtb1	beta-carotene hydroxylase-1
Coffee	lycocyc	lycopene epsilon cyclase
Potato	zep	zeaxanthin epoxidase
Coffee	zep	zeaxanthin epoxidase
Tomato	zep	zeaxanthin epoxidase
Potato	AF321537	lycopene epsilon cyclase
Petunia	AY593974	phytoene desaturase
Tomato	Cnr	Colorless non-ripening
Tomato	DQ335097	phytoene synthase
Tomato	crtiso	carotenoid isomerase
Tomato	crtle1	LELYCOCYC lycopene epsilon-cyclase
Tomato	gtom5	LEYFGTOM5 GTOM5 mutant phytoene synthase
Coffee	lycocyc	lycopene epsilon cyclase
Toma		obsolete-phytoene synthase 1
Potal		phytoene desaturase
Coffe		phytoene desaturase
Tobacco	pds	phytoene desaturase
Tomato	pds1	phytoene desaturase
Pepper	pds1	phytoene desaturase 1
Coffee	pds	phytoene synthase

Candidates identified in other databases are here

Information missing from other databases is here...

[http://sgn.cornell.edu/phenome/locus_display.pl?locus_id=4476](#)

[File](#) [Edit](#) [View](#) [Favorites](#) [Tools](#) [Help](#)

[Google](#)
[Search](#) [+](#) [-](#) [Bookmarks](#) [Sign In](#)

[http://tomatomap.net/](#) [SGN Tomato locus: zeax...](#)

[Home](#) [RSS](#) [Print](#) [Page](#) [Tools](#)



sol genomics network

[home](#) | [forum](#) | [contact](#) | [help](#)
[search](#)
[maps](#)
[sequencing](#)
[tools](#)

[log in](#)

Tomato 'zeaxanthin epoxidase'

Locus details

[Download GMOD XML](#) | [Note to Editors](#) | [Annotation guidelines](#)
[\[New\]](#) [\[Edit\]](#) [\[Delete\]](#)
Locus name **zeaxanthin epoxidase**Symbol **zep**Gene activity **zeaxanthin epoxidase**
 Description **Enzyme that catalyzes the conversion of zeaxanthin to violaxanthin.
Mutants are wilted due to abscisic acid deficiency**
Chromosome **2**

Arm

2



Tomato-EXPEN 2000

Locus synonyms 3: hp3 LeZEP ZE [\[Add/Remove\]](#)
Locus editors: [Andrew Thomson](#) [Navot Galnaz](#)

Comment on the databases:

Information is not always complete/up to date.

Display is not always optimal, and several steps may be needed to go from pathway > gene > potential marker.

Sequence data has error associated with it. eSNPs are not the same as validated markers.

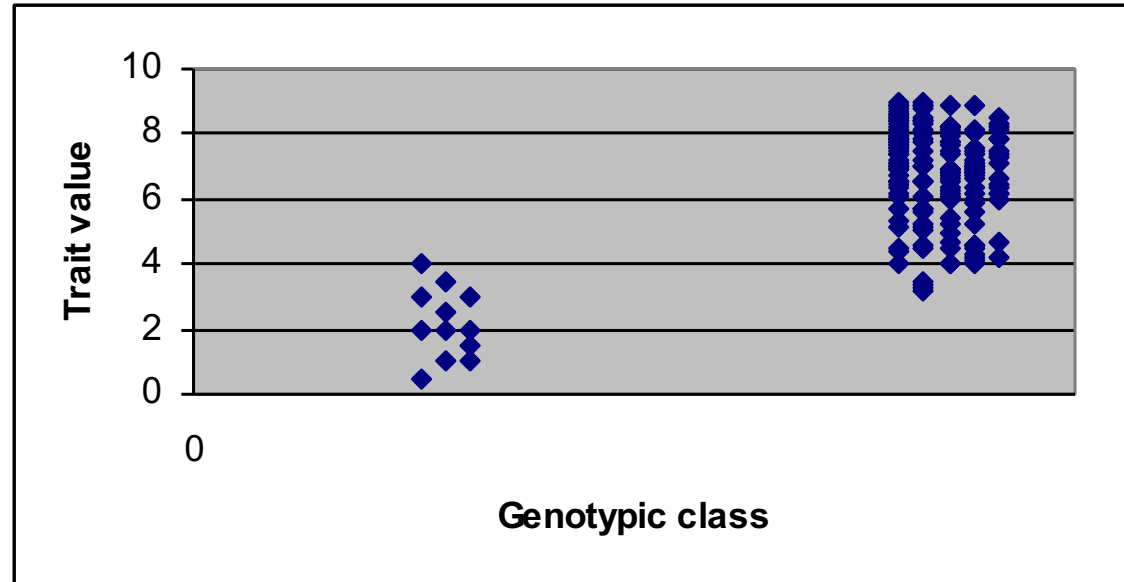
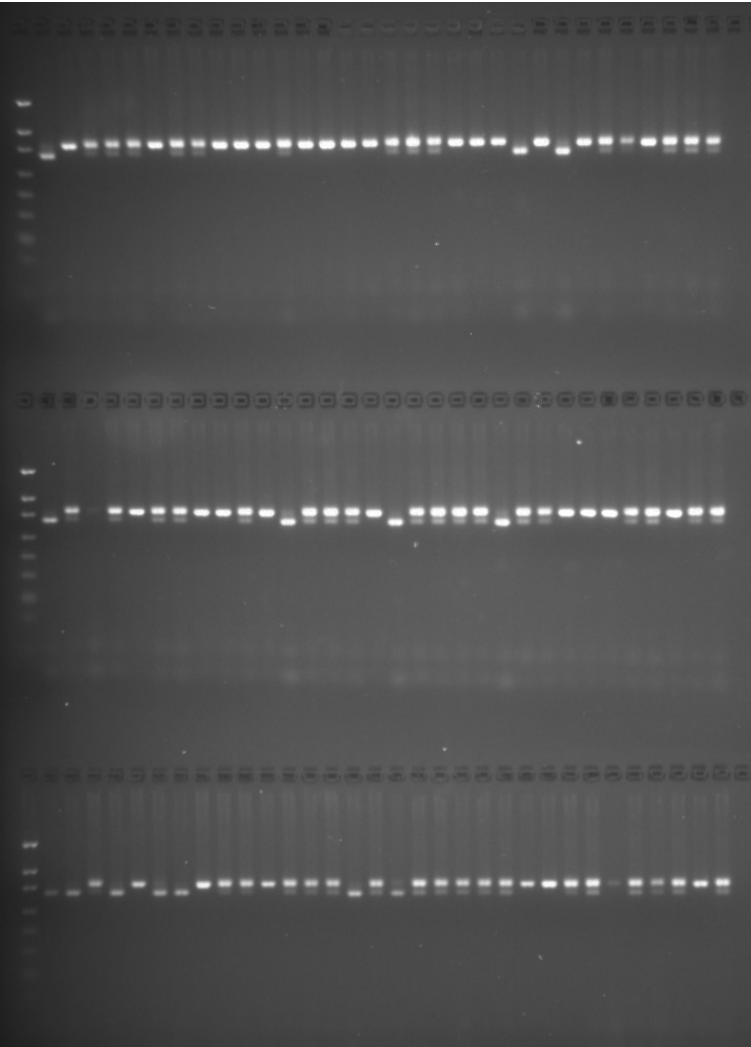
```
*****  
T252213.TA496 CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
T278626.TA496 CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
T554112.TA496 CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
ST307657.RicG CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
ST281082.RicG CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
ST308731.RicG CCCACCTCCAGCACCCCTAATTGATCCGGAAGAGTTAGGAAAA  
ruler 110.....120.....130.....140.....150
```

```
***** ** *****  
T252213.TA496 GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
T278626.TA496 GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
T554112.TA496 GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
ST307657.RicG GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
ST281082.RicG GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
ST308731.RicG GTTAAACCCCGCTGTTACATTCGGCCCTTTCCTAGCCCGAAAAGTATCGT  
ruler .....360.....370.....380.....390.....400
```

Germplasm data may also have error (e.g. PI 128216)

There is a wealth of information organized and available.

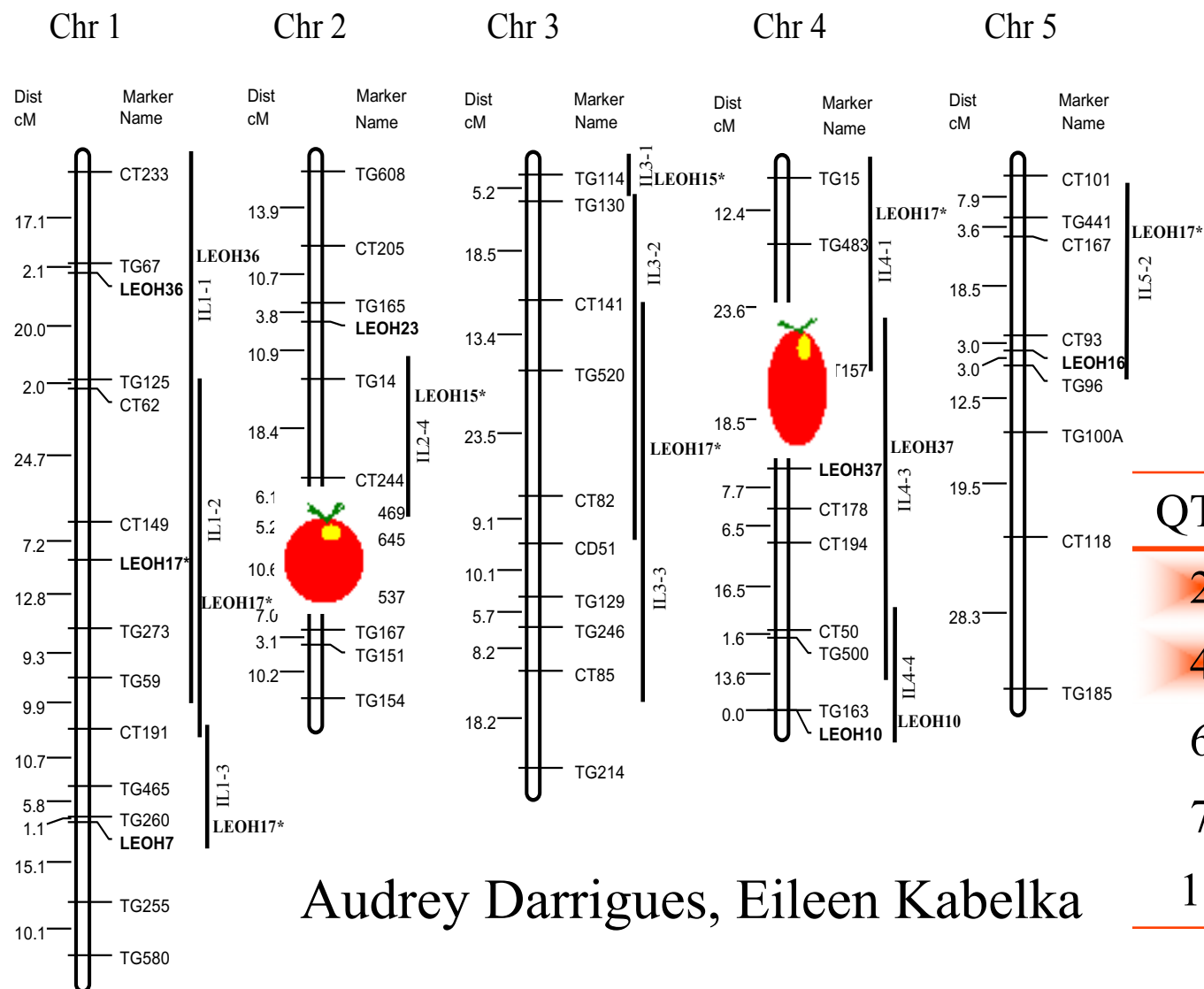
Establishing linkage between a “marker” and “trait”



Analysis is conceptually simple (t-test; ANOVA; regression)

In practice, more complex statistical models are adopted such that missing data is tolerated, significance level is accurately estimated, etc...

Example: QTL for color uniformity in elite crosses



QTL	Trait	Origin
2	L, YSD	<i>S. lyc.</i>
4	YSD	<i>S. lyc.</i>
6	L, Hue	og ^c
7	L, Hue	<i>S. hab.</i>
11	L, Hue	<i>S. lyc.</i>

Audrey Darrigues, Eileen Kabelka

How much phenotypic variation is explained by a QTL?

Analogous to the concept of “heritability” (variation due to genetics/total variation).

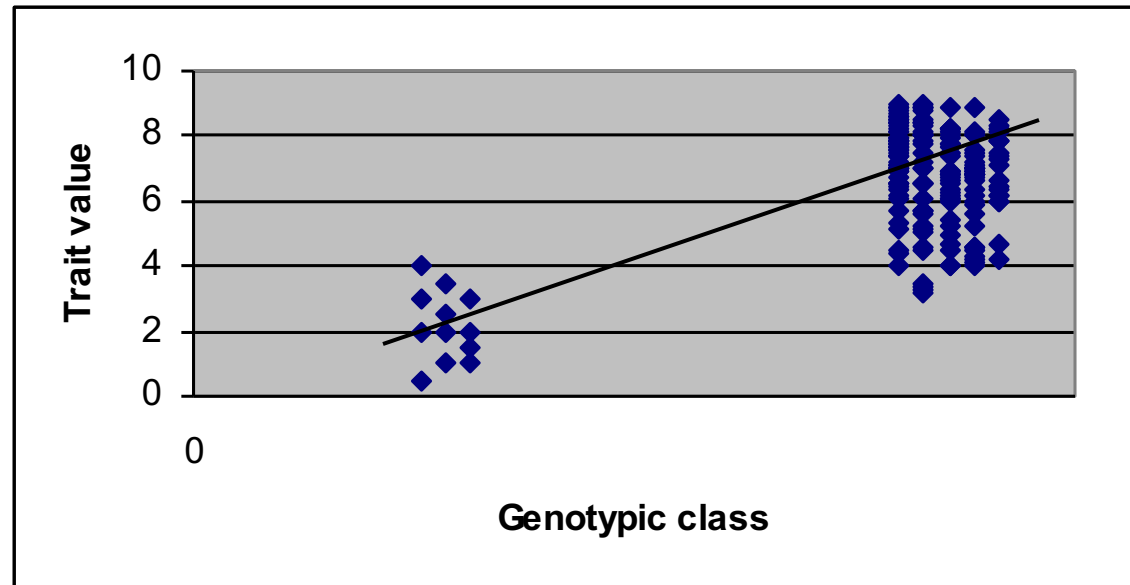
Two approaches
to estimate:

From ANOVA:

$$V_{\text{marker}}/V_{\text{total}}$$

From regression:

$$r^2$$



r^2 for marker is dependent on two things: the strength of the linked gene and linkage distance. Therefore r^2 can be increased by identifying markers closer to the QTL or gene of interest.

Source	DF	Expected MS
Genotypes	N-1	$\sigma^2 + b\sigma^2(G)$
Marker	1	$\sigma^2 + b[\sigma^2(G_{QTL}) + 4r(1-r)g^2] + bc(1 - 2r)^2g^2$
Gen(marker)	N-2	$\sigma^2 + b[\sigma^2(G_{QTL}) + 4r(1-r)g^2]$
Error	N(b-1)	σ^2

b is the number of replicates

r is the recombination separating the marker from the QTL

c is a coefficient related to the population size

$$c = N - (n_1^2 + n_2^2)/N$$

($n_1 + n_2 = 1$; representing the number in each marker class)

g is the genetic effect (in BC pop's additive and dominance effects are confounded).

$\sigma^2(G_{QTL})$ is the part of the error variance that cannot be explained by the QTL.

$$V_{\text{marker}}/V_{\text{total}}$$

F-test for significance is Marker/Gen(marker)

$$= bc(1 - 2r)^2g^2$$

Significance depends on population size, recombination, the strength of the genetic effect relative to the error variance and the part of the error variance that cannot be explained by the QTL

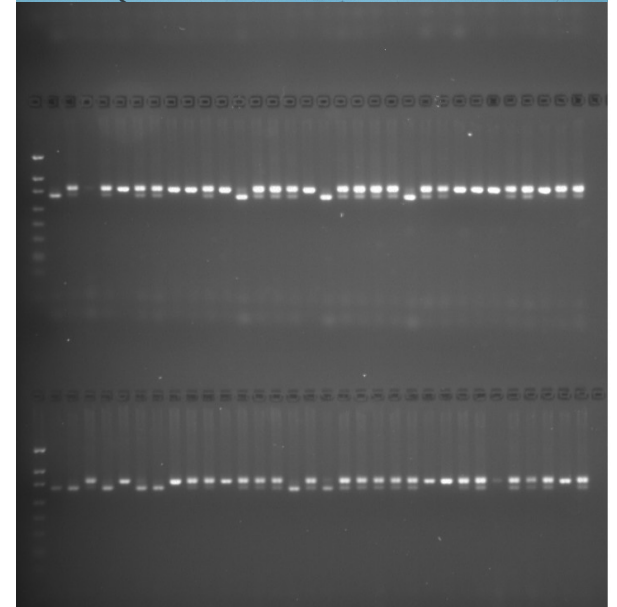
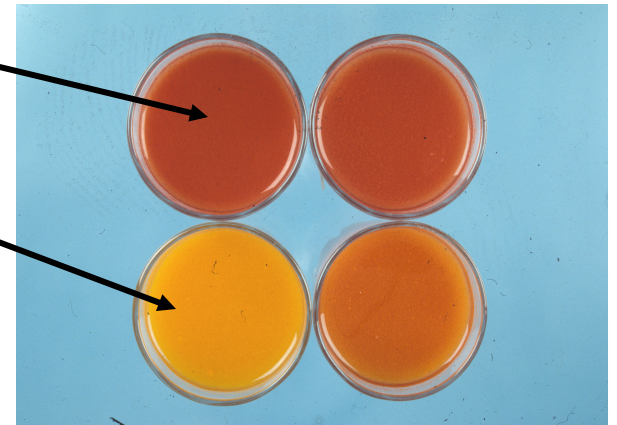
The proportion of variance explained by polymorphic marker on chromosome 2

<u>Trait</u>	<u>V_m/V_T</u>
L	0.25
Hue	0.15
L uniformity	0.28
Hue uniformity	0.32

Marker assisted selection is a form of indirect selection

Selection for t^v

Selection for dg



MAS is a form of indirect selection:

The relative efficiency of selection can be expressed as $= r(\text{gen}) \{H_{\text{indirect}}/H_{\text{trait}}\}$

H_{indirect} marker or indirect phenotype (proportion of accurate genotypes when scoring a marker)

H_{trait} Heritability of the trait

$r_{(\text{gen})}$ Genetic covariance for traits.

$r_{(\text{gen})}$ for a marker and a phenotypic trait = $\text{Var}(\text{marker})/\text{Var}(\text{Total})$ or $r_{(\text{marker-trait})}$

Comparison of direct selection with indirect selection (MAS).

Line-mean heritability (H) for color							
Trait	H		H		Indirect Selection		
	plant	SE	line	SE		Prop Vp	SE
BRIX	0.14	0.09	0.40	0.26			
Color-L	0.11	0.04	0.57	0.22	L-MAS	0.25	0.15
Color-Hue	0.07	0.04	0.39	0.23	Hue-MAS	0.15	0.06
Color unif. -L	0.14	0.11	0.63	0.25	Ldiff-MAS	0.28	0.16
Color unif. -Hue	0.13	0.10	0.64	0.23	Hdiff-MAS	0.32	0.14

Relative efficiency of selection:

$$r_{(\text{gen})} \times \{H_i/H_d\}$$

Line performance over locations >

MAS > Single plant

$$r_{(\text{gen})} \times \{H_i/H_d\}$$

For Hue uniformity, when applied to early generations:

$$= 0.32/0.13 = 2.46$$

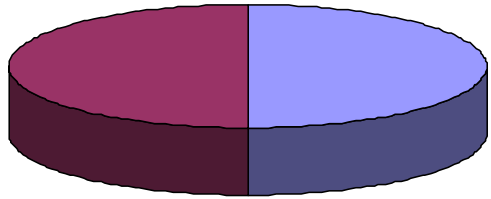
later generations:

$$= 0.32/0.64 = 0.5$$

The best time to use genetic markers : early generation selection

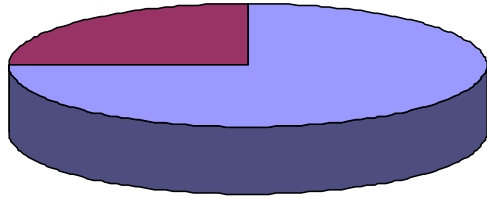
Accelerating Backcross Selection

F1 50:50

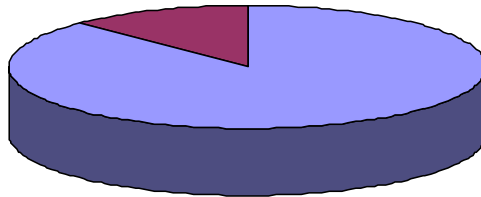


Expected proportion of
Recurrent Parent (RP)
genome in BC progeny

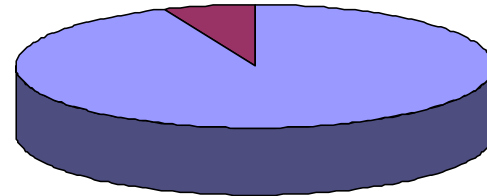
BC1 75:25



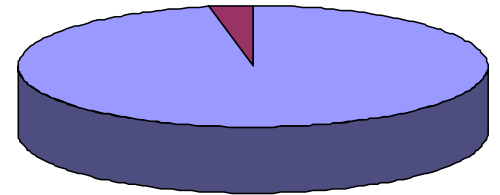
BC2 87.5:12.5



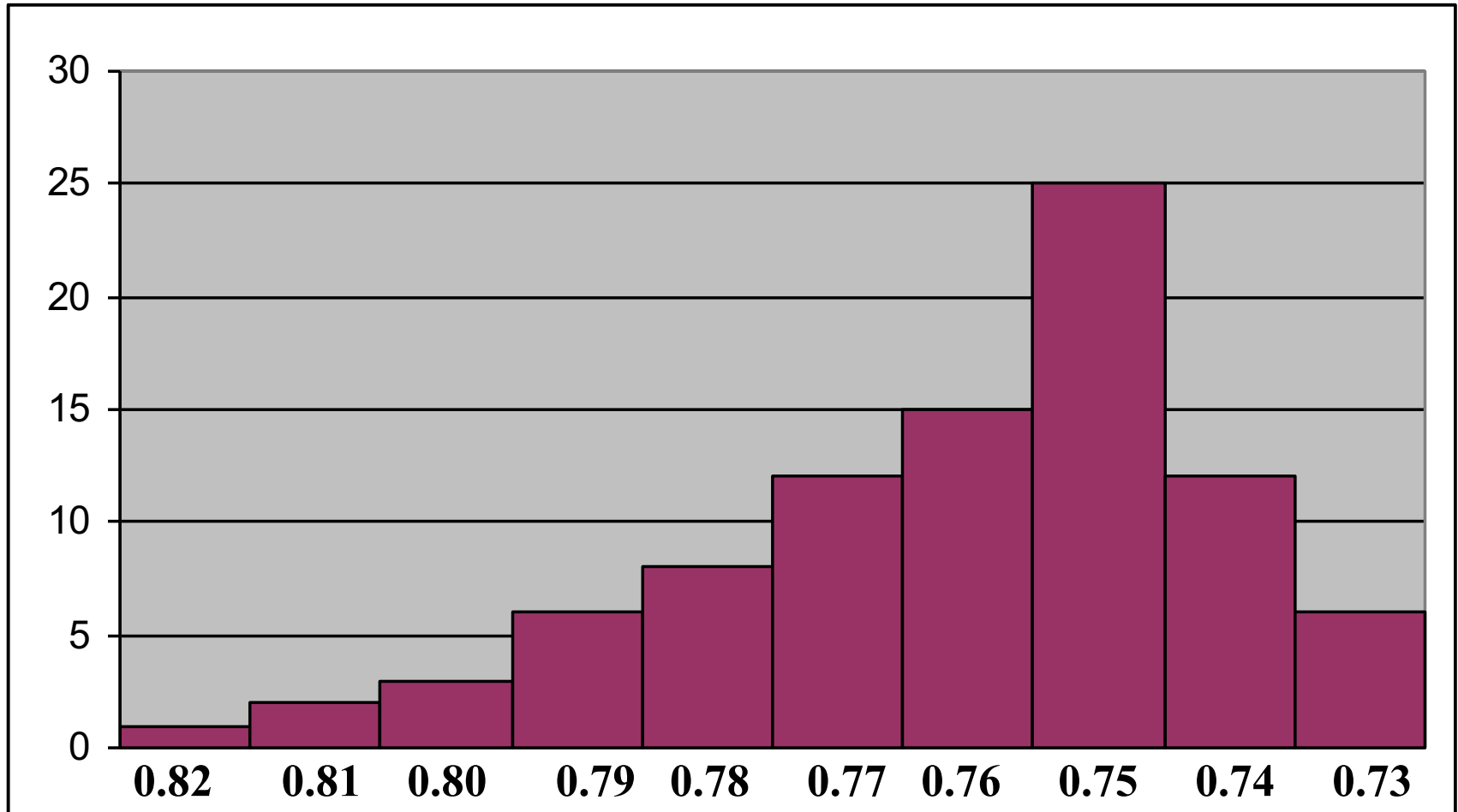
BC3 93.75:6.25



BC4 96.875:3.125



Frequency distribution: Proportion of recurrent parent genome in BC1 population



Two-stage selection					
Select for target allele		Select for RP genome at unlinked markers			
Three-stage selection					
Select for target allele		Select for RP recombinants at flanking markers		Select for RP genome at unlinked markers	
Four-stage selection					
Select for target allele		Select for RP recombinants at flanking markers		Select for RP genome on carrier chromosome	Select for RP genome at unlinked markers

References:

Frisch, M., M. Bohn, and A.E. Melchinger. 1999. Comparison of Selection Strategies for Marker-Assisted Backcrossing of a Gene. *Crop Science* 39: 1295-1301.

For effective background selection we need:

Markers for our target locus (C > T SNP for Zep)

Markers on the target chromosome (Chrom. 2)

Markers unlinked to the target chromosome (~2 per chromosome arm)

Progeny needed for Background Selection During MAS

Q10 of RP genome in percent								
	Population Size							
	20	40	60	80	100	125	150	200
Two-Stage								
BC1	76.7	78.7	79.7	80.3	80.7	81.3	81.7	82.2
BC2	90.3	91.9	92.8	93.3	93.6	93.9	94.0	94.6
BC3	95.8	96.2	97.1	97.3	97.4	97.5	97.6	97.8
Three-Stage								
BC1	71.2	72.7	73.4	73.6	73.3	73.2	72.8	72.2
BC2	86.1	87.2	88.5	89.3	90.2	90.7	91.3	91.8
BC3	94.4	95.7	96.5	96.9	97.2	97.3	97.5	97.6

Q10 indicates a 90% probability of success

From Frisch et al., 1999.

Important point: we can save ~1 generation in a BC program through background selection

Take home messages:

The best time to use genetic markers : early generation selection

Selecting for recurrent parent genome can save >1 BC

Restructuring breeding programs to maximize use of MAS may include:

- 1) Increasing genotypic replication (population size) at the expense of plot replication (consider augmented designs).**
- 2) Collecting and archiving objective data.**
- 3) Restructuring selection pipelines to make optimal use of genotyping reagents.**



References:

Brewer et al., 2006. Plant Physiology 141: 15-25

Darrigues et al., 2008. JASHS, 133: 579-586

Frisch, et al., 1999. Crop Science 39: 1295-1301.

Knapp and Bridges, 1990. Genetics 126: 769-777.