

MG

Understanding the gene pool structure of cultivated potatoes for association mapping of trait important to breeding.

MB

Gene pool structure of cultivated potatoes assessed by SSR marker analyses

Jorge Núñez¹, María del Rosario Herrera¹, Guillermo Trujillo¹, Frank Guzmán¹, David Spooner², Marc Ghislain¹

1 Applied Biotechnology laboratory, International Potato Center, PO Box 1558, Lima12, Peru.

2 USDA, ARS, Vegetable Crops Research Unit, Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison WI 53706-1590, USA.

2nd Solanaceae Genome Workshop 2005, Ischia, Italy



Generation Challenge Program and CIP

The International Potato Center CIP



Background: gene pool structure

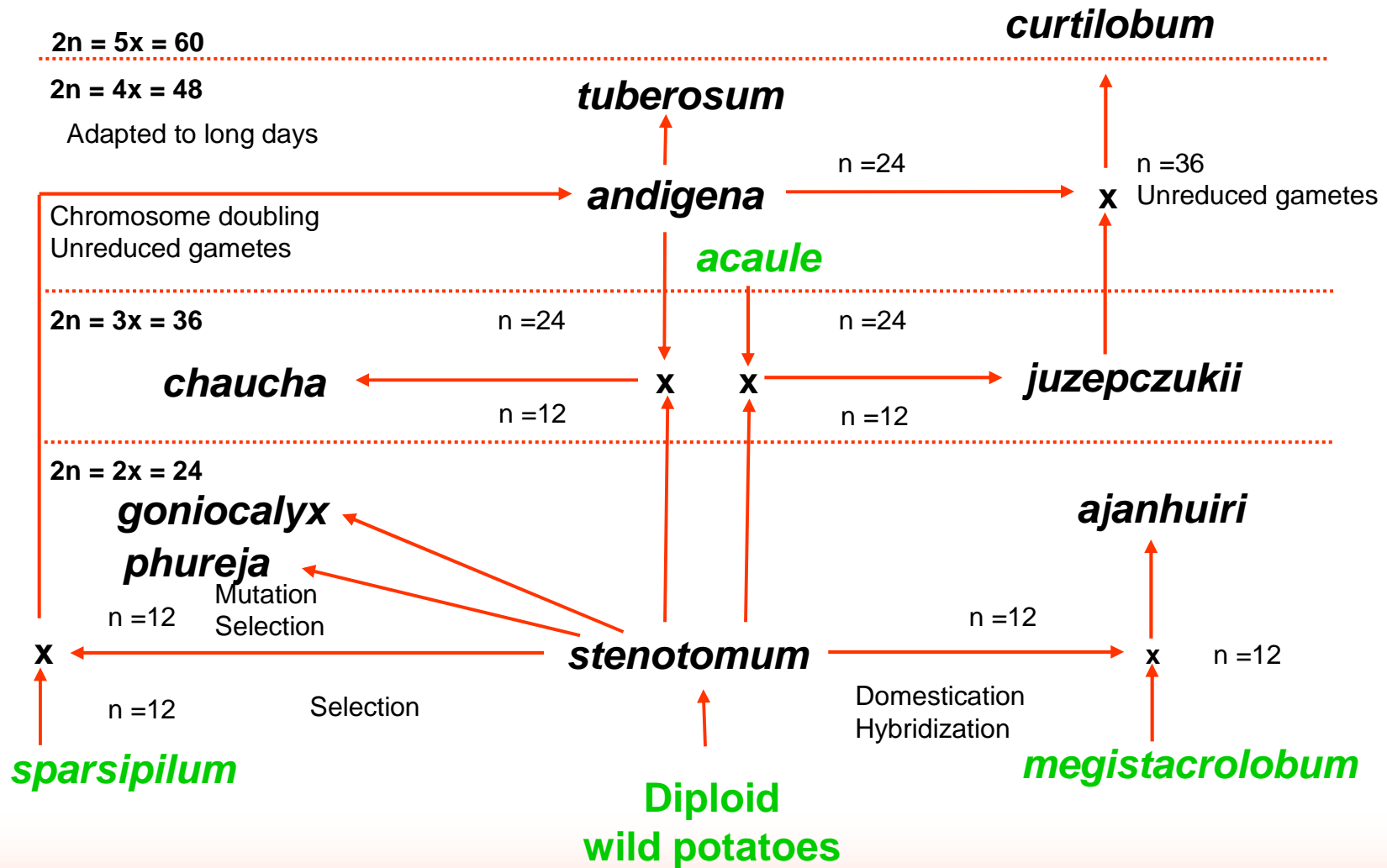
Different views on cultivated potato taxonomy:

Solanaceae, *Solanum*, sect. *Petota*, subsect. *Potatoe*, Series *Tuberosa*:

- Either 7 (9) species: *S. tuberosum* 2 sub species *andigenum* and *tuberosum* (or *S. andigenum* and *S. tuberosum*) , *S. stenotomum* 2 sub species *stenotomum* and *goniocalyx* (or *S. stenotomum* and *S. goniocalyx*), *S. phureja*, *S. chaucha*, *S. ajanhuiri*, *S. juzepczukii*, *S. curtilobum* according to Hawkes, 1990 (Ochoa, 1999);
- Or 1 species *S. tuberosum* with 8 taxonomic groups: *Andigenum*, *Chilotanum*, *Stenotomum*, *Phureja*, *Chaucha*, *Ajanhuiri*, *Juzepczukii*, *Curtilobum* according to Huamán & Spooner, 2002.



Background: inter-specific hybridization



Adapted from Hawkes, 1994

The International Potato Center CIP



Objectives of the SSR characterization of the cultivated potato germplasm

1. Increase the number of SSR markers for potato germplasm analyzes, determine genetic location of the new SSR markers on potato genetic maps, and assess these as taxonomic tool.
2. Analyze the gene pool structure of the cultivated potato for future association mapping studies by genotyping a sample of 1000 accessions with 50 SSR markers.



Materials: a Composite Genotyping Set (CGS)

	holdings	with (x) SSR	of the CGS
Landraces			
andigenum (adg)	3322	537 (10)	250
goniocalix (gon)	81	78 (11)	41
chilotanum (chi)	157	120 (18)	30
stenotomum (stn)	260	230 (11)	51
ajanhuiri (ajh)	22	22 (53)	22
chaucha (cha)	151	151 (53)	151
curtilobum (cur)	20	20 (53)	21
juzepczuki (juz)	35	33 (53)	37
phureja (phu)	203	104 (53)	104
unclassified (sol)	94	42 (53)	42
modern cultivars			69
Advanced cultivars	284		56
Breeding lines	1710		123
Mapping populations	236		85
Total	6575	1337	1082

Landraces were sampled using the 9 species classification; *adg*, *gon*, *stn*, *tbr* (*chi*) were sampled for maximizing SSR allele capture using at least 10 SSR markers (95%); *phu* was included entirely; minor groups were sampled randomly.

Potato SSR marker genetic maps

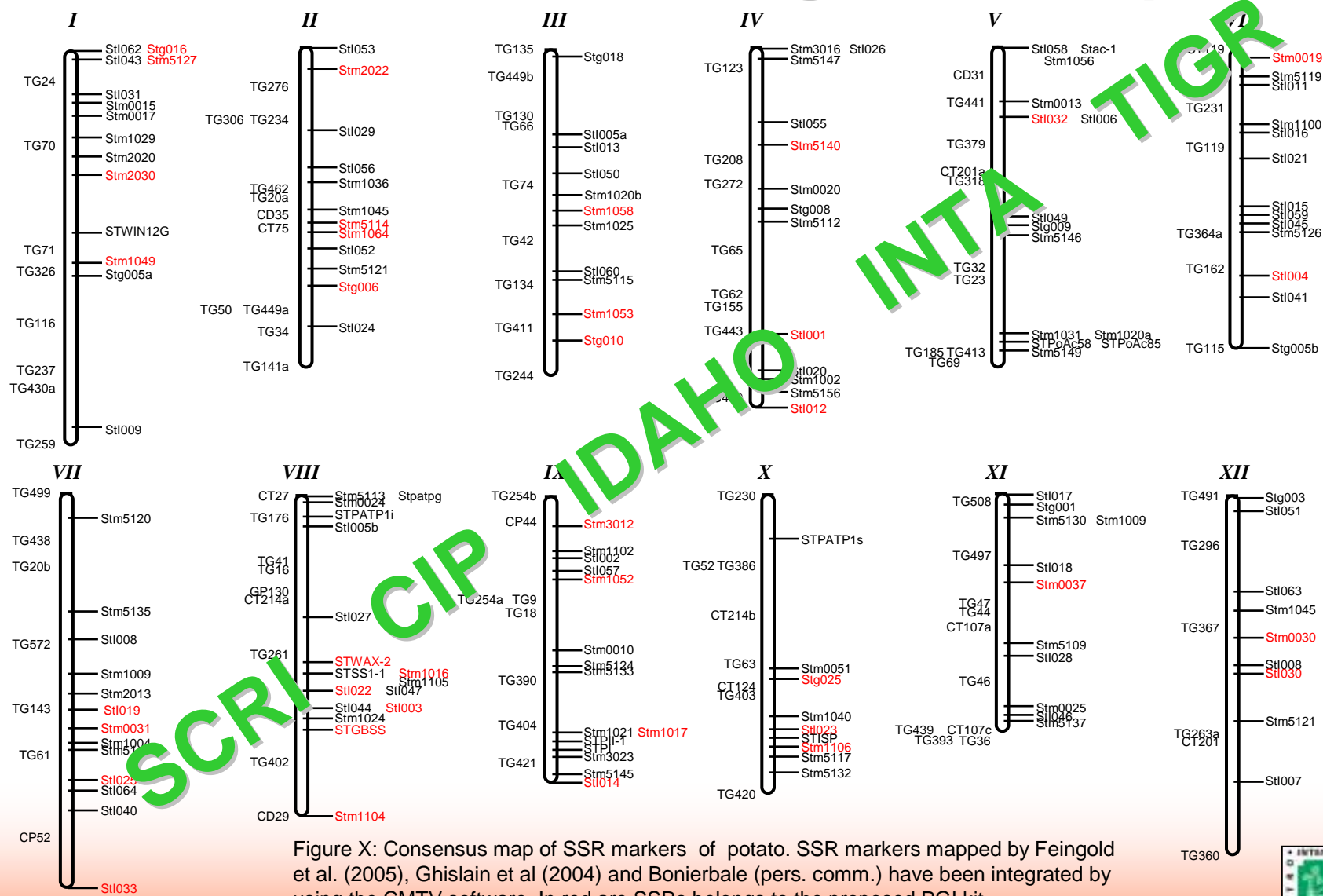


Figure X: Consensus map of SSR markers of potato. SSR markers mapped by Feingold et al. (2005), Ghislain et al (2004) and Bonierbale (pers. comm.) have been integrated by using the CMTV software. In red are SSRs belongs to the proposed PGI kit.



SSR data editing

1. Out of the 53 SSR markers, 7 were not considered for the following reasons:
 - 5 SSR (STM3023, STM2013, STG0018, STG0020 and STI0034) appeared to be multi loci (exceeding allele number/ploidy) which causes precise allele identification impossible;
 - 2 SSR markers (STM1031 and STPoAC58) displayed null alleles;
2. LI-COR generated data allowed to fully identify alleles (1 band for 2x indicates homozygous locus; 2, 3, or 4 bands in 3x, 4x, 5x were translated into alleles by measuring relative band intensity);
3. No binning of SSR data were performed (no allele sequence evidence to justify it – to do).

SSR marker characterization

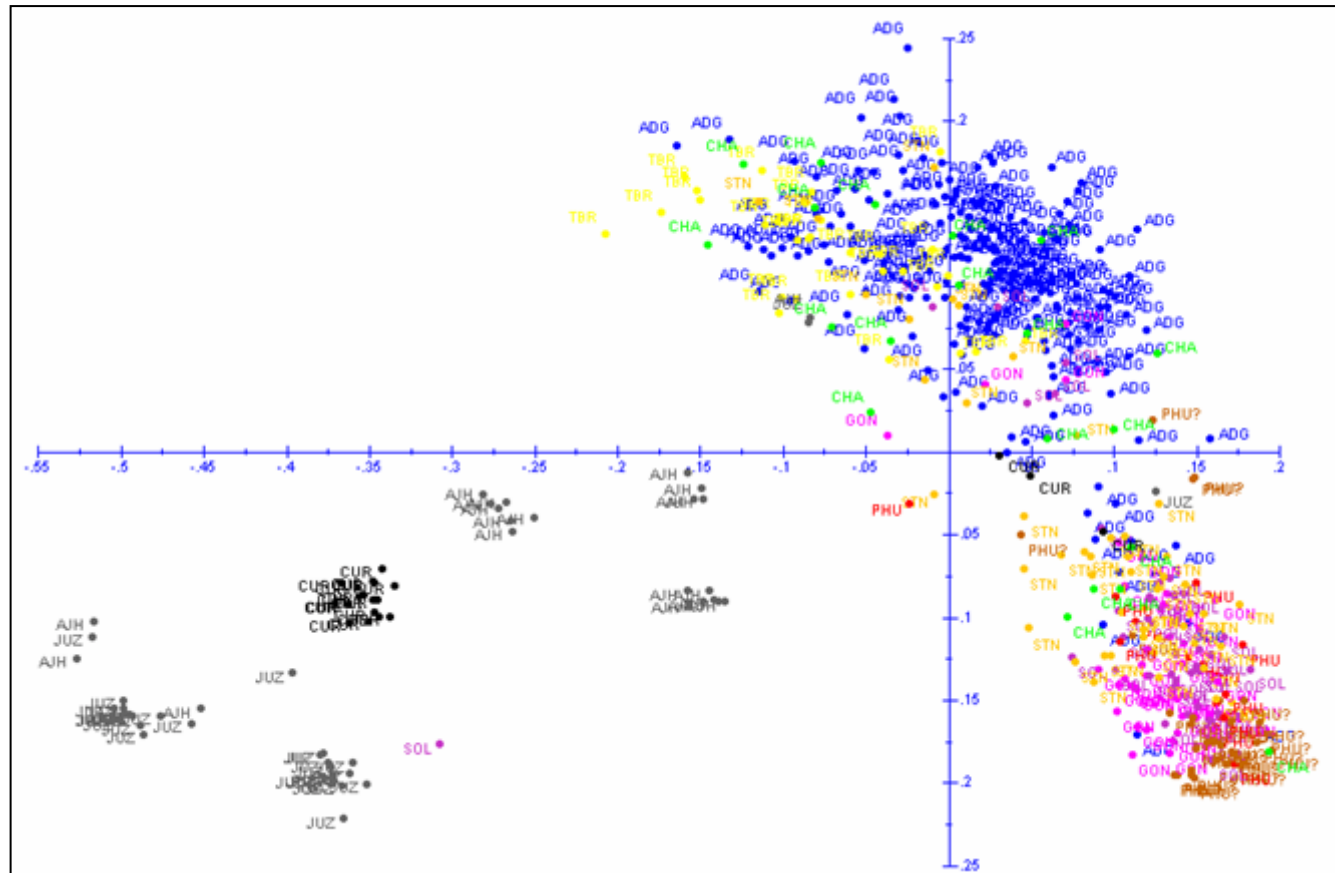
Using the 46 SSR markers and the original passport data for taxonomic identification:

Species	AJH	PHUcore	PHU?	PHU	GON	SOL	STN	CHA	JUZ	ADG	TBR	CUR	CGS
Ploidy	2x	2x	2x	2x	2x	~2x	2x	3x	3x	4x	4x	5x	2x-5x
PIC min	0.219	0.000	0.000	0.000	0.000	0.045	0.000	0.095	0.057	0.180	0.361	0.172	0.296
PIC max	0.772	0.830	0.696	0.755	0.803	0.804	0.856	0.871	0.750	0.859	0.853	0.823	0.892
% Ploidy discordance	77.3	15.4	18.9	18.0	31.7	59.5	49.0	60.0	9.1	2.2	0.0	15.0	20.9
Alleles min. by SSR	2	1	1	1	1	2	1	2	2	2	2	2	2
Alleles max. by SSR	8	8	8	10	9	10	11	14	7	19	12	8	22
Unique alleles (Freq. Abs. >0)	3	1	1	2	2	0	3	2	0	90	29	1	-
Unique alleles (Freq. Abs. >1)	2	0	0	0	0	0	0	0	0	53	17	1	-
% Rare alleles (Freq.<0.01)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	28.6	0.0	0.0	31.9
% Rare alleles (Freq.<0.05)	21.5	8.7	34.2	41.6	47.1	42.4	40.5	33.4	37.1	51.8	29.6	27.4	56.7
Total of alleles	195	172	161	202	225	255	289	302	202	440	304	190	504
Total of genotypes	22	13	37	50	41	42	49	20	33	224	30	20	531

- SSR markers appeared to be good indicators of ploidy. In all cases verified by chromosome counting, an exceeding allele number for 2x accession was indicative of mis-classification;
- ADG presented the maximum allele diversity.
- Very few unique alleles, however, allele distribution is uneven among the various taxonomic groups with {ajh, juz, cur}, {adg, chi, stn} being more alike and while phu displayed distinct allele distribution from the previous ones.

Gene pool structure analyses I

Factorial analysis
(DARwin 4.0)
on 531
landraces X
46 SSR
markers: All
ploidy
confounded.



Observations:

Evidences for structure in cultivated potato:

4x adg+tbr(chi) – 2x phu+stn+gon+sol – 2x ajh – 3x juz – 5x cur

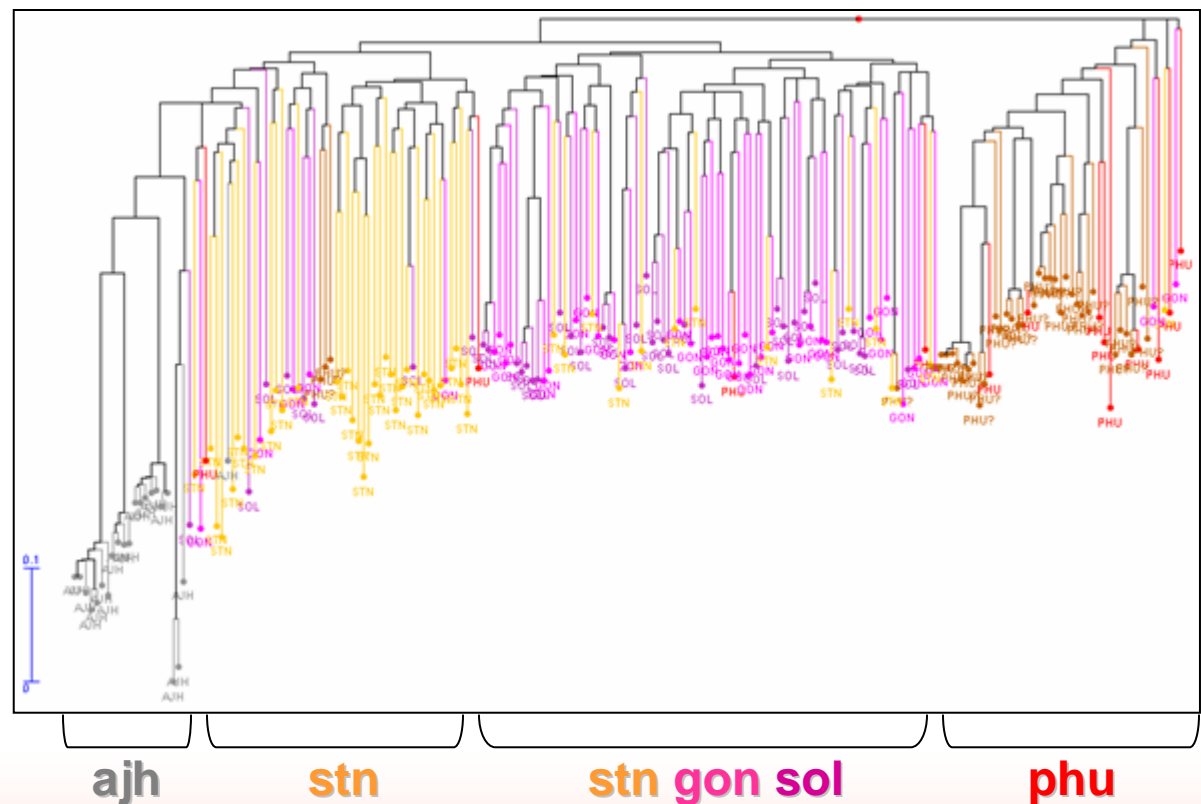
Gene pool structure analyses IIa

Cluster analyses (DARwin 4.0) on 531 landraces X 46 SSR markers:

2x material:

- AJH
- PHU (○ PHU?)
- STN
- GON

Observations:
Good separation into 4 diploid cultivar groups.



Gene pool structure analyses IIb

Cluster analyses (DARwin 4.0) on 531 landraces X 46 SSR markers:

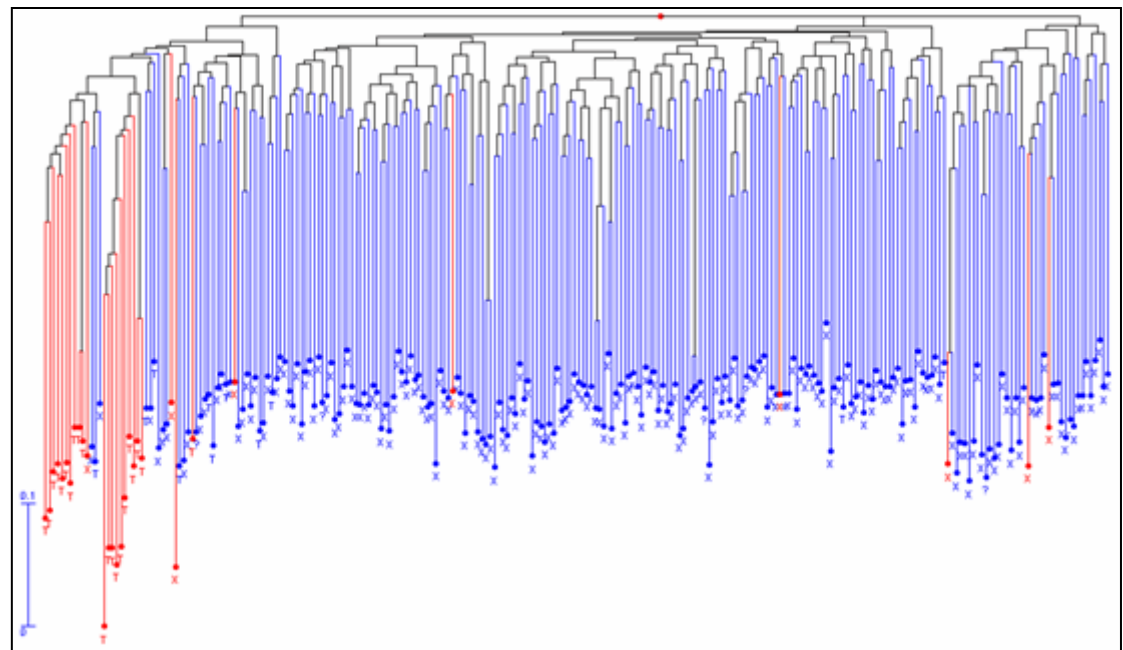
4x material:

○ ADG

○ TBR

Observations:

Clear grouping of landraces of the Chilotanum group with landraces of Andigenum group having the cp marker for tbr (chi).



tbr (chi)

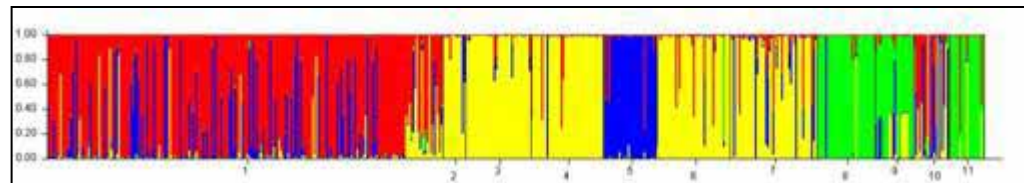
adg

T: cp marker for tbr (chi)

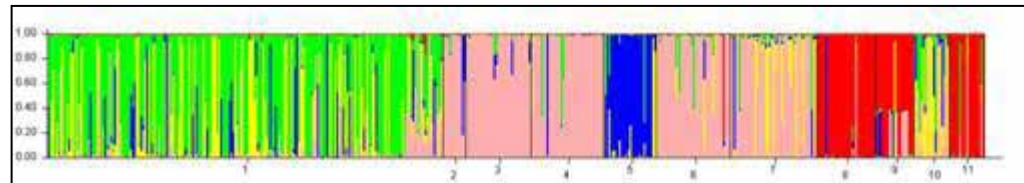
Gene pool structure analyses III

Population structure analyses (Structure 2.1) on 531 landraces
X 46 SSR markers: 100,000 Burn-in period + 200,000 reps
{Assumptions: ancestry admixture model; correlated allele frequency}.

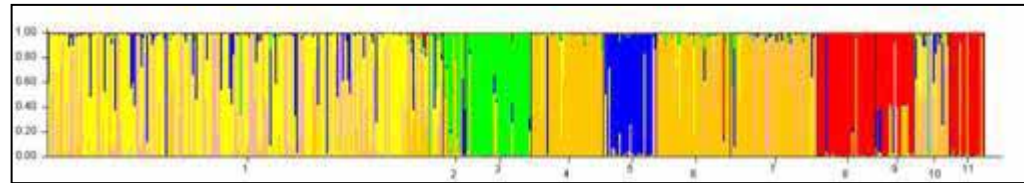
k=4



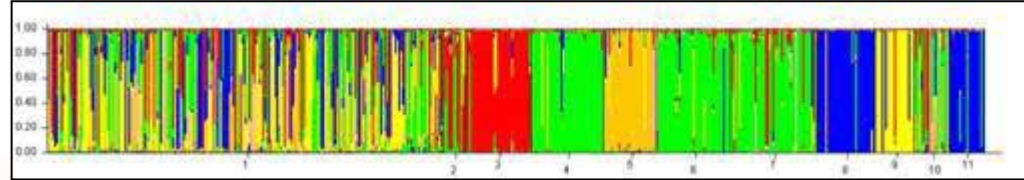
k=5



k=6



k=10



Observations:
***Clear sub-
populations,
some distinct
from cultivar-
groups.***

Adg

Phu

Gon

Tbr

Sol

Stn

Juz

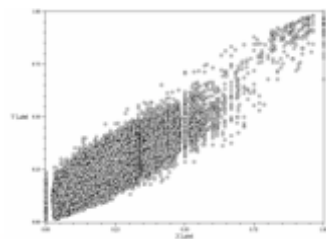
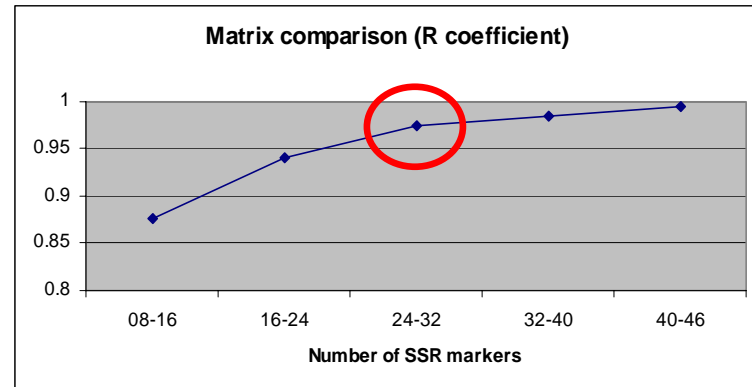
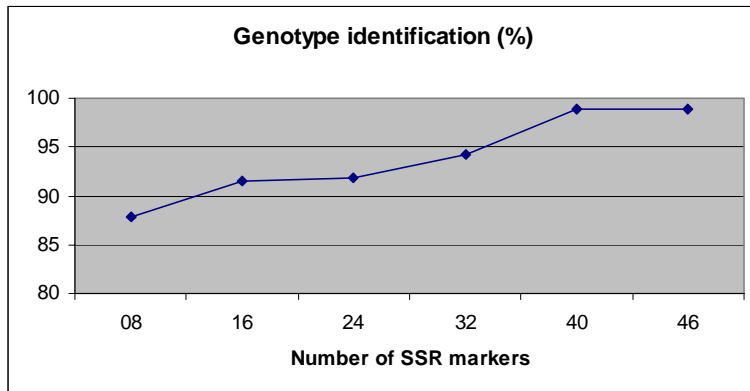
Ajh

Cha

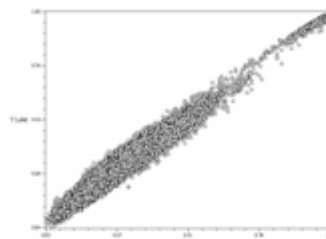
Cur

Recommendation for a new PGI kit

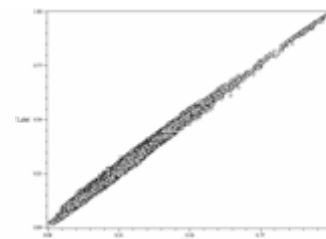
The 46 SSR were evaluated in an additive way , each 8 SSR starting with the most polymorphic markers (highest PIC), and corresponding similarity matrix were compared.



8 vs 16



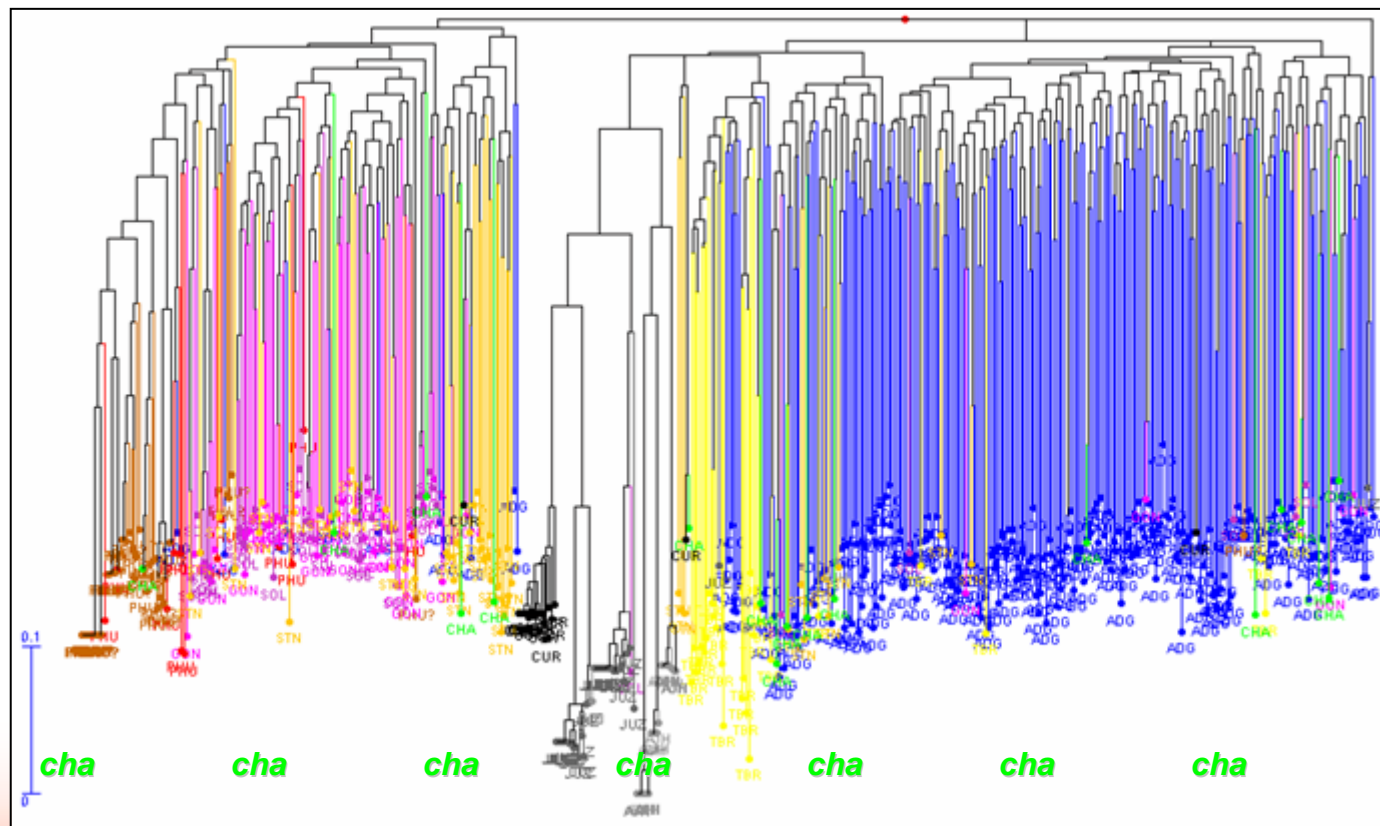
24 vs 32



40 vs 46

Cluster analyses with new PGI kit

Cluster analyses (Darwin) on 531 landraces X 24 SSR markers:
 $r = 0.97$ (vs 32 SSR markers)
92% of genotype identification



cha

cha

cha

cha

cha

cha

cha

phu

stn gon
sol

ajh juz
cur

tbr (chi)

adg

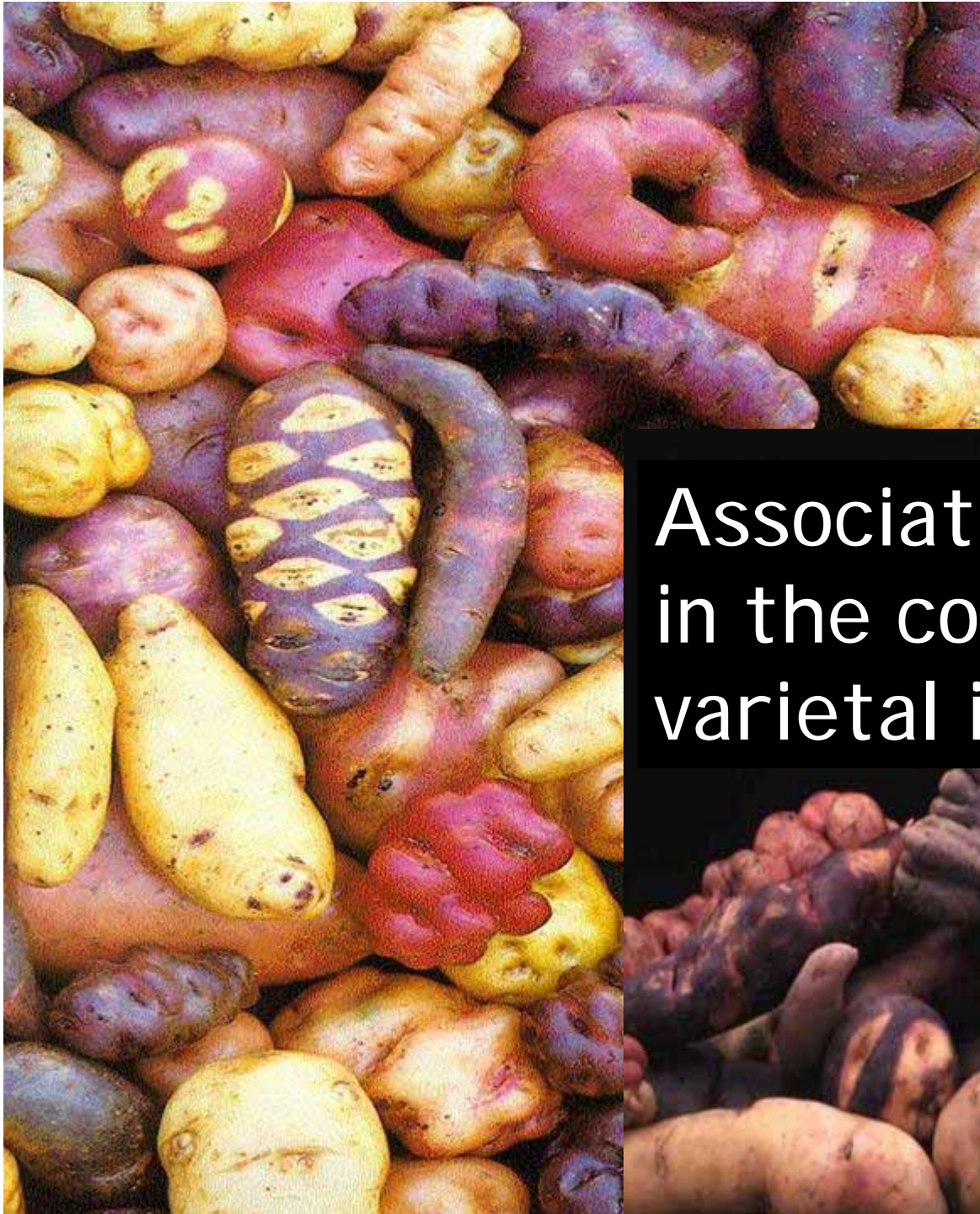
The International Potato Center CIP



Conclusions on the gene pool structure of the cultivated potatoes

- **46(53) SSR markers were characterized on 531(716) landraces (PIC value, map location, allele distribution, new PGI kit - 24 SSR markers);**
- **The cultivated potato gene pool appears to be structured into several sub-populations: (1) phu; (2) stn-gon-sol; (3) adg; (4) chi; (5) ajh; (6) juz; (7) cur;**
- **The taxonomic classification in either 7(9) species or 1 species with 8 cultivar-groups is unresolved because some sub species appear to be intermixed with others (gon with stn, cha with adg stn and phu) while others (ajh, juz, cur) appear well separated as species.**

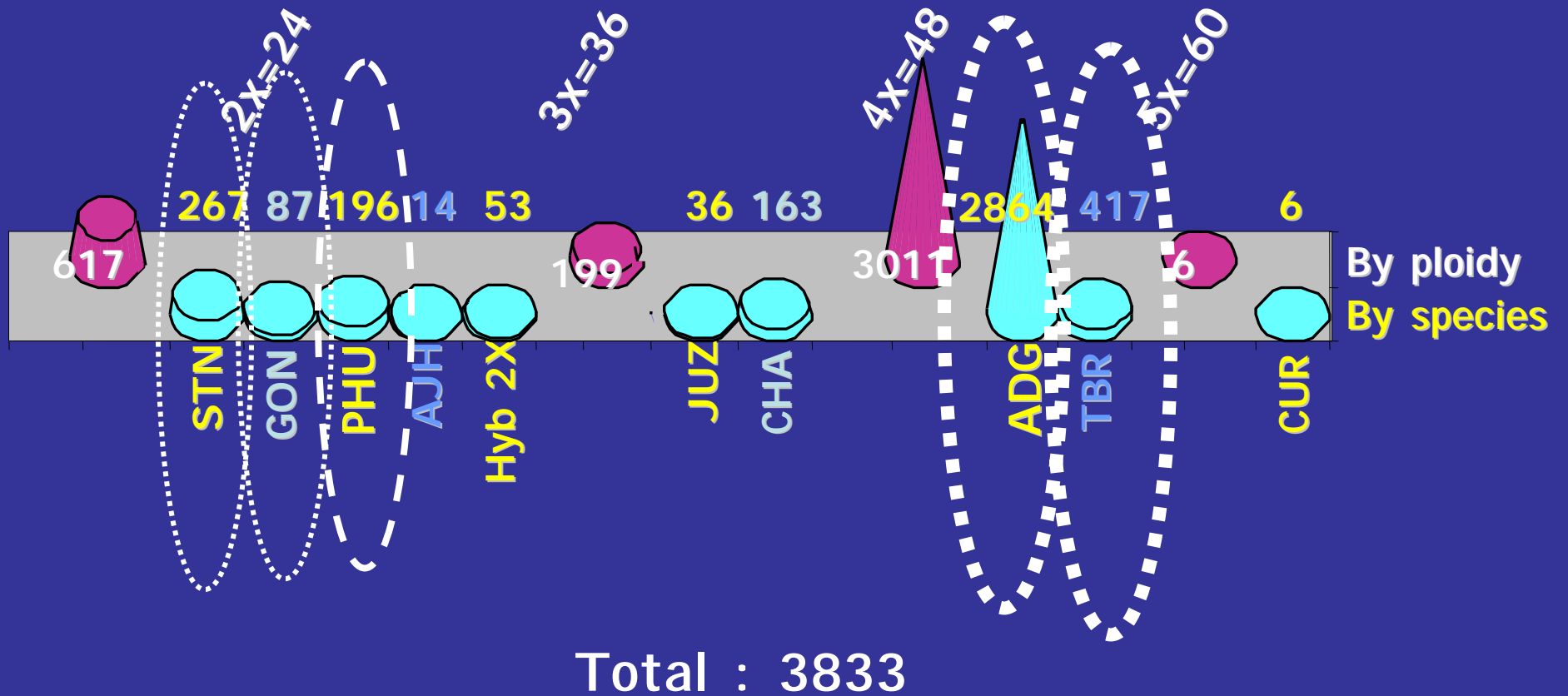




Association analysis
in the course of
varietal improvement



Representation of native cultivated potatoes at CIP



Breeding Objectives

- Durable disease resistance
 - Fungi- Late blight: *P. infestans*
 - Viruses: PVY, PLRV, PVX
 - Bacterial wilt: *Ralstonia solanacearum*
- Morphology and Phenology
 - Uniform, early tuberization & bulking (shorter vegetative period);
 - Tuber appearance
- Quality and nutrition
 - Rapid, high starch accumulation & low reducing sugars
 - Increased micronutrient concentration (Fe, Zn, Vitamin C, carotenoids)
- Adaptation to stressful environments
 - drought; warm temperatures
 - long day conditions (tuberization- photoperiod response); back to phenology

Reliable energy
productivity/unit
time, area & input

Preference &
Market traits

Nutritional quality

SP1 'Association' Project Goals

- Ultimate: To set the basis for efficient identification of useful genes/superior alleles assembled in germplasm collections or achieved in breeding programs
- Immediate: To determine population structure and the extent of LD in potato germplasm and address implications for association studies



Project Objectives

- Describe structure of LD in a tetraploid population known for diverse traits
 - For each population
 - Explore the extent of LD and its relationship to distance between pairs of SSRs from the genome
 - Evaluate rates at which LD decays in genome regions i.e., *among linked loci*
 - .. and *within genes*, via sequence data from several candidate genes

Really??
Awesome!!

Work plan

1. Assemble pedigree and phenotypic data; document phenotyping methods
2. Conduct gap-filling genotyping
3. Develop additional molecular data needed to estimate of structure and degree of LD
4. Complete phenotypic data sets



Factors affecting LD and their implications on Association Mapping in Potato

<i>Factor</i>	<i>Reason</i>	<i>Implications</i>	
		<i>Advantage</i>	<i>Issue</i>
Vegetative Propagation	Fewer meiosis	Higher LD	Lower Resolution
Bottleneck	Small number of ancestors	Extensive LD	Lower Resolution
Admixture	Complex breeding history/ introgression/ limited gene flow	-	Spurious associations
Recombination Hot Spots	≠ recombination rates across genome	-	Lower predictivity

Exhibit 1: 2x population (*S. phureja*)

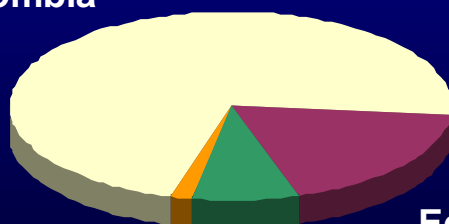
- ❖ 'Traditional' (modern-pedigree-wise unrelated) germplasm
- ❖ Andean landrace/farmers' cultivars
- ❖ Outstanding quality and nutritional traits
- ❖ Moderate levels of resistance to late blight and bacterial wilt
- ❖ Diploid, heterozygous, clonally propagated
- ❖ Likely arose from chance hybridizations among wild spp. and/or local cvs
- ❖ Less 'bottlenecked' than improved germplasm (?)
- ❖ Limited 'admixture' (prior assessment; no complex breeding history ??)



Origin & Evaluation status of *S. phureja* collection at CIP

Country of Origin

Colombia



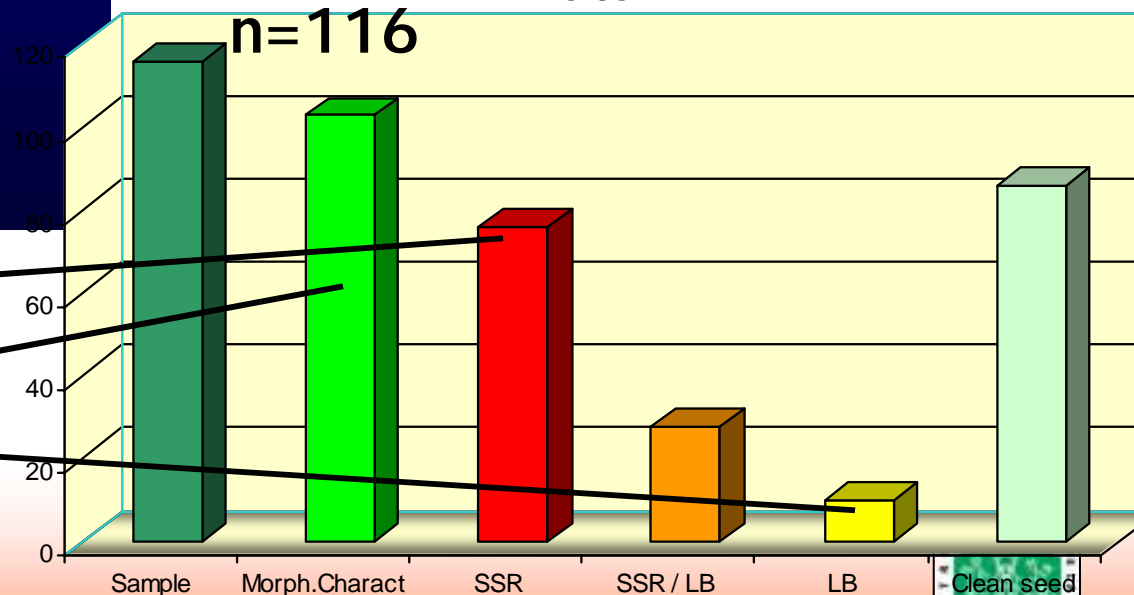
Bolivia Peru

Ecuador

Availability of Phenotypic & Genotypic Data

n=116

53 SSR loci
Morphology
Late blight
Bacterial wilt
(Micronutrients)

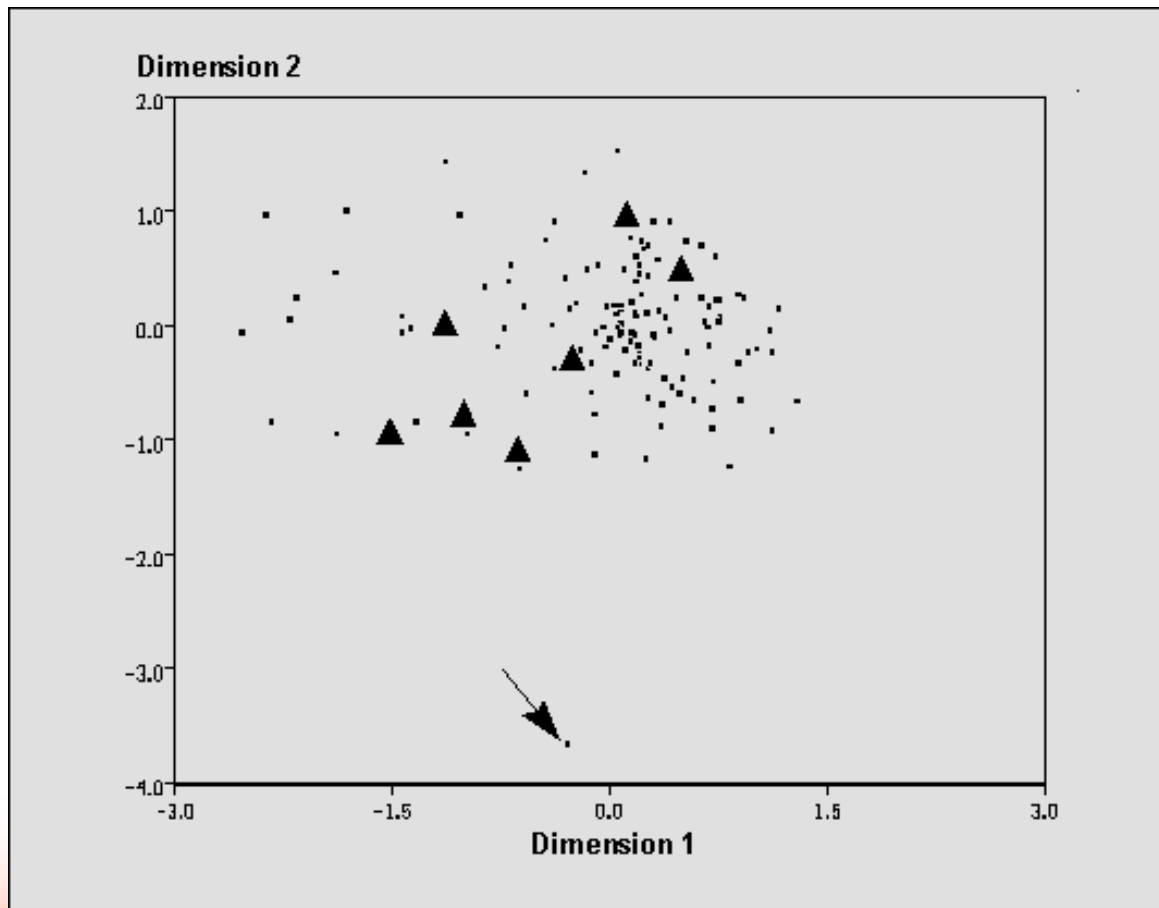


The International Potato Center CIP



Population structure

Previous assessment: Multidimensional scaling of 131 *Solanum phureja* accessions (RAPD data)



- Little apparent sub-structure
- No diversity x country patterns

(Ghislain et al. 1996)

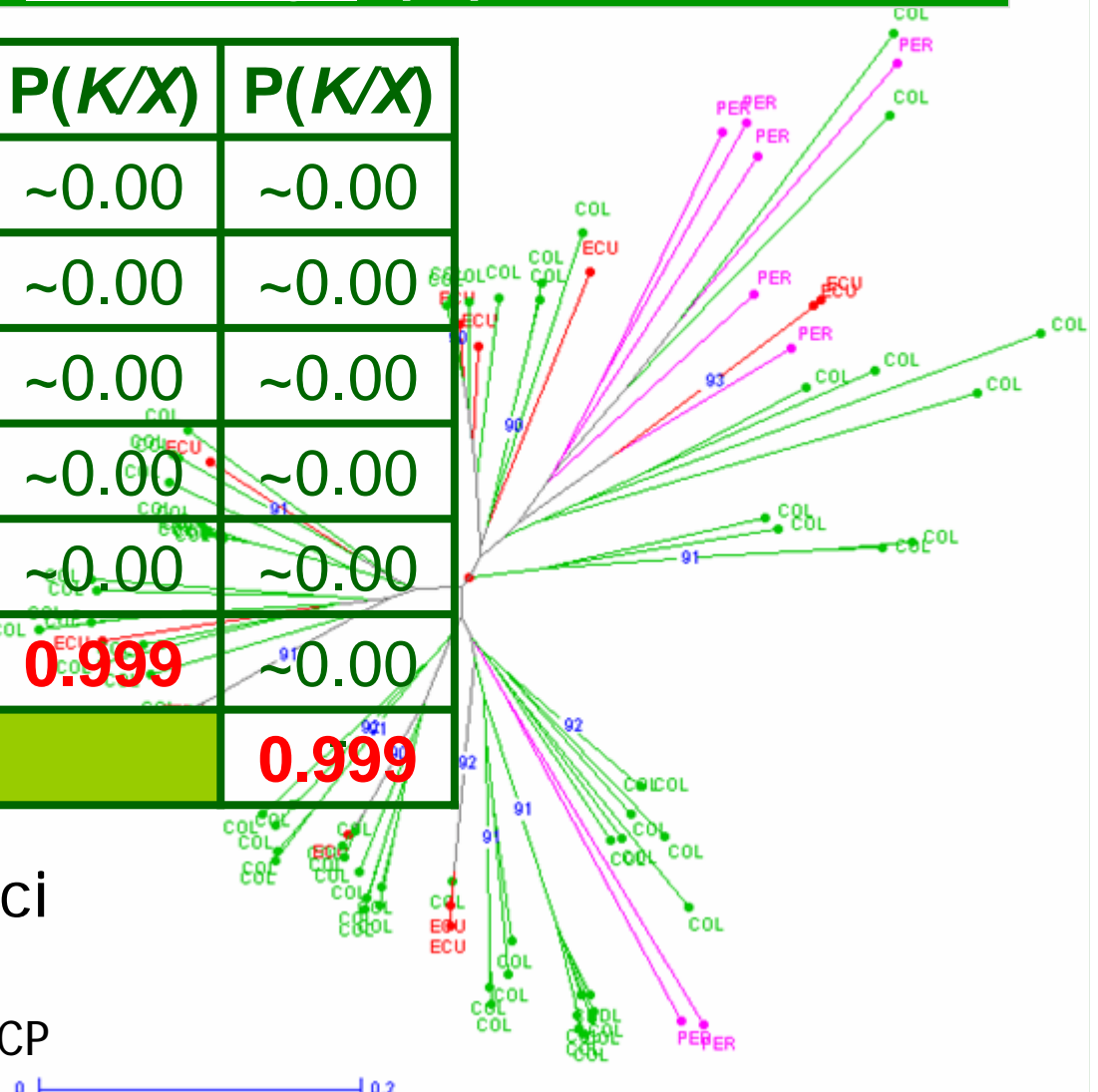
Population structure

Ongoing SP1/ Association Project: Inference of substructure in *S. phureja* 'population'

K	Ln P(X/K)	P(K/X)	P(K/X)	P(K/X)
1	-5625	~0.00	~0.00	~0.00
2	-5192	~0.00	~0.00	~0.00
3	-4841	~0.00	~0.00	~0.00
4	-4661	~0.00	~0.00	~0.00
5	-4525	1.00	~0.00	~0.00
6	-4339		0.999	~0.00
7	-4327			0.999

98 accessions - 46 SSR loci

Ghislain et al. CIP/GCP



0 |-----| 102

Next steps (*2x Phureja*)

- Check population structure
 - Prune redundant individuals from putative sub-groups or at least those suggested by the tree
 - Refine set of SSRs: Eliminate one/two in closely-linked pairs; Weighting? F and T² statistics?
- Settle interpretation of single band 'genotype': homozygote vs. heterozygote (assuming alternate allele as null)
- Fill gaps in SSR and phenotypic datasets
- Calculate R² among markers to estimate LD
- Sequence candidate genes (haplotype) & estimate LD decay
- Consider adding in remaining *2x* groups - *Stenotomum* and *Goniocaylx*
- Identify a combined *2x* reference set?

211 diploid genotypes – 34 SSR

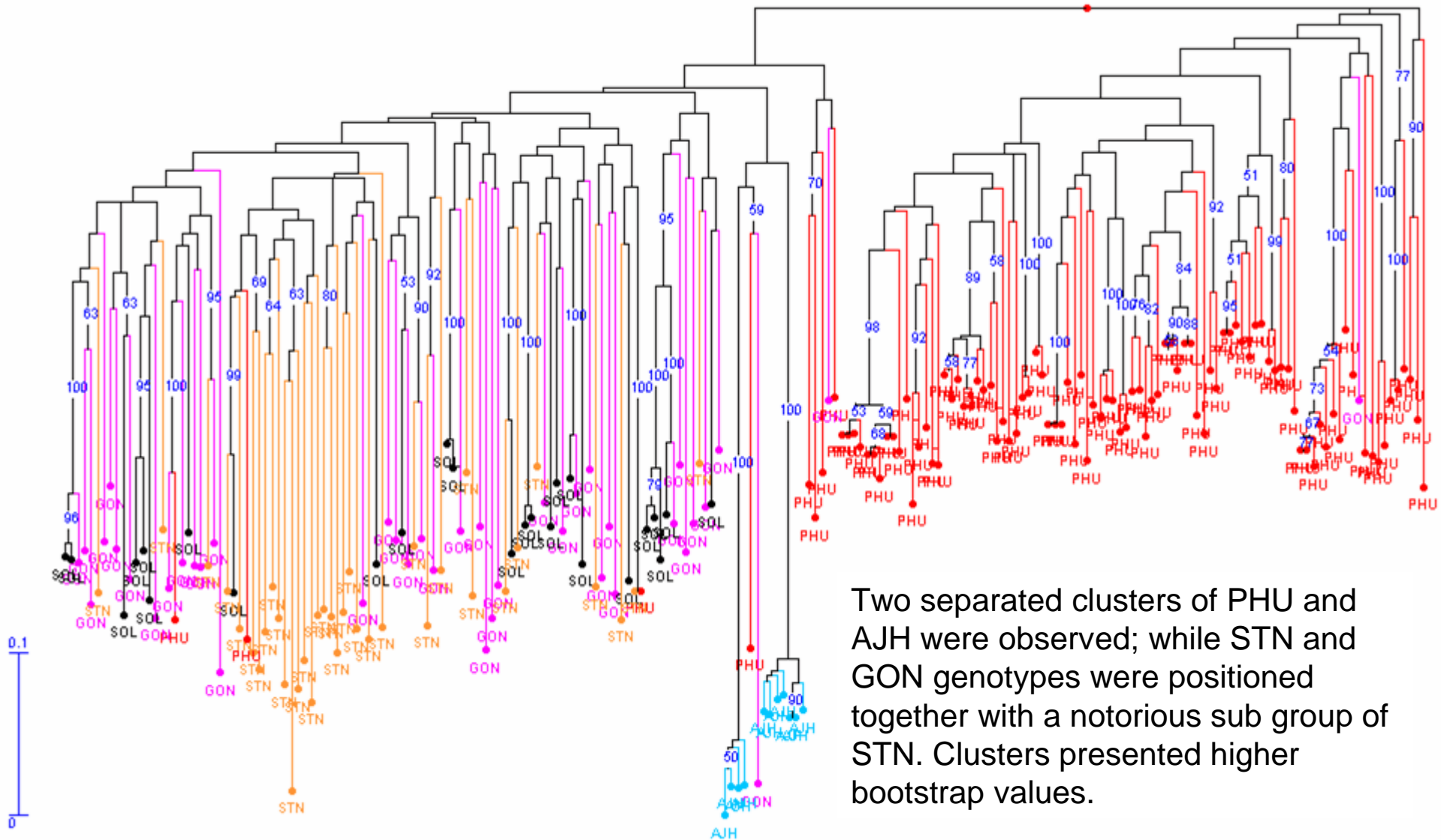


Exhibit 2: Improved 4x Breeding Population 'B1'

- ❖ Founded from 57 native (Andigena) accessions
- ❖ 5 cycles of genotypic recurrent selection for resistance, morphological, and phenological traits
 - ✓ Well-documented breeding program; multi-generation pedigrees available
 - ✓ Extensive, high-quality phenotypic data available
- ❖ Expect Bottleneck and Admixture: *Phureja* < 'B1' < modern variety collections
- ❖ Locus-specific bottlenecks due to selection cause LD between the selected alleles at a locus and linked loci
→ Ideal for association mapping using candidate genes.

**Advanced tetraploid potato breeding population
Derived from Andigena through 5 cycles of
recurrent selection for resistance and agronomic
traits**



e.g. Phenotypic data ..quantitative traits

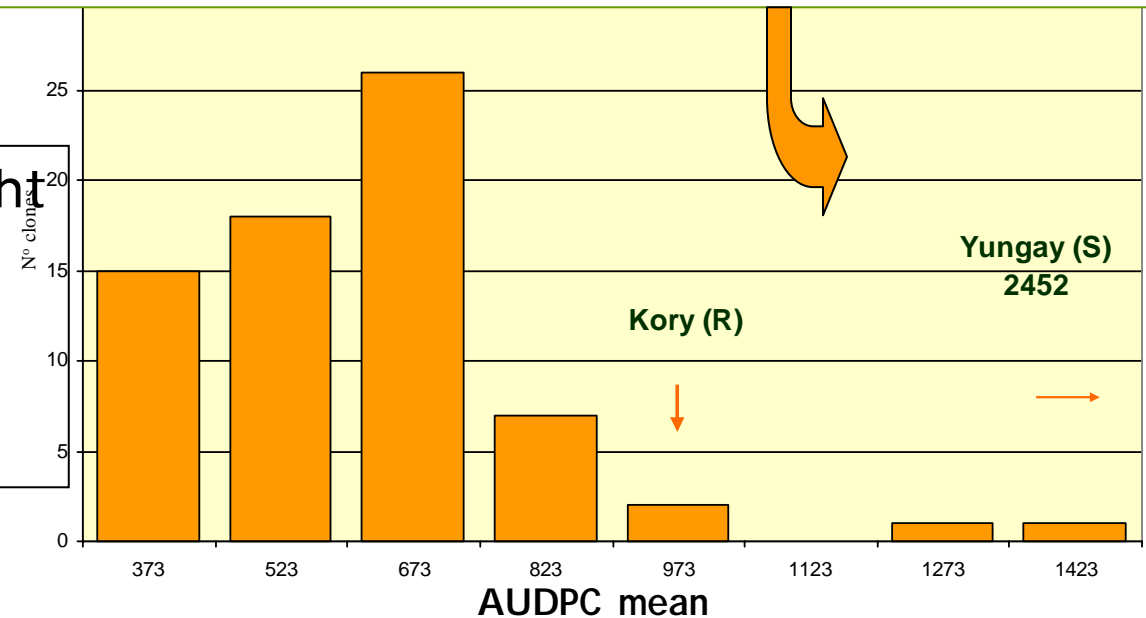


PHENOTYPE No 1
Quantitative
resistance to late
blight

70 genotypes- + n additional 'susceptibles'

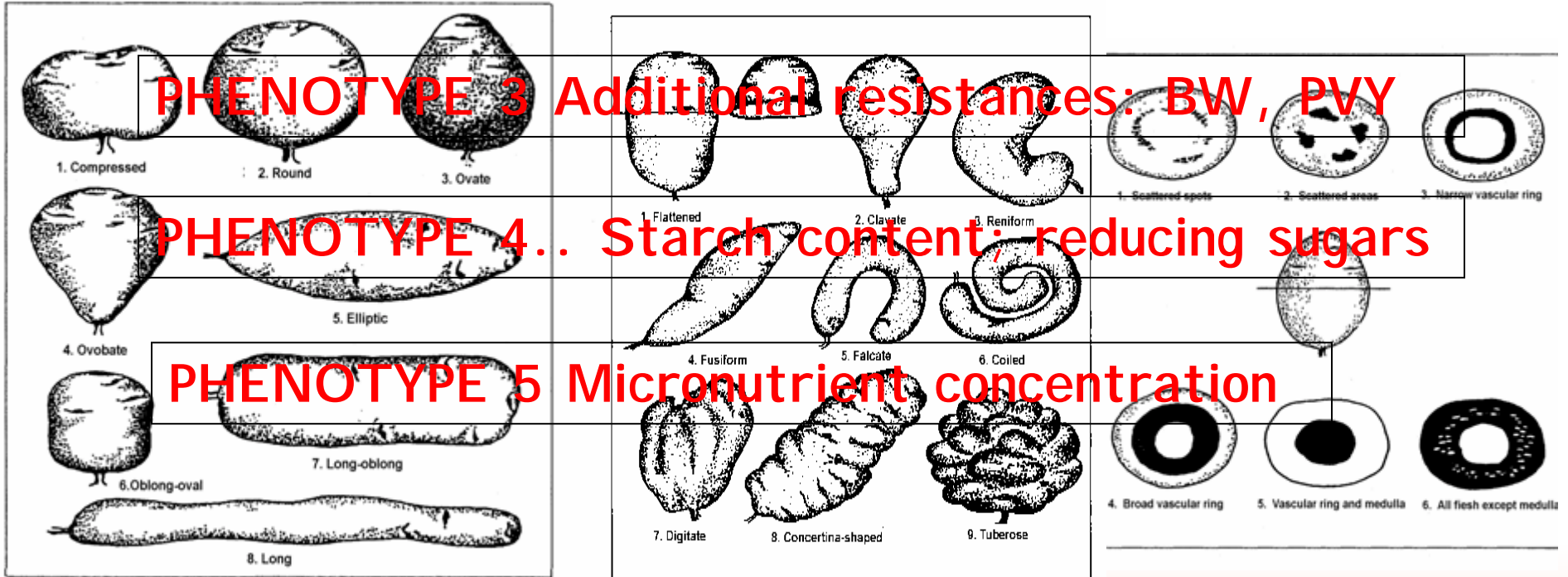
Field Resistance to Late Blight

- Replicated trials
- 2 Locations
- Complex isolates
- % LAI → AUDPC



Phenotypic data ..quantitative traits

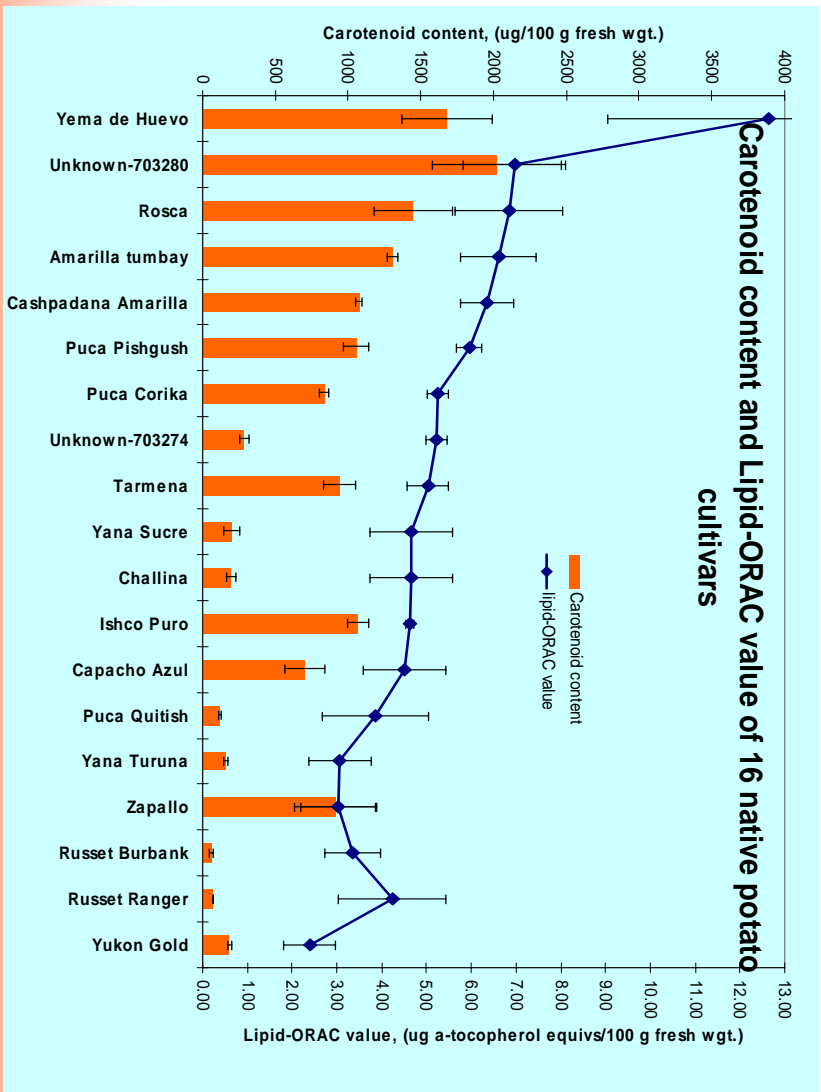
PHENOTYPE 2 Morphology



... PHENOTYPE X Drought tolerance



Phenotypic data: Morphology; 1° & 2° metabolites



Needs & challenges

- Haplotype phase (Greater number of unknown haplotype phases in tetraploids than diploids)
 - Issue: Several clones/products per individual need to be analyzed if all haplotypes are to be detected (approx. 13 randomly-selected clones for 90% chance of detecting all alleles)
 - Option: Use of programs developed for DNA pools of diploid genotypes?
- No programs available to perform association studies on polyploids?