Analysis of training population effects on genomic selection in *Brassica napus* L.
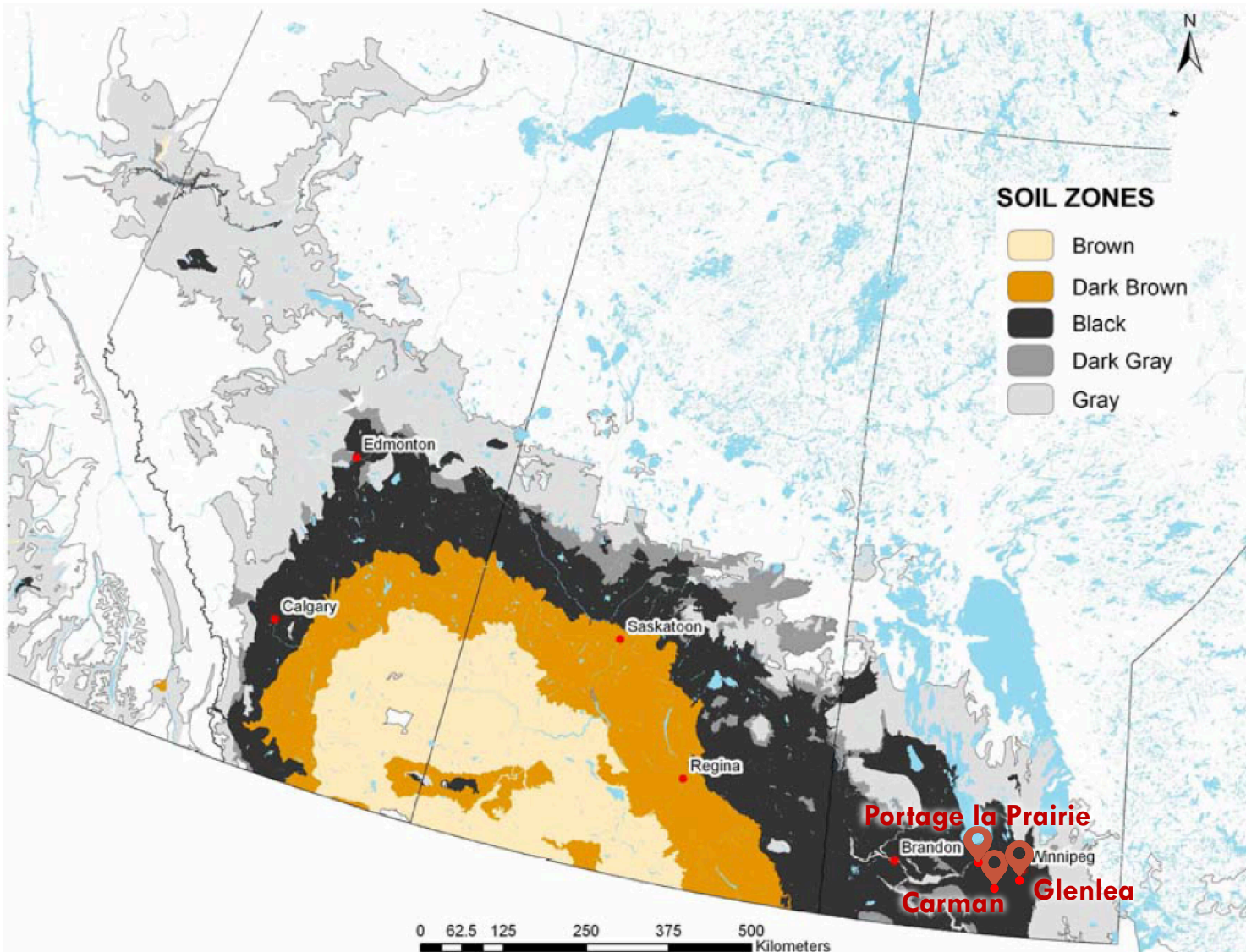
Jia Sun, Erin Higgins, Robert W. Duncan

Macdonald, Manitoba

University of Manitoba

Background

- Canadian Prairie soil zones

- GS in *B. napus*

  - *Oil content, Plant Height, Gluc Content (Werner et al., 2018)*

  - *Testcross Performance (Jan et al., 2016)*

Image: Les Fuller, The soil zones of the Prairie Provinces 2010

Objectives

- Analyze population structure of the training population (TP)

- Association analysis on agronomic traits

- TP effects on genomic selection prediction accuracy

Phenotypic data collection

- 92 parents (31 B lines, 61 R lines)
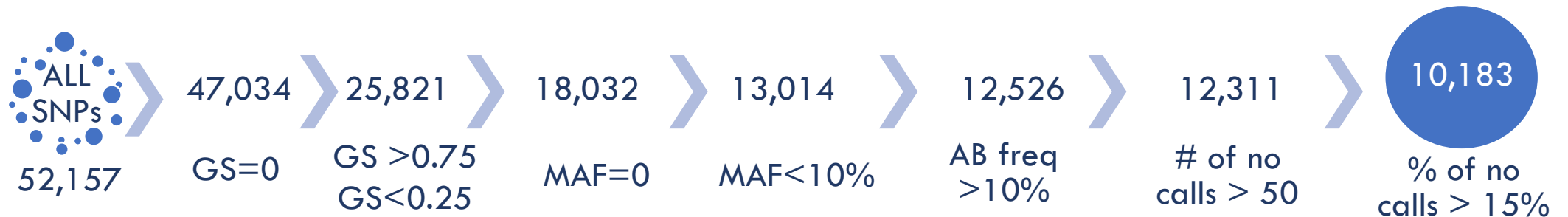- 5 site years across southern Manitoba, Canada

Year	Location	Traits evaluated	Notes
2016	Glenlea	flr, mat, ht, yld, pc, oc, gluc	Double-row nursery plots
2017	Carman	flr, mat, ht, yld, pc, oc, gluc	10 m^2 plots
2017	Portage la Prairie	flr, mat, ht, yld, pc, oc, gluc	10 m^2 plots
2018	Portage la Prairie	flr, mat, ht, yld, pc, oc, gluc	10 m^2 plots
2018	Glenlea	flr, mat, ht, yld, pc, oc, gluc	10 m^2 plots

UNIVERSITY of Manitoba

Data analysis

- SAS Studio (UNIVERSITY EDITION 2.8 9.4 M6)

  Phenotypic data diagnostics, ANOVA, LS means calculation

- R Studio (V. 1.2.1335)

  PCA analysis, Cluster analysis, Association analysis, rrBLUP

- GenomeStudio 2.0

  Marker filtration

Genotypic data

- All 92 TP individuals have been genotyped with *Brassica* 60K Illumina Infinium SNP array (AAFC Saskatoon, Canada)
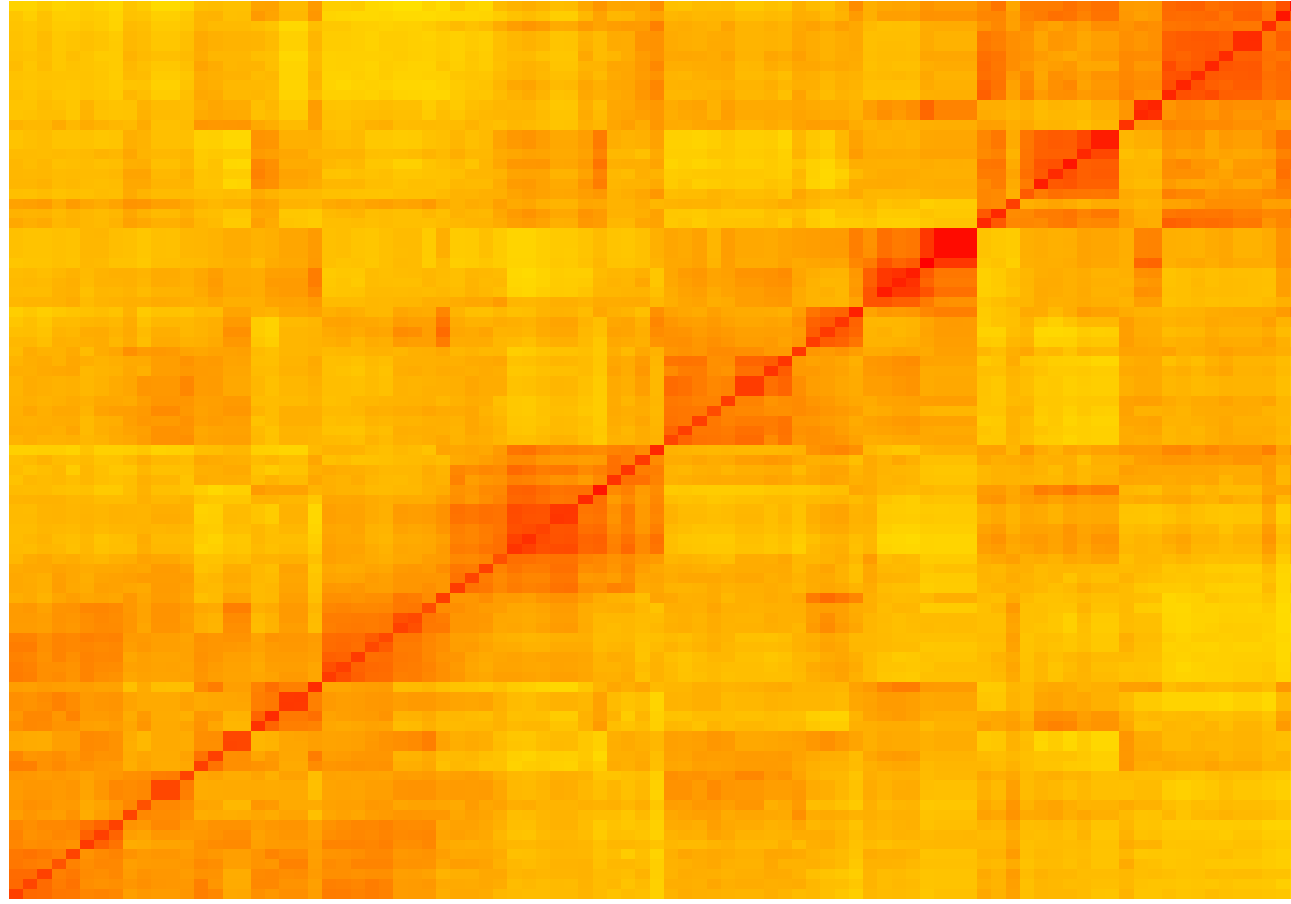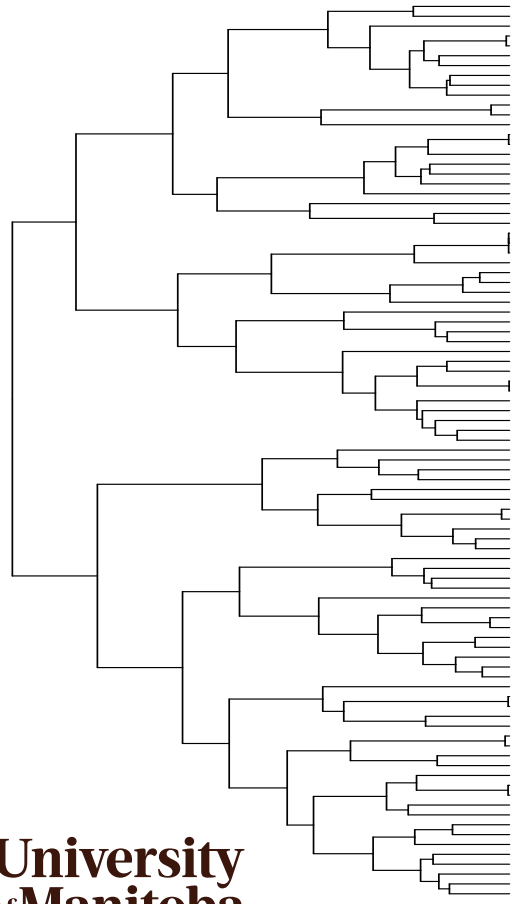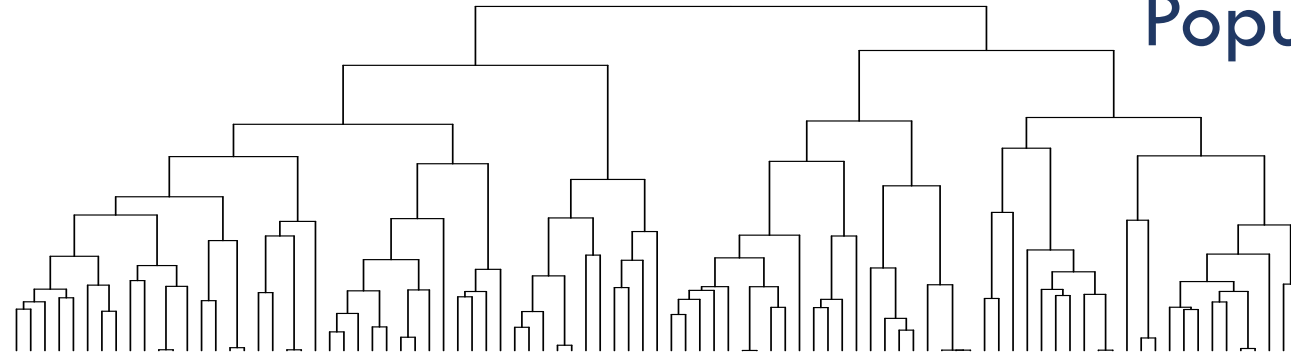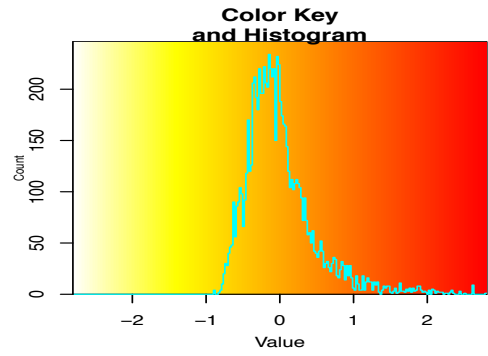
- Training population:

ALL SNPs	47,034	25,821	18,032	13,014	12,526	12,311	10,183
52,157	GS=0	GS >0.75 GS<0.25	MAF=0	MAF<10%	AB freq >10%	# of no calls > 50	% of no calls > 15%

- GS model: rrBLUP

Yield

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Genotype	91	97511556	1071556	6.24	<.0001
Block(SiteYear)	10	7697077	769708	10.9	<.0001
SiteYear	4	124212464	31053116	36.3	<.0001
SiteYear*Genotype	364	62956719	172958	2.45	<.0001
Residual	923	65165831	70602		

Population structure

R lines
B lines
R lines
B lines
R lines
B lines

R lines

B lines

R lines

B lines
R lines

PCA analysis

	PC1	PC2	PC3
Standard deviation	2.01	1.16	0.78
Proportion of Variance	0.58	0.19	0.09
Cumulative Proportion	0.58	0.77	0.86

	PC1	PC2	PC3
Yield	0.86	-0.17	-0.19
Flowering time	-0.68	-0.58	0.06
Maturity time	-0.69	-0.54	-0.09
Plant height	0.48	-0.80	0.09
Protein content	-0.87	0.02	-0.36
Oil content	0.92	-0.15	0.25
Glucosinolate content	-0.73	0.11	0.60
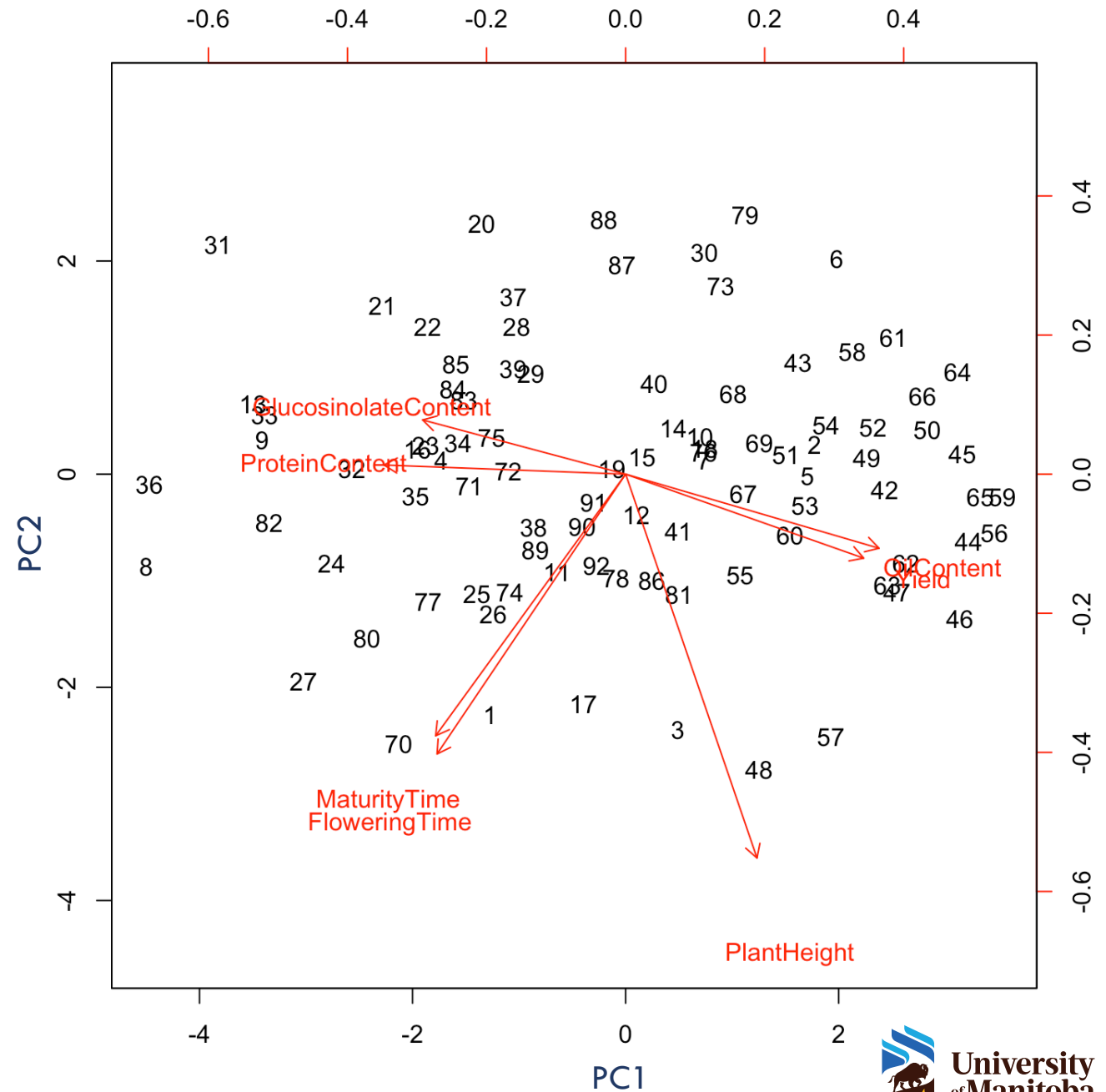
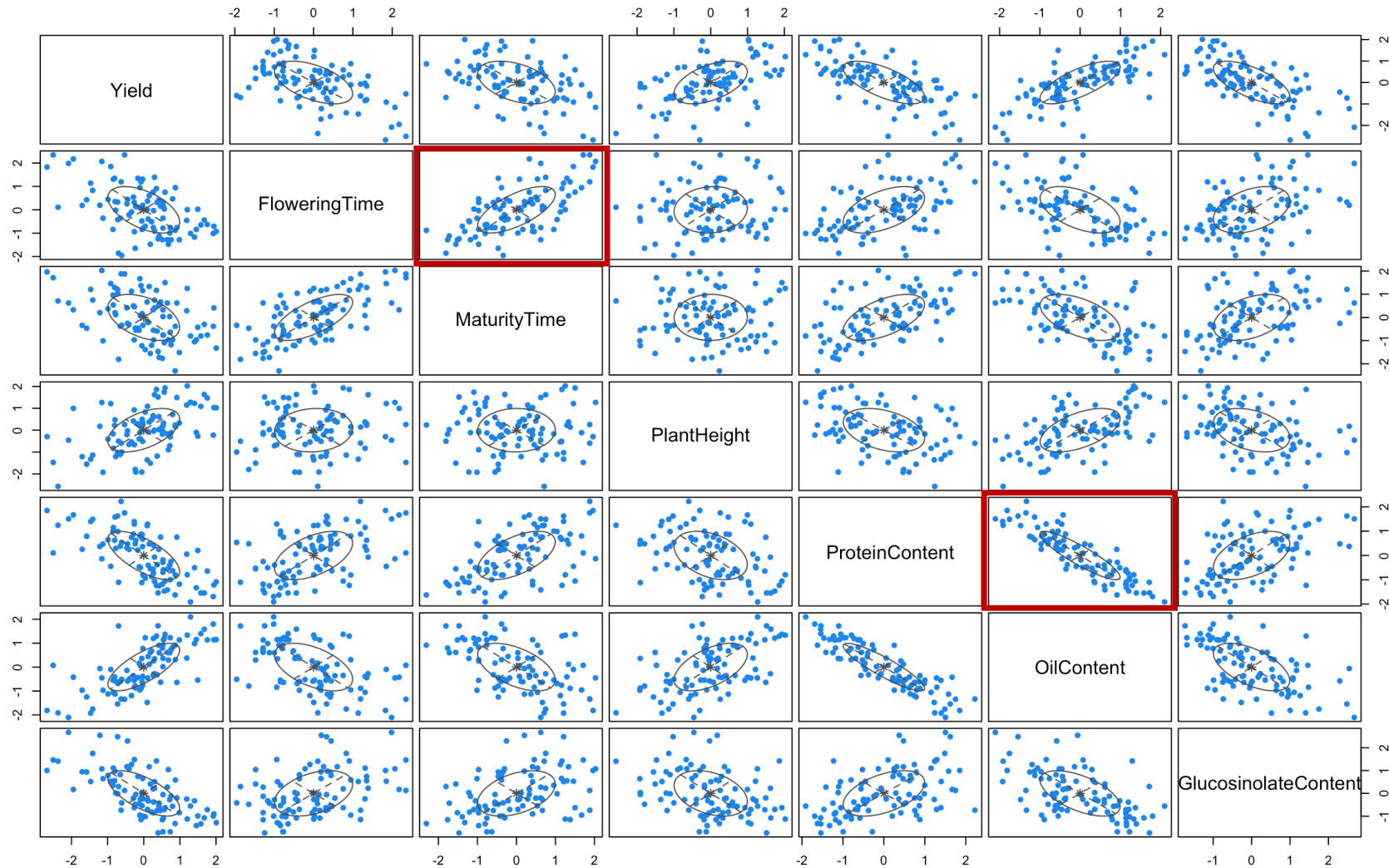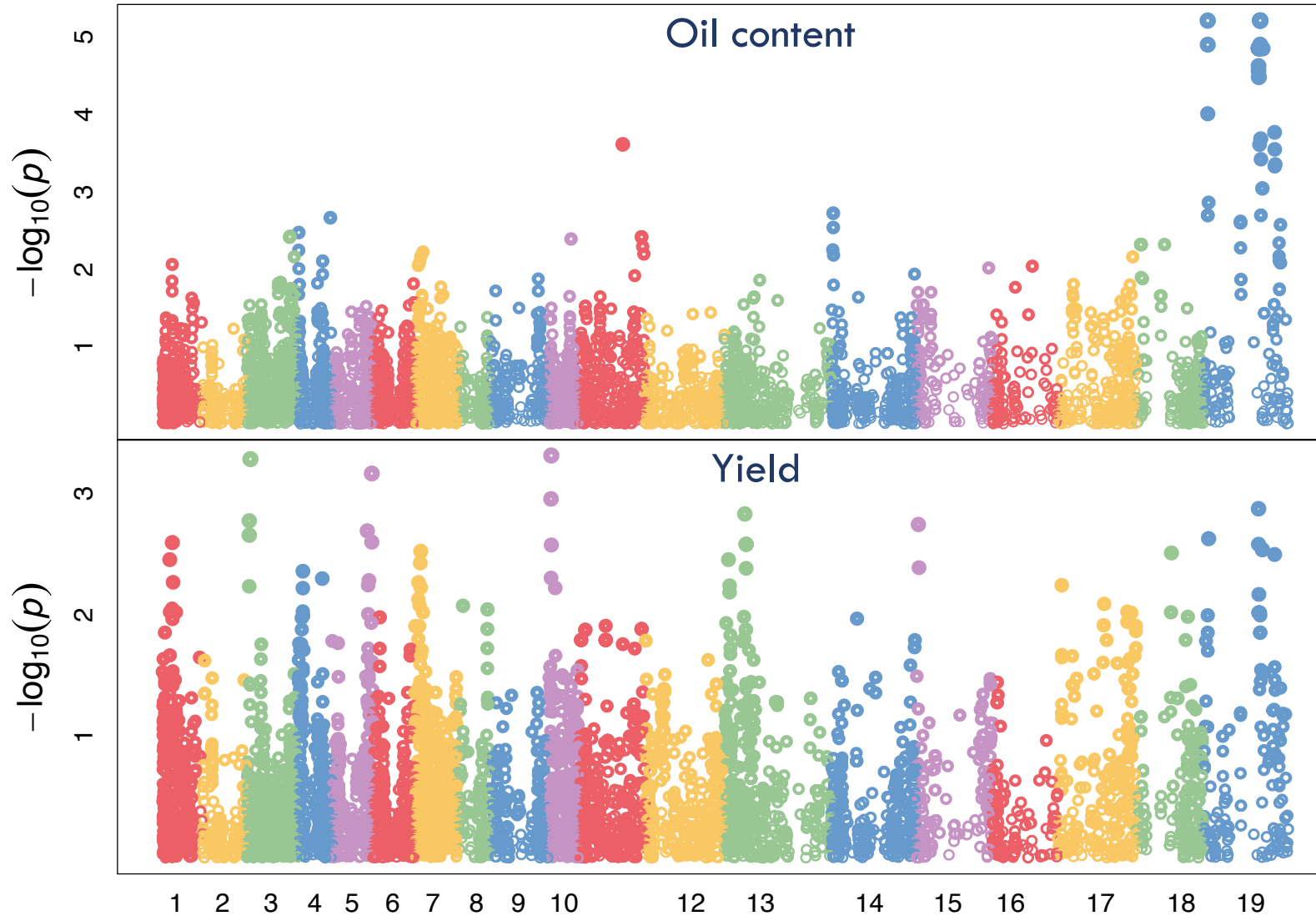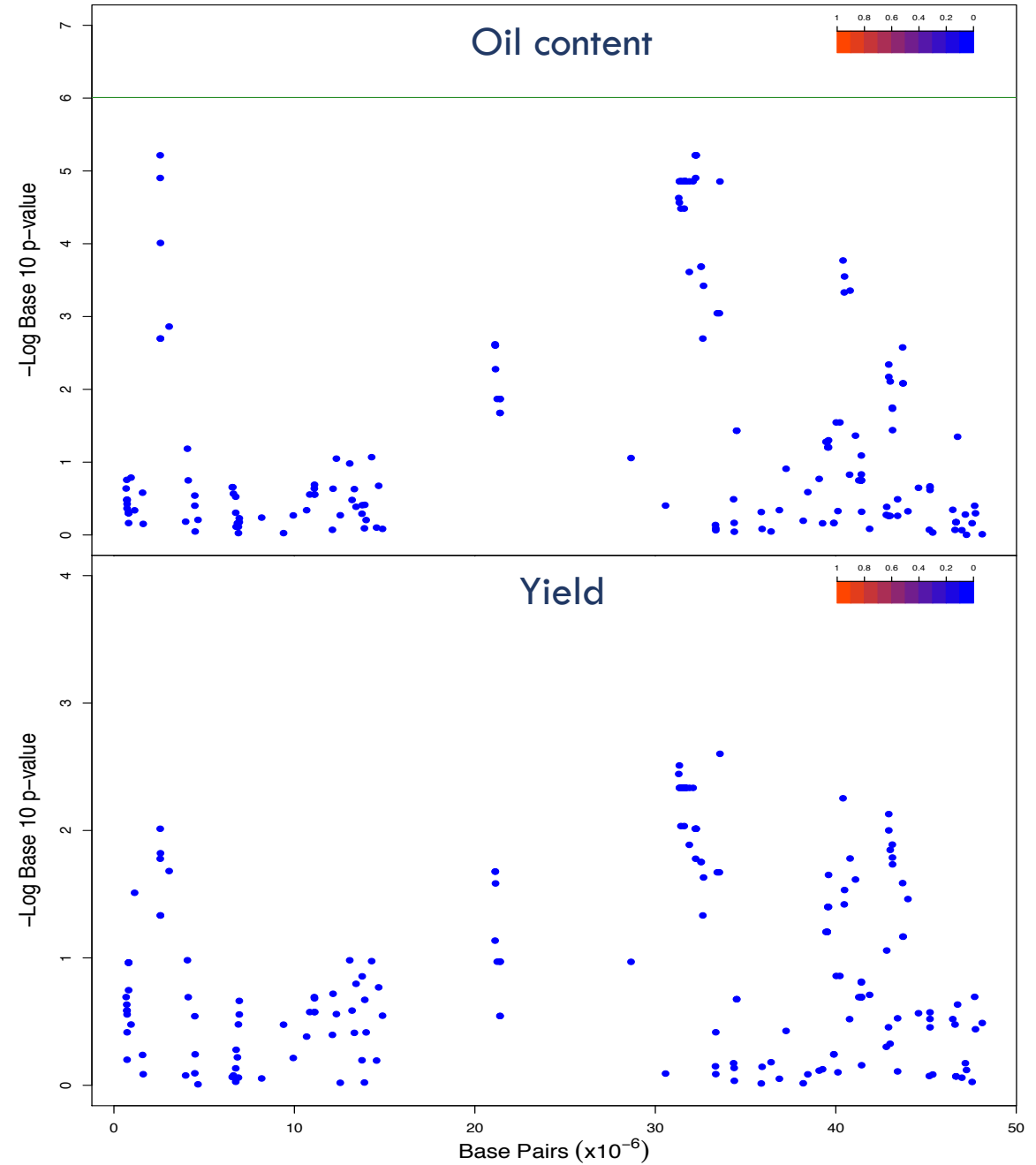# PCA analysis
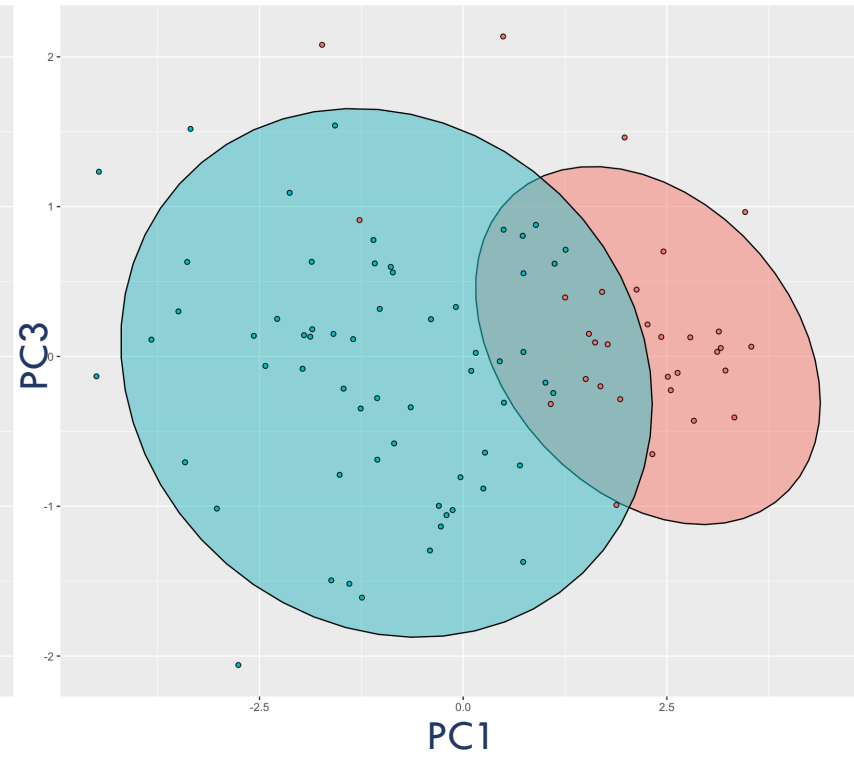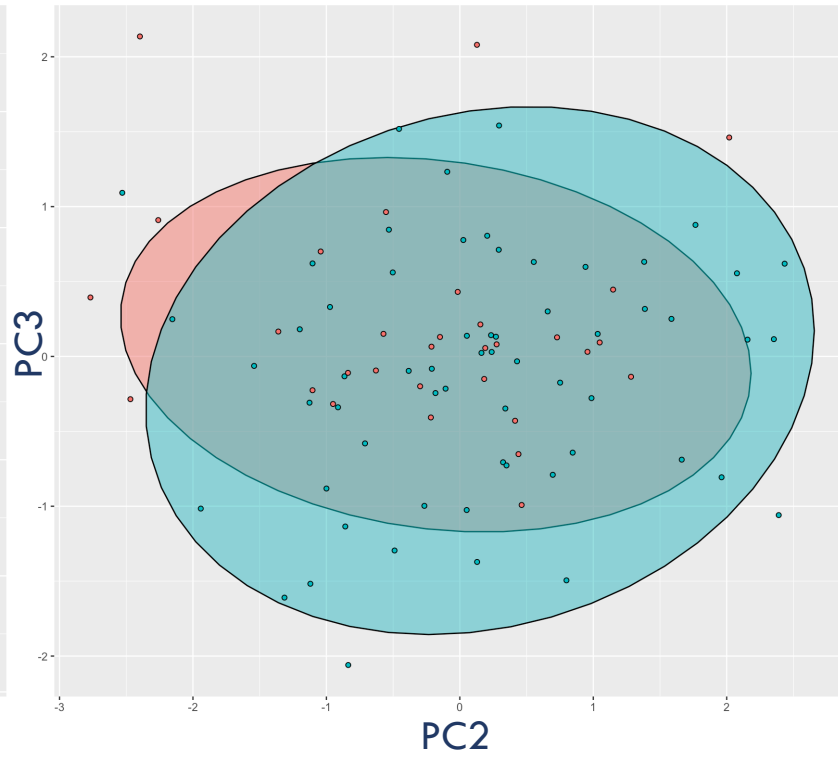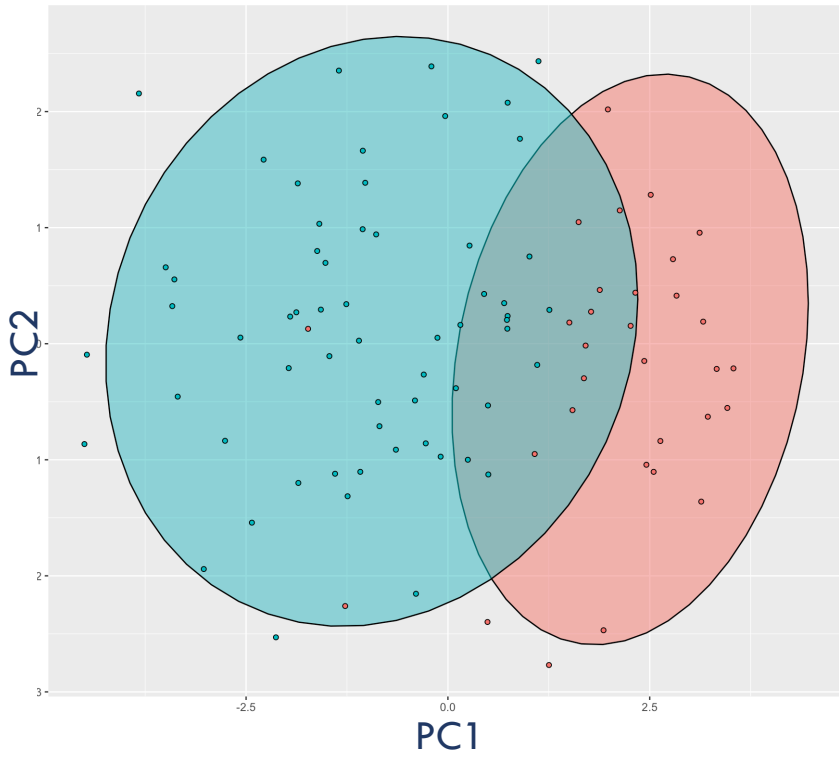
Cluster analysis: Model-based clustering

Association analysis

Chromosome 19

rrBLUP

- Prediction accuracy: correlation between observed & predicted values

Iterations	60/40			80/20		
	1	500	1000	1	500	1000
Yield	0.14	0.32	0.33	0.18	0.36	0.37
Flowering Time	0.12	0.22	0.23	0.69	0.25	0.26
Maturity Time	0.49	0.40	0.39	0.47	0.39	0.40
Plant Height	0.54	0.48	0.47	0.59	0.52	0.52
Oil Content	0.55	0.61	0.61	0.56	0.62	0.63
Protein Content	0.64	0.66	0.65	0.75	0.67	0.68
Glucosinolate Content	0.48	0.44	0.45	0.47	0.50	0.48

University of Manitoba

Future work

- Analyses on the hybrid population

    383 hybrids, 20+ site years

- Training different GS models

- Marker density effects

Acknowledgements

Carman, Manitoba 2017

Thank you!

P.C. AP Fei Xu