Max Planck Institute for Plant Breeding Research

PLANT & ANIMAL GENOME XX
The Largest Ag-Genomics Meeting in the World.
January 14-18, 2012
San Diego, CA
www.intl-pag.org

KWS

Genetic diversity, population structure, and linkage disequilibrium in elite sugar beet germplasm

Jinquan Li[1], Ann-Katrin Lühmann[2], Britta Schulz[2], Knuth Weißleder[2], Benjamin Stich[1,2*]

[1]Max Planck Institute for Plant Breeding Research, Germany;
[2]KWS SAAT AG, Germany

Importance of the analysis of population structure and linkage disequilibrium(LD) for breeding

- Exploring the breeding history
- Essential for the efficient organization and utilization of germplasm
- Important to identify essentially derived varieties
- Identifying potential target genes under selection
- Necessary for the design of association mapping experiments

Population structure indicated by STRUCTURE and PCoA based on modified Roger's distance (MRD) estimates

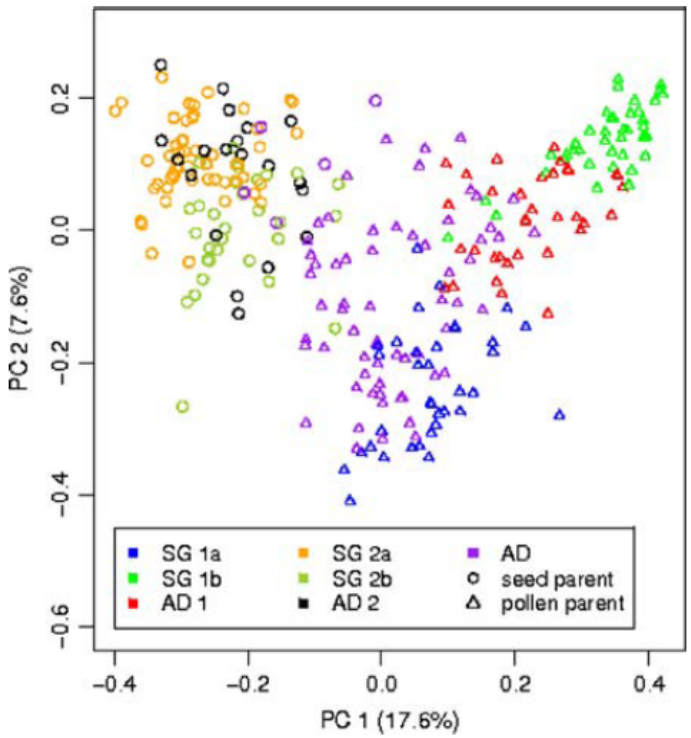
Genotyped with 23 SSR markers by KWS SAAT AG

Group	*N*
Entire germplasm set	289
SG 1	115
SG 1a	35
SG 1b	45
AD 1	35
SG 2	106
SG 2a	55
SG 2b	31
AD 2	20
AD	68

All inbreds were from pollen parent (PP) heterotic pool

All inbreds were from seed parent (SP) heterotic pool

5 inbreds were from SP pool and 63 from PP pool

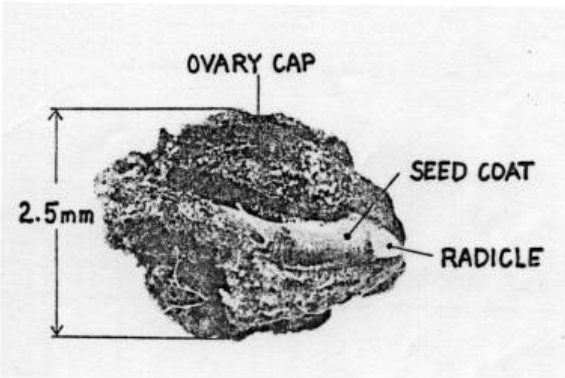


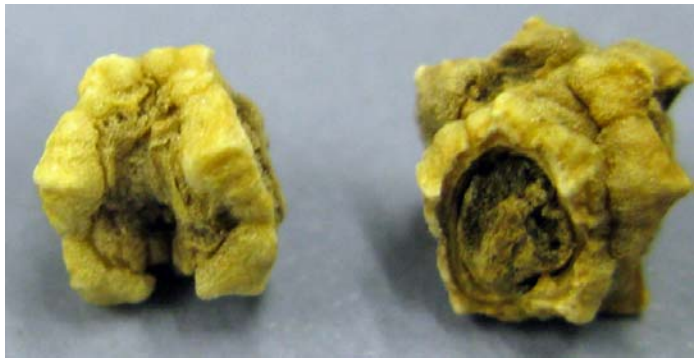
As indicated by PCoA, most of the inbreds from the SP and PP pool are located in two distinct clusters

Li et al. Euphytica, 2010, 175: 35-42

SP and PP heterotic pools of sugar beet

- Two important pools for hybrid breeding:
  - SP heterotic pool (♀): monogerm germplasm
  - PP heterotic pool (♂): multigerm germplasm

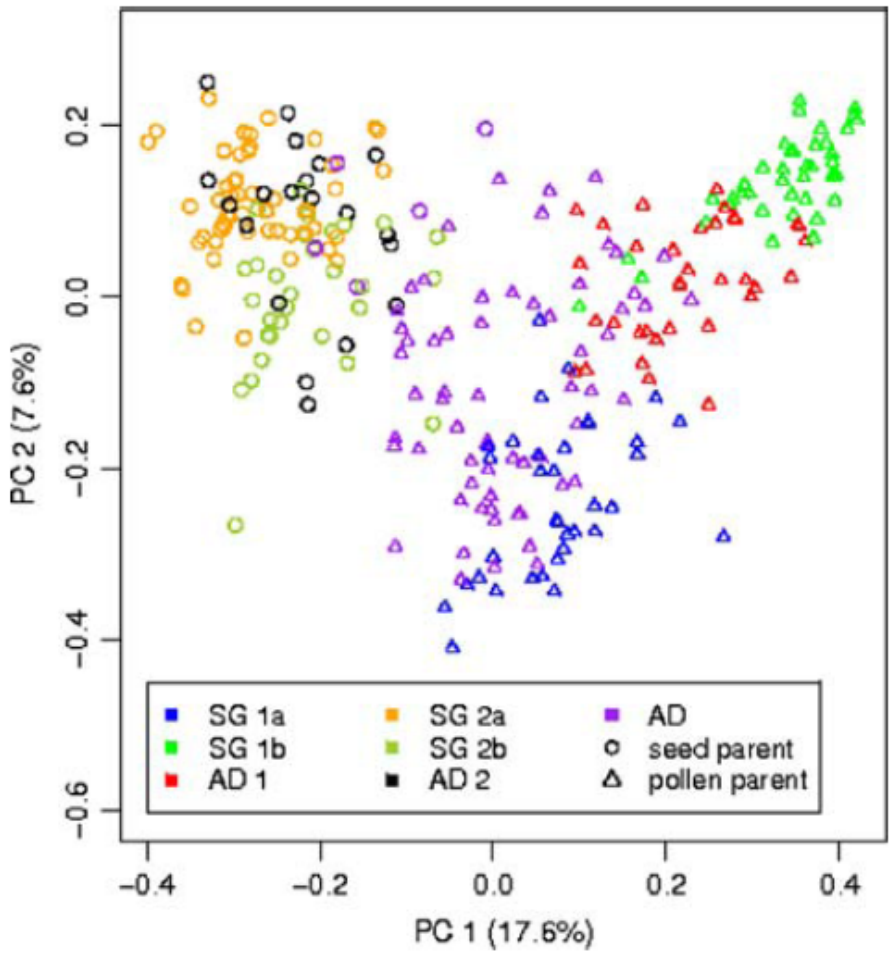




provided by Olaf Brinkman, KWS

- SP pool was developed out of the PP pool in the late seventies of the 20th century.
- Since then, two heterotic pools have been developed separately.

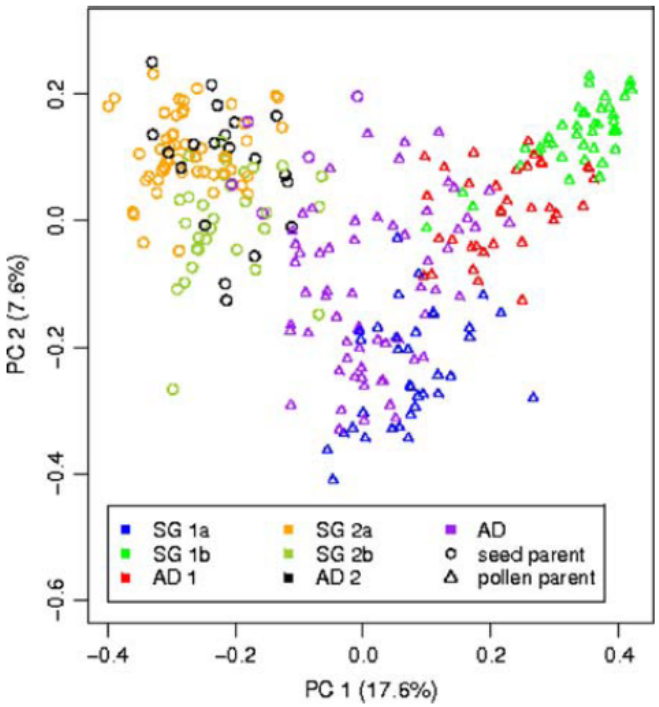
Population structure indicated by STRUCTURE and PCoA



As indicated by PCoA, most of the inbreds from the SP and PP pool are located in two distinct clusters

It indicates that only thirty years of recurrent reciprocal selection have the potential to lead to differentiated populations

Genetic diversity within the elite germplasm set



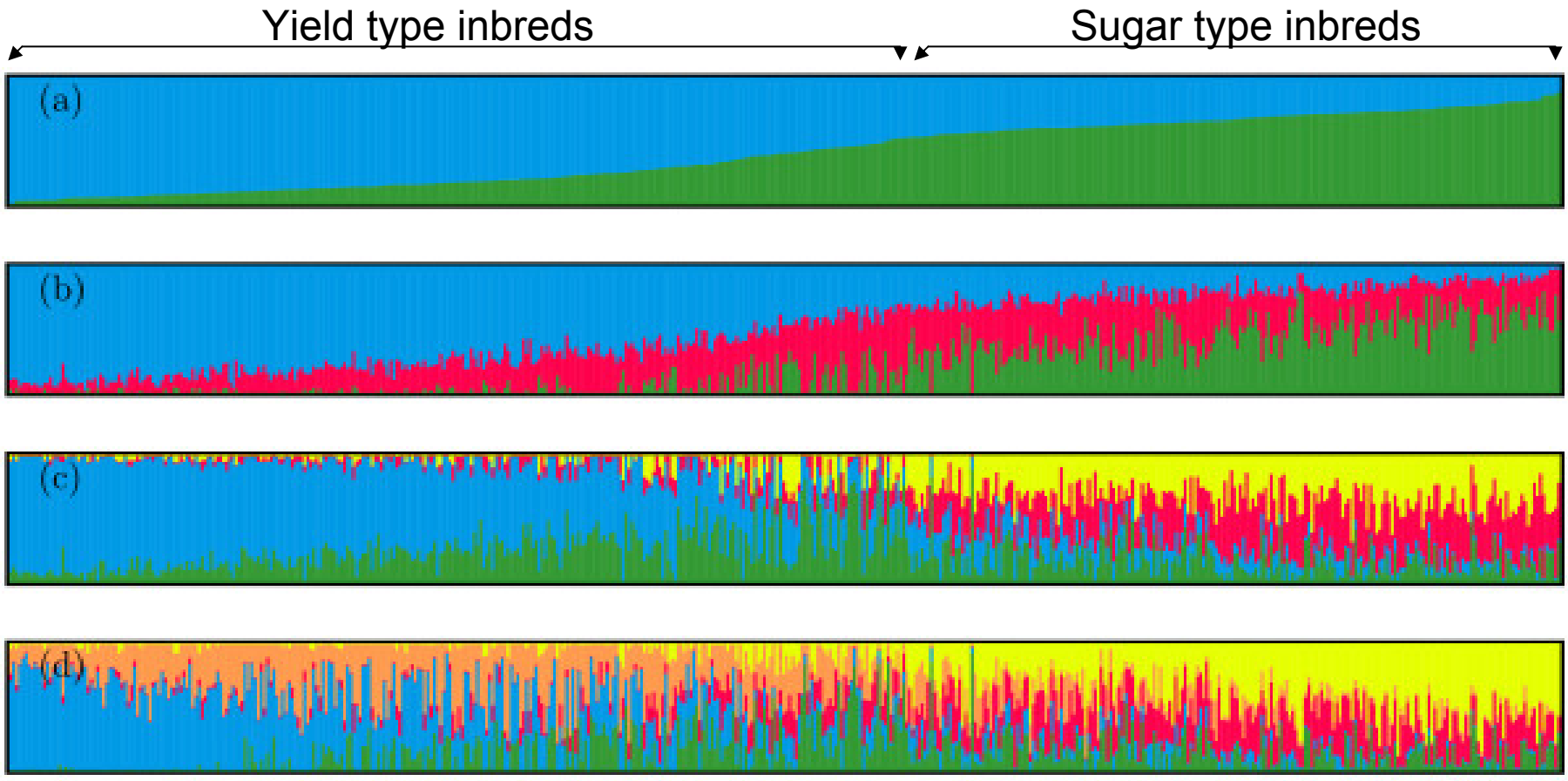
Group	*N*	*A*	*I*	*Gst*
Entire germplasm set	289	5.0	0.60	-
SG 1	115	4.3	0.52	0.128
SG 1a	35	3.6	0.55	0.019
SG 1b	45	2.9	0.30	0.440
AD 1	35	3.6	0.49	0.087
SG 2	106	3.0	0.42	0.309
SG 2a	55	2.5	0.35	0.170
SG 2b	31	2.5	0.33	0.215
AD 2	20	2.7	0.39	0.070
AD	68	4.4	0.59	0.019

The results revealed that SG 1 and AD (mainly from PP pool) have a higher gene diversity than SG 2 (from SP pool), while SG 2 is more divergent than the other two groups

Materials

- 502 accessions of the PP heterotic pool of sugar beet
- Genotyped with 328 SNPs markers by KWS SAAT AG
- A total of 26, 33, 41, 35, 40, 42, 39, 32, and 40 SNPs were distributed on the 9 linkage groups

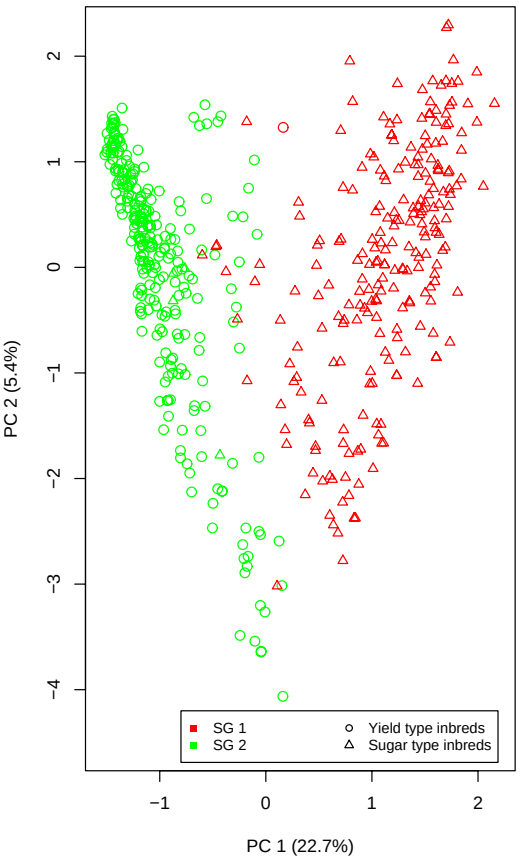
Population structure indicated by STRUCTURE



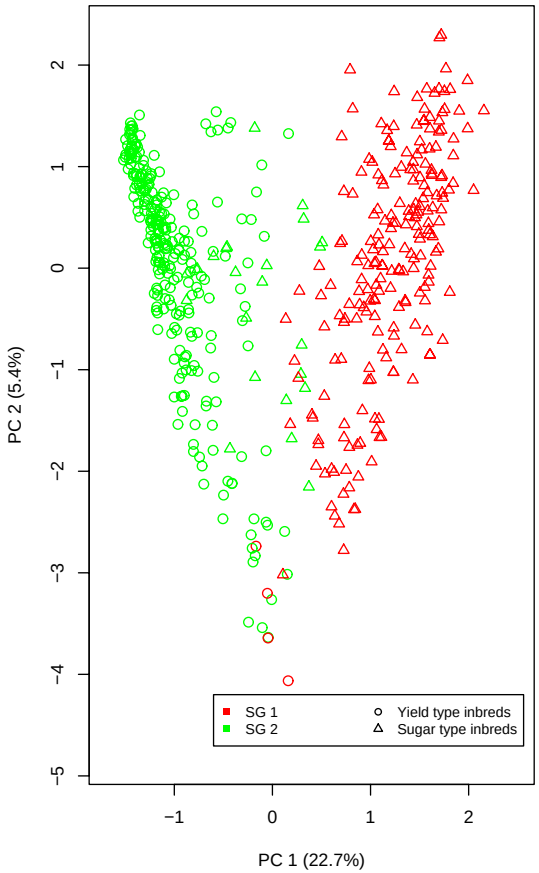
Two subgroups were identified by STRUCTURE. SG 1 mainly consits of inbreds from sugar type and SG 2 mainly from inbreds of yield type

In production, because there is almost invariably a negative correlation between root yield and sugar percentage, there are yield and sugar types of sugar beet (Bosemark 2006)

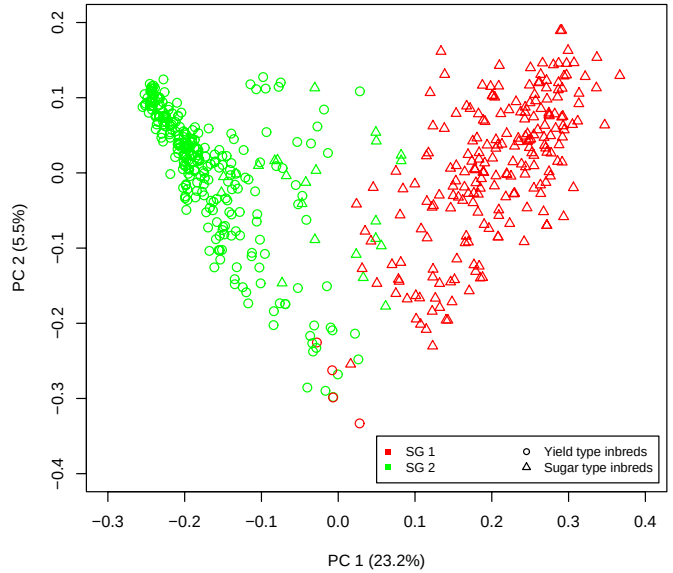
Population structure indicated by PCA, PCoA, LAPSTRUCT, and MCLUST



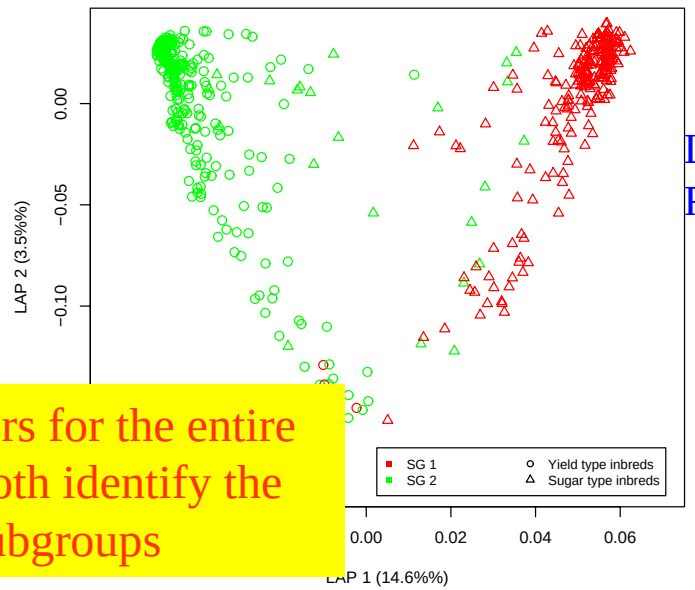
PCA+MCLUST



PCA+Structure



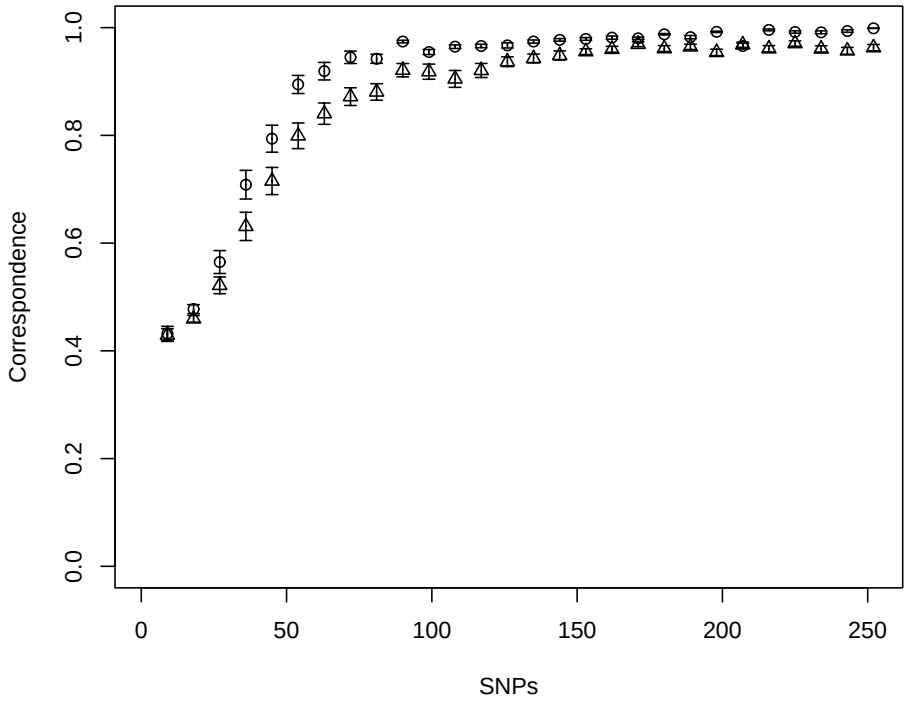
PCoA

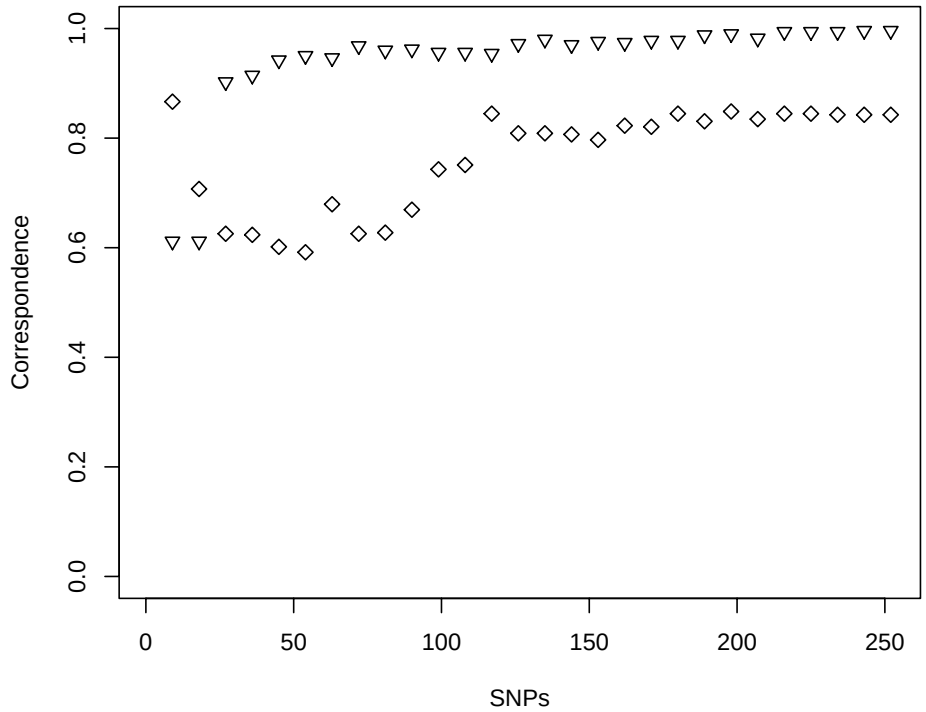


LAPSTRUCT

PCA, PCoA, Lapstruct identified two distinct clusters for the entire germplasm set. MCLUST and STRUCTURE can both identify the number of subgroup and assign the individuals to subgroups

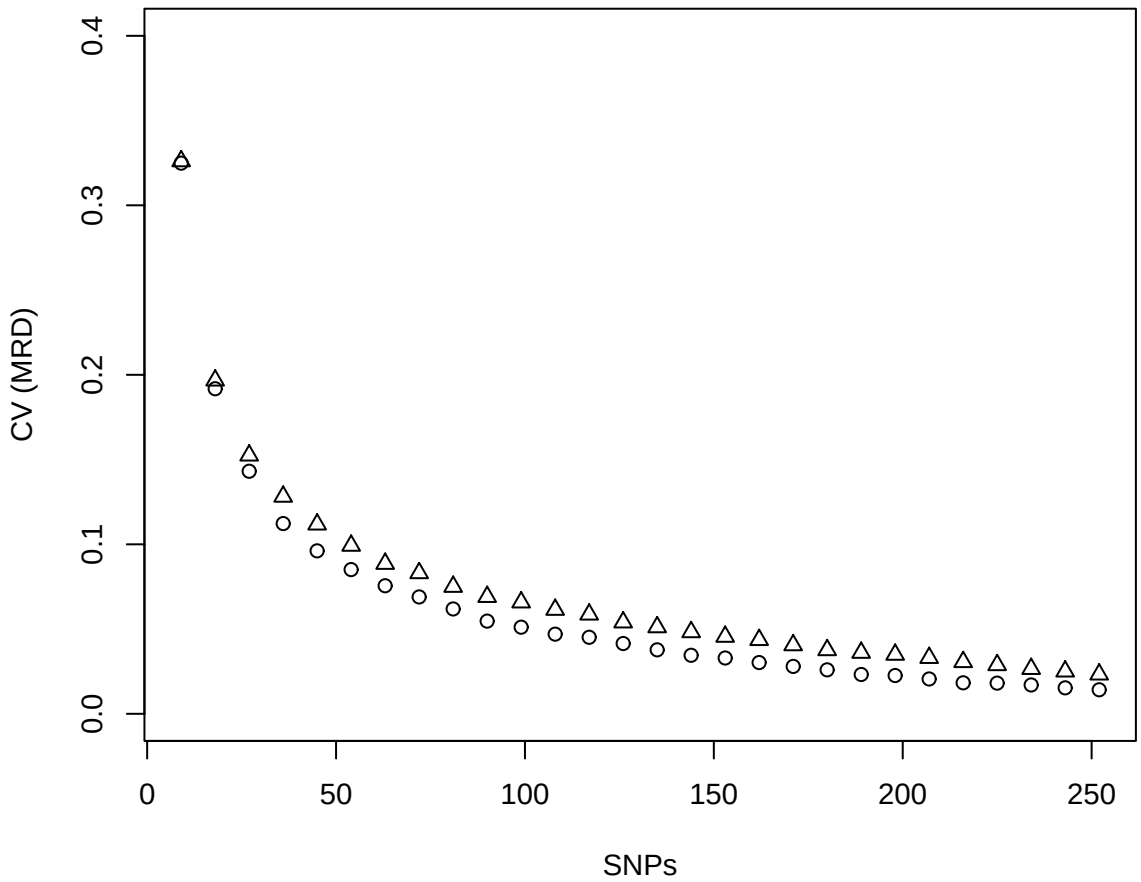
Comparison of different numbers of SNPs for detecting population structure





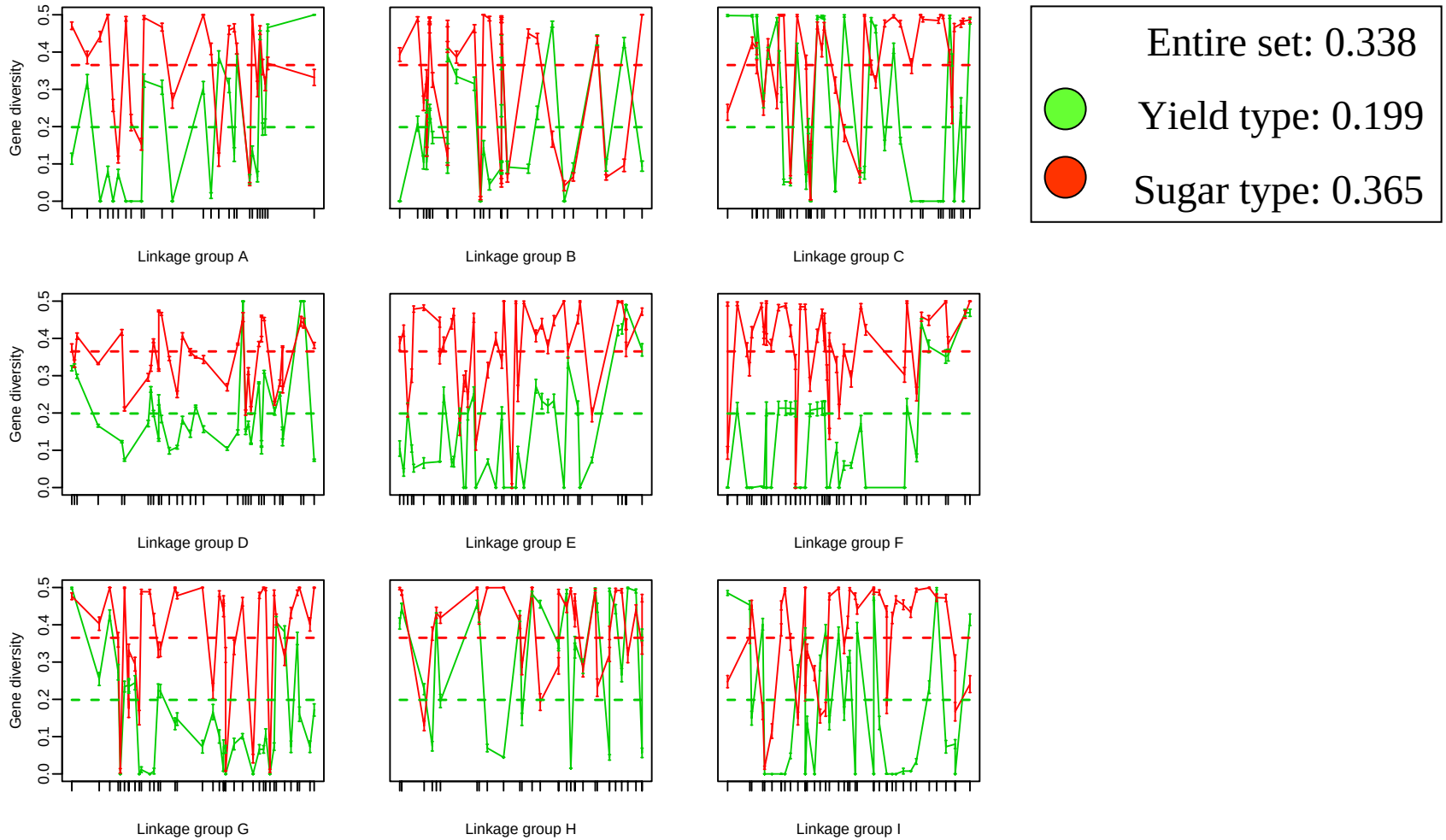
Correspondence increased as the number of SNP markers increased. When SNPs reached about 100 SNPs, not much further gain could be obtained. The correspondence could be even higher when the SNPs were selected with respect to their PIC values

Comparison of different numbers of SNPs for detecting population structure



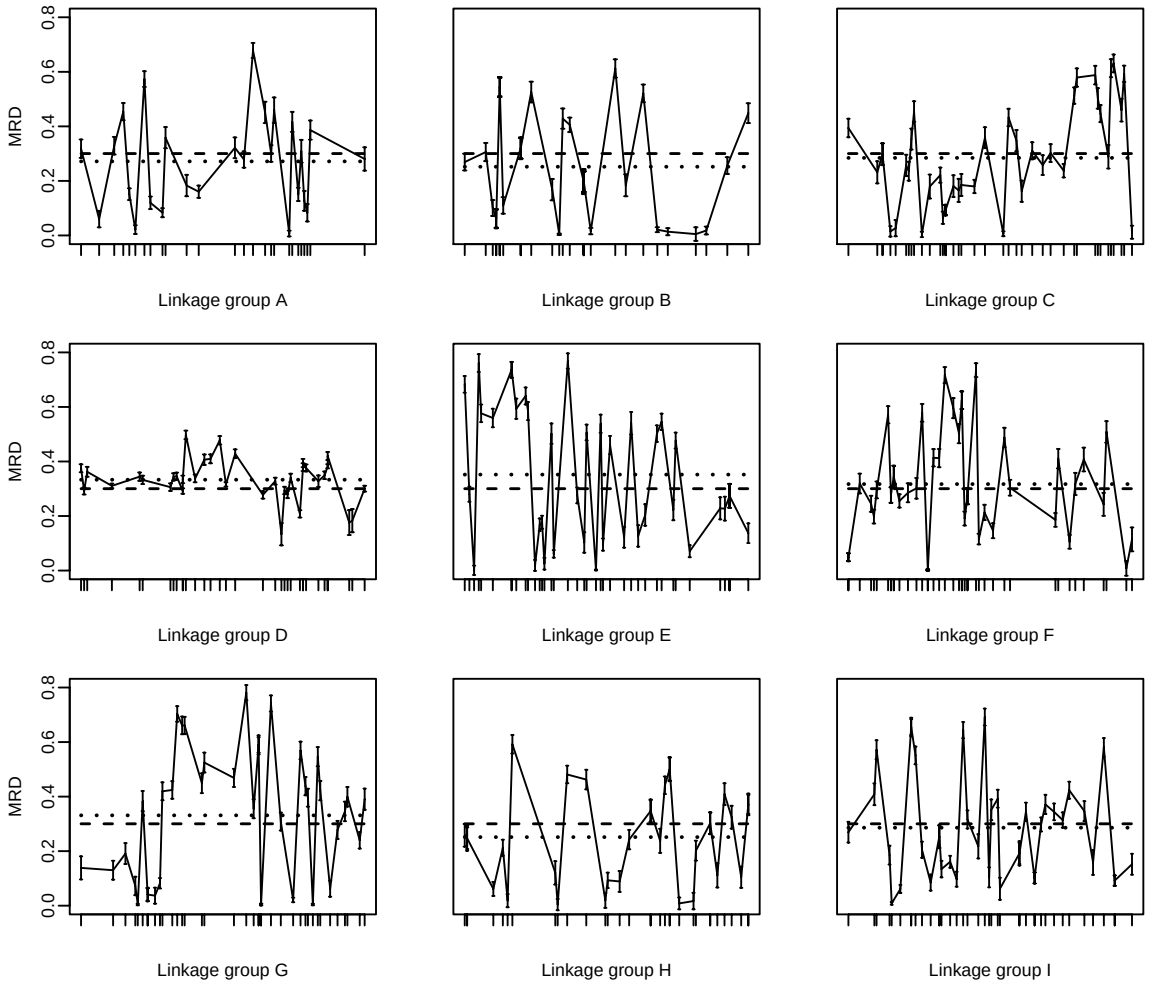
CV for MRD decreased as the number of SNP markers increased. When SNPs reached about 100 SNPs, not much further gain could be obtained

Genome-wide distribution of genetic diversity for yield and sugar type inbreds



For most genome regions, sugar type inbreds showed higher gene diversity than yield type inbreds

MRD between yield and sugar type inbreds across the genome



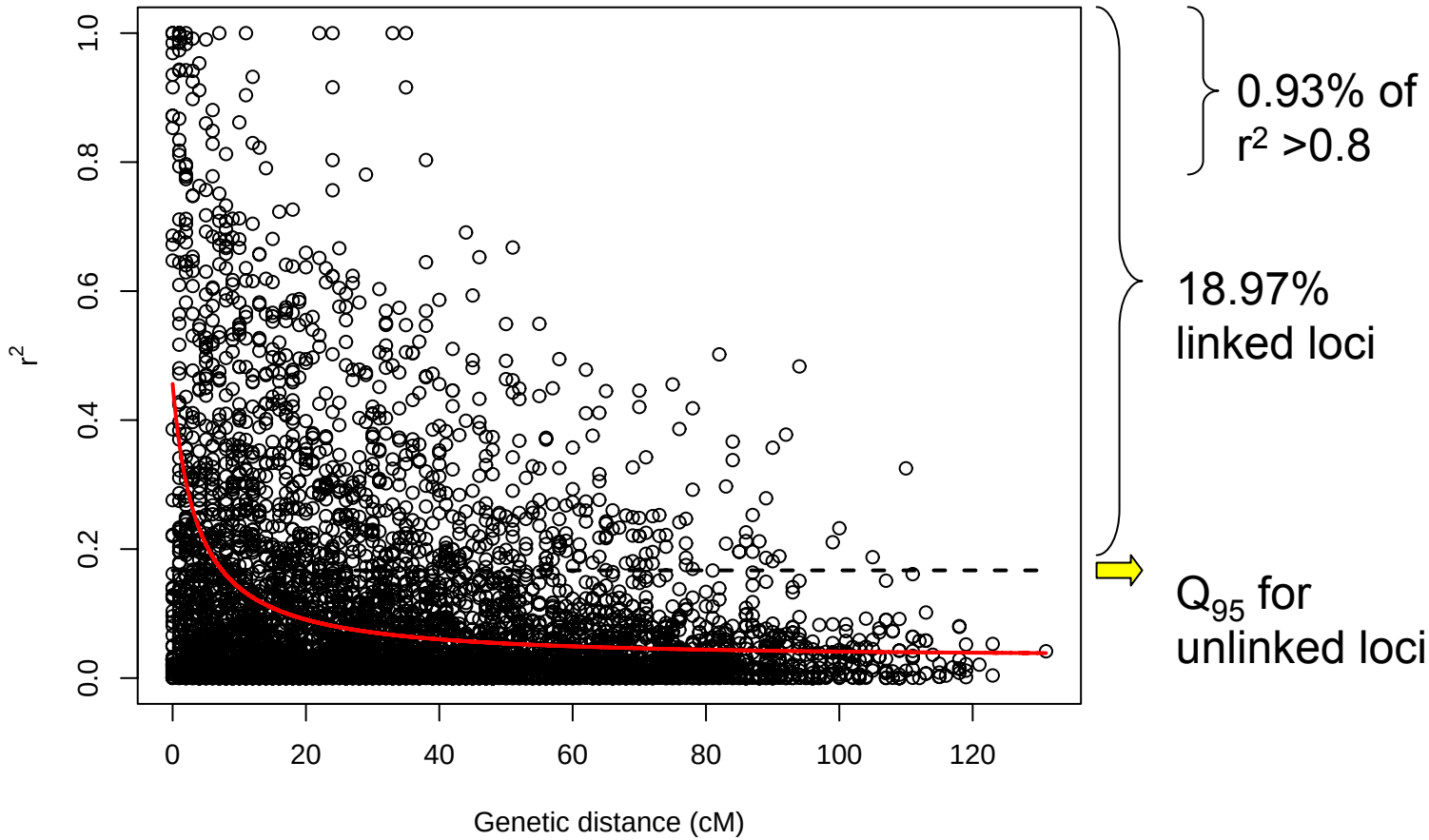
All individuals: 0.562

Yield and sugar type:

0.311

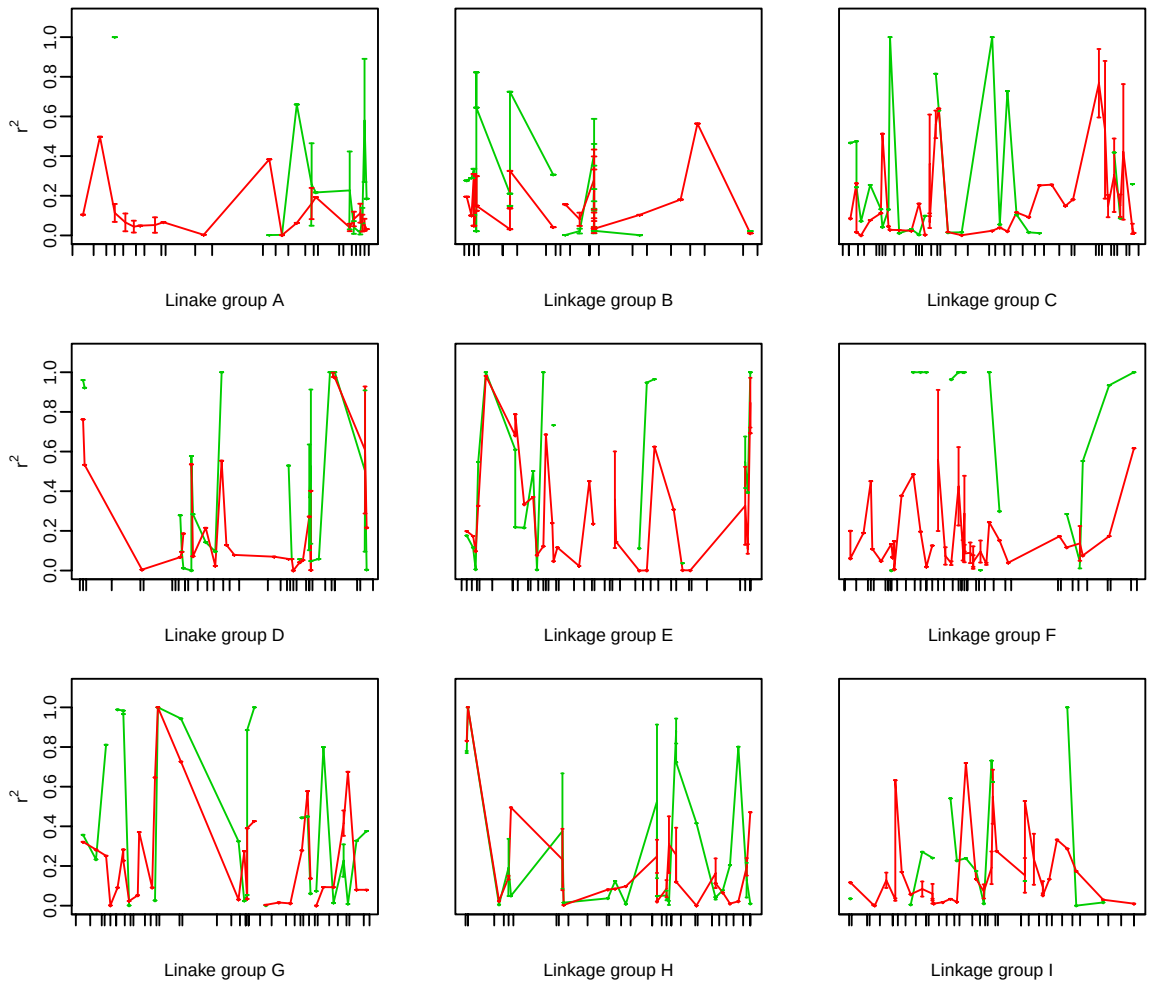
Different degree of divergence was found across genome, which might be due to artifical selection

Extent of LD for the entire germplasm set



LD decays to the Q_{95} threshold at about 7.4cM. 18.97% linked loci pairs are in significant LD

Extent of LD for all linked loci pairs within 5cM intervals across the genome



Green/red lines indicate average r^2 of yield/sugar type inbreds. The vertical line at each point represents the standard error

We observed different LD levels along the linkage groups. This observation suggests that different numbers of markers are required in different genome regions

Effective population size (*Ne*) inferred from the decay of LD

Effective population size (*Ne*) for the entire germplasm set, yield type inbreds, and sugar type inbreds across all linkage groups

	A	B	C	D	E	F	G	H	I	All
Entire germplasm set	137.4	68.0	62.9	89.2	23.0	52.8	28.3	57.3	80.0	52.7
Yield type	47.1	30.7	16.8	31.6	15.5	12.3	16.5	29.2	23.6	21.2
Sugar type	210.7	87.4	91.8	83.3	36.0	92.7	48.2	66.2	81.6	72.7

Ne was higher for sugar type than yield type inbreds and varied among linkage groups

Summary

1. Two distinct subgroups related to the SP and PP heterotic pools were identified, which indicated that the SP pool are clearly sparated from PP pool despite only 30 years of separation
2. Furthermore, two distinct subgroups were identified for the inbreds from PP heterotic pools, which is in accordance to breeding history
3. MCLUST analysis had high correspondence with the germplasm types and STRUCTURE results
4. About 100 SNPs could identify the same population structure as the whole SNPs set did
5. Gene diversity and MRD varied considerably across genome and between pools
6. The percentage of marker pairs with $r^2 > 0.8$ was low and, thus, more markers for genome-wide association mapping are required

Li et al. BMC Genomics, 2011,12: 484

Acknowledgment

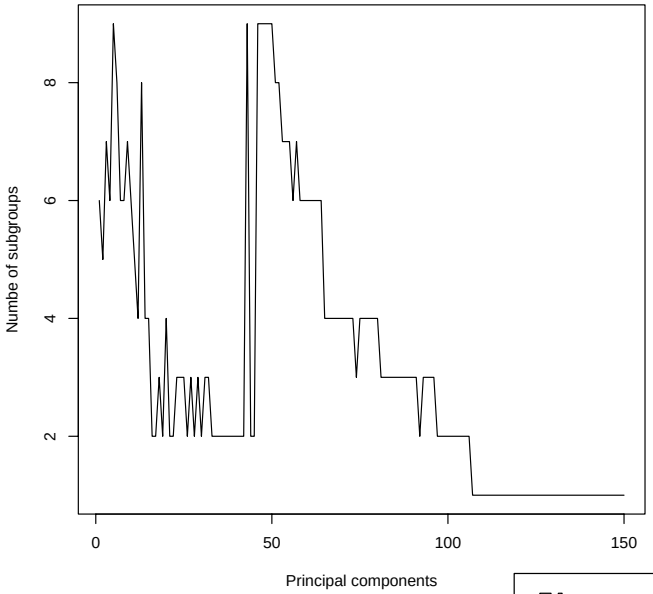
KWS SAAT AG

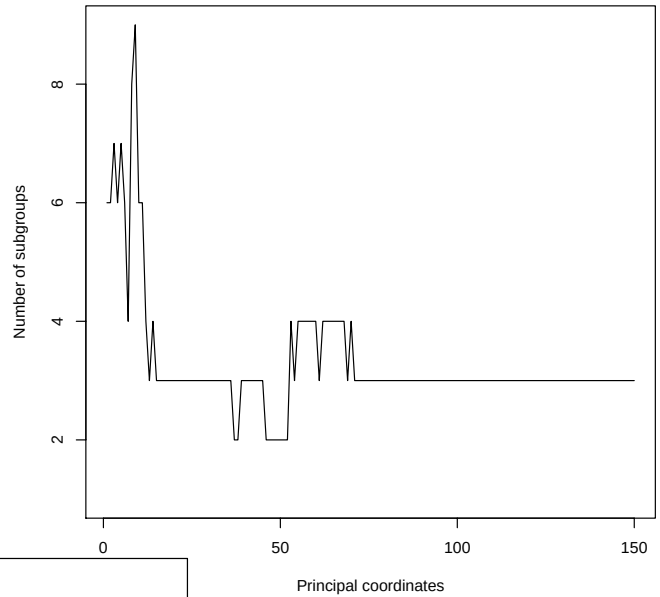
Max Planck Society

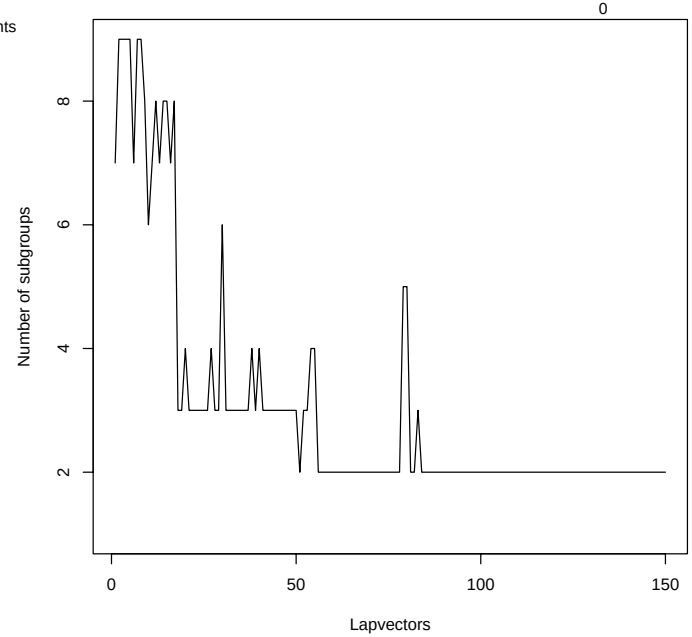
Bundesministerium für Bildung und Forschung (BMBF)

Thank you very much for your attention!

Numbers of subgroups identified by MCLUST based on PCA component, PCoA coordinate, and Lapvectors

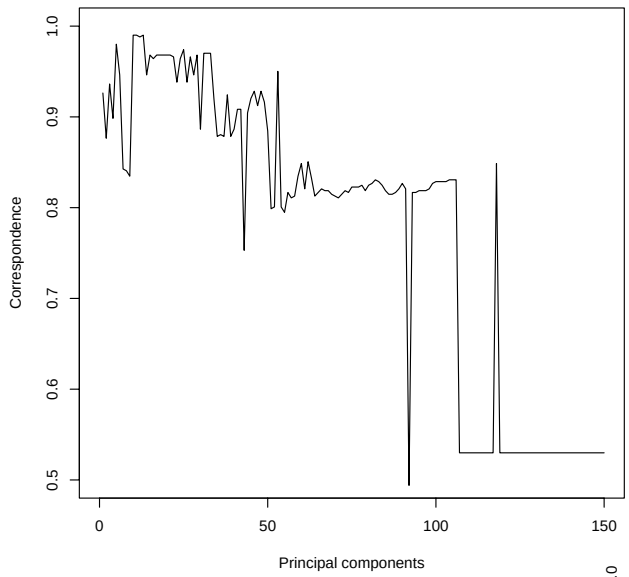


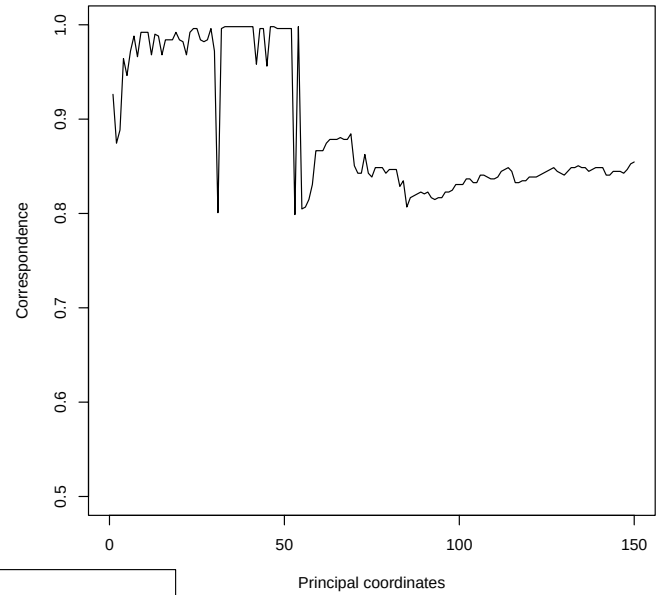




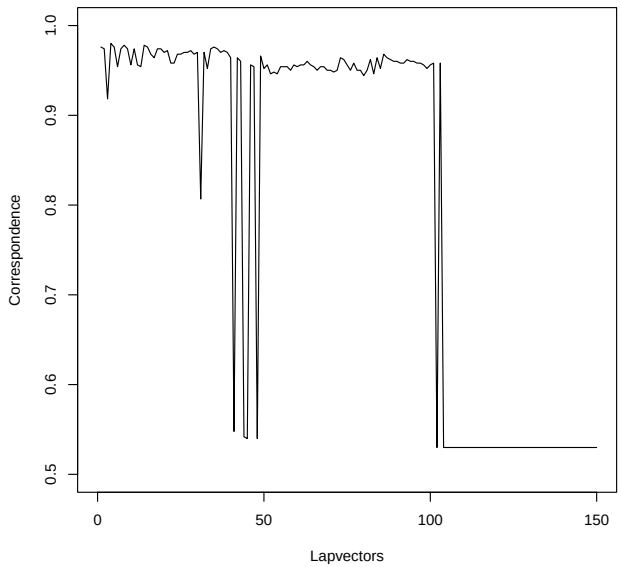
The number of subgroups identified by MCLUST varied from 1 to 9 when 1-150 PC, PCo, or Lapvectors were applied.

Correspondence between the known germplasm types and the assignment by MCLUST



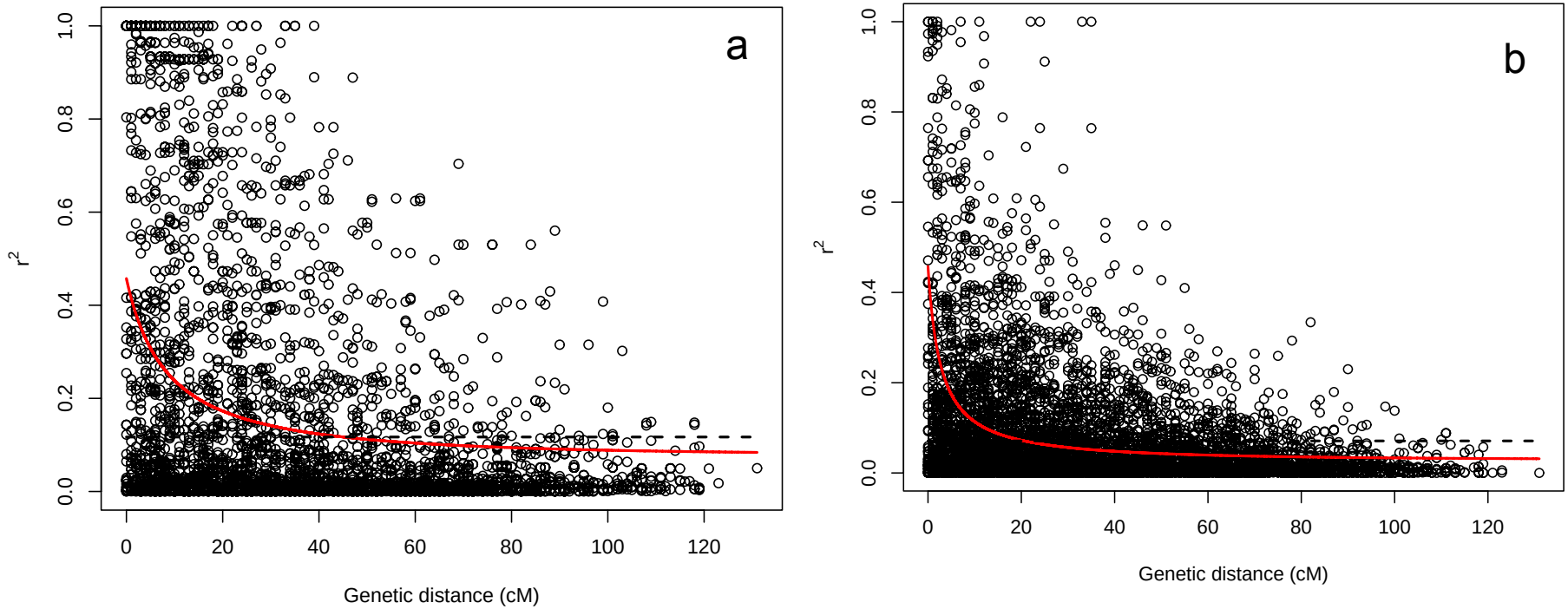


When the number of subgroups was set to two, we observed for 10-40 PC, 10-50 PCo, and 1-100 Lapvectors >95% correspondence with the germplasm type



Our observation suggested that it should be carefully inferred the number of subgroups

Extent of LD for yield and sugar type inbreds



LD (r^2) for linked loci pairs vs. genetic distance within yield (a) and sugar (b) type inbreds . Dashed line is Q_{95}, and red line is nonlinear regression

LD decays to threshold at about 45.1cM for yield type inbreds, and about 20.6cM for sugar type inbreds. 31.8% and 32.0% linked loci in yield and sugar type inbreds are in significiant LD. Percentage for $r^2 > 0.8$ are 6.2% and 0.7%