Development of a Statistical Framework for Association Mapping in Recurrently Selected Populations

Yogasudha Veturi
Department of Plant and Soil Sciences
University of Delaware
29th June 2012

INTRODUCTION

THEORY AND APPROACHES

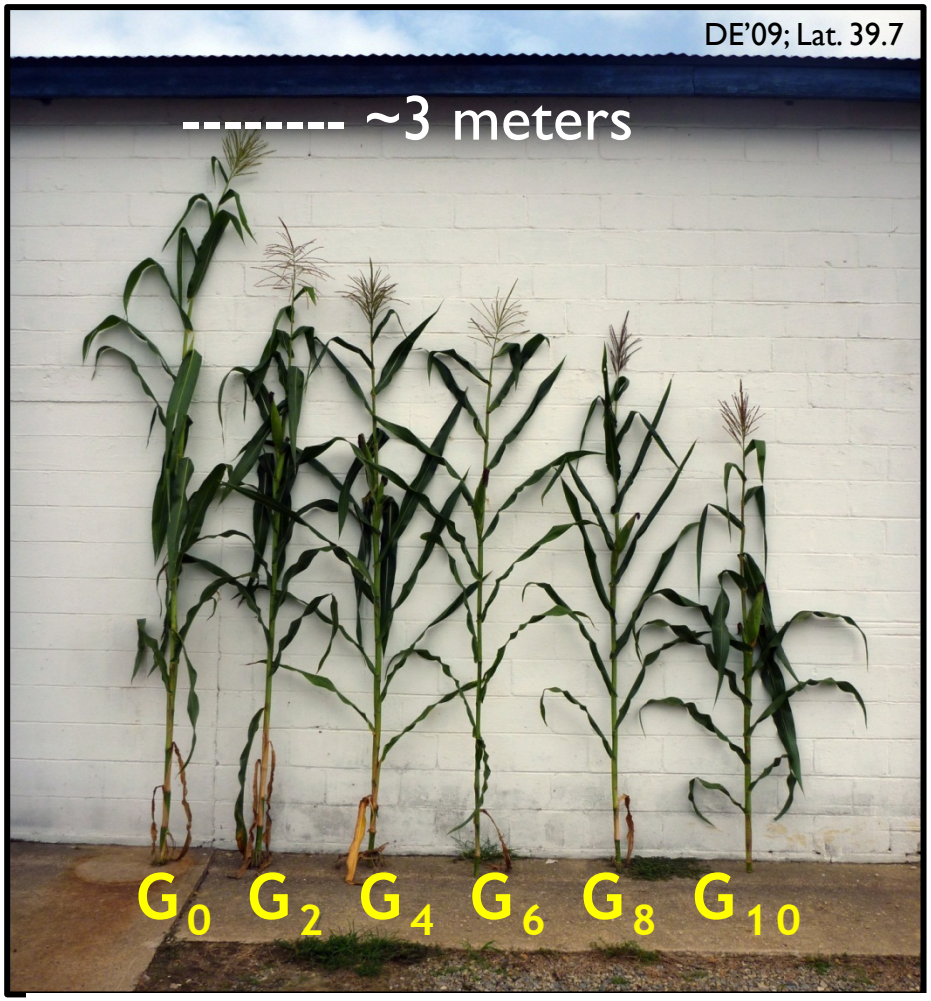
COANCESTRY IN RECURRENTLY SELECTED POPULATIONS

ASSOCIATION MAPPING IN RECURRENTLY SELECTED POPULATIONS

Response to artificial selection

Quantitative traits vary on a continuous scale, which if often attributable to the effects of multiple genes and their interaction with the environment.

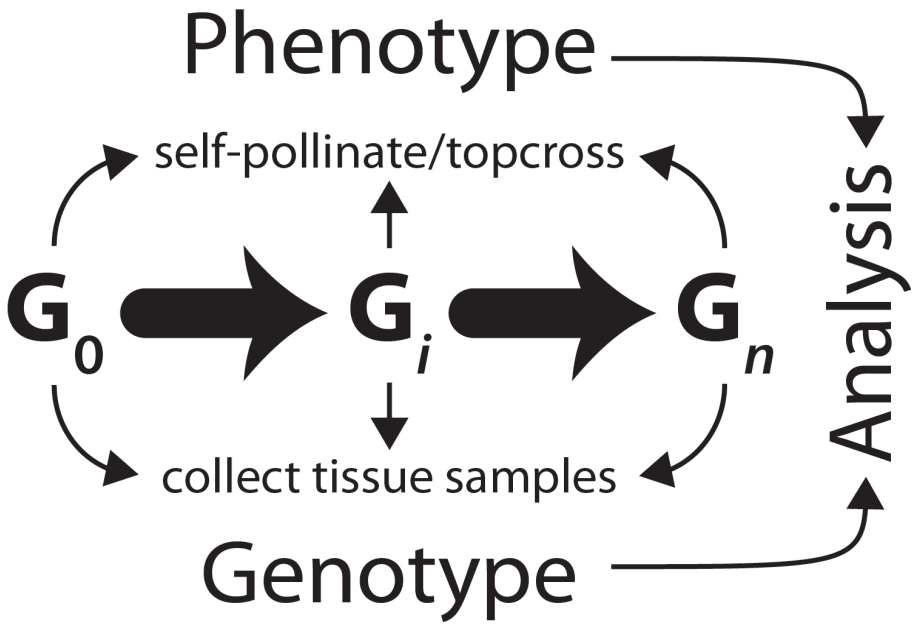
Recurrent selection is a common method of population improvement used for quantitative traits



A representative plant from cycles of Tusón in DE.

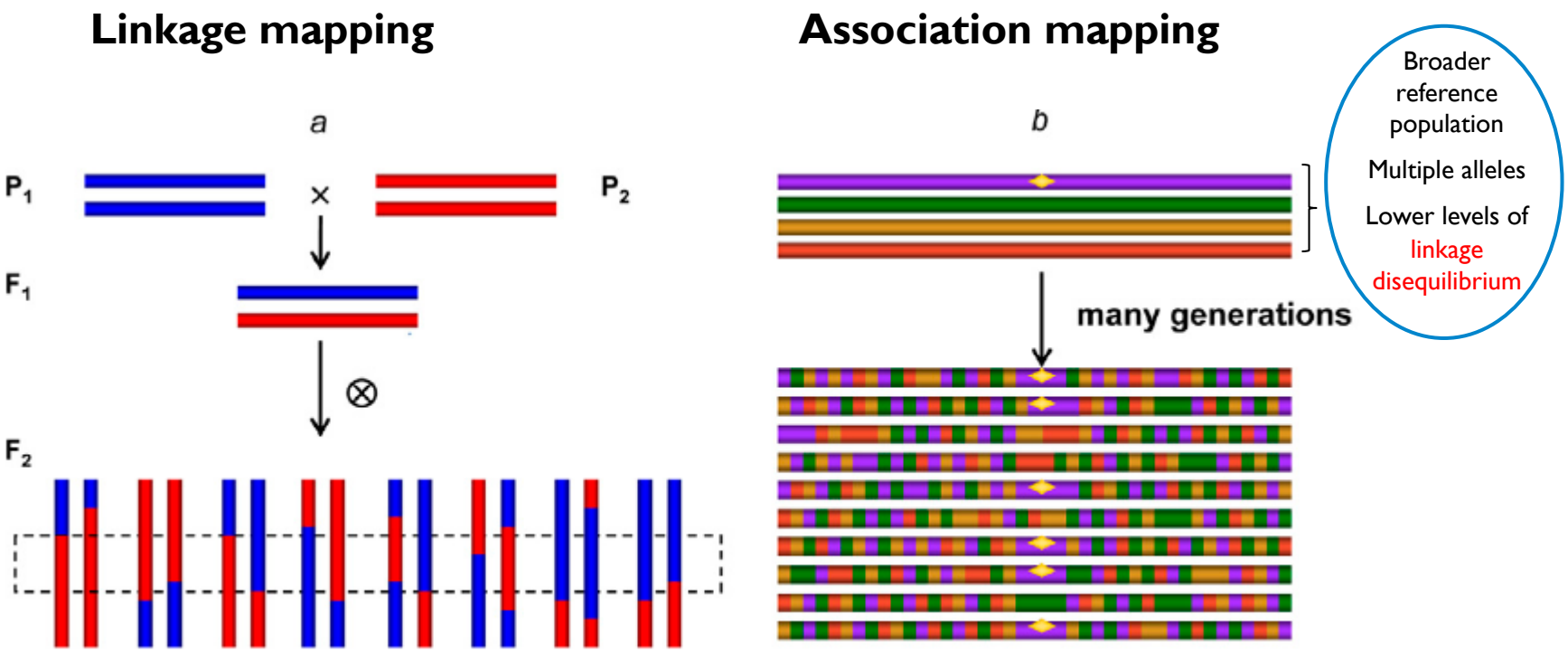
General approach and objective

- The genetic architecture underlying **response to selection** has not been widely characterized.
- We have developed a **new genetic design** combining association mapping and selection mapping in artificially selected maize populations.



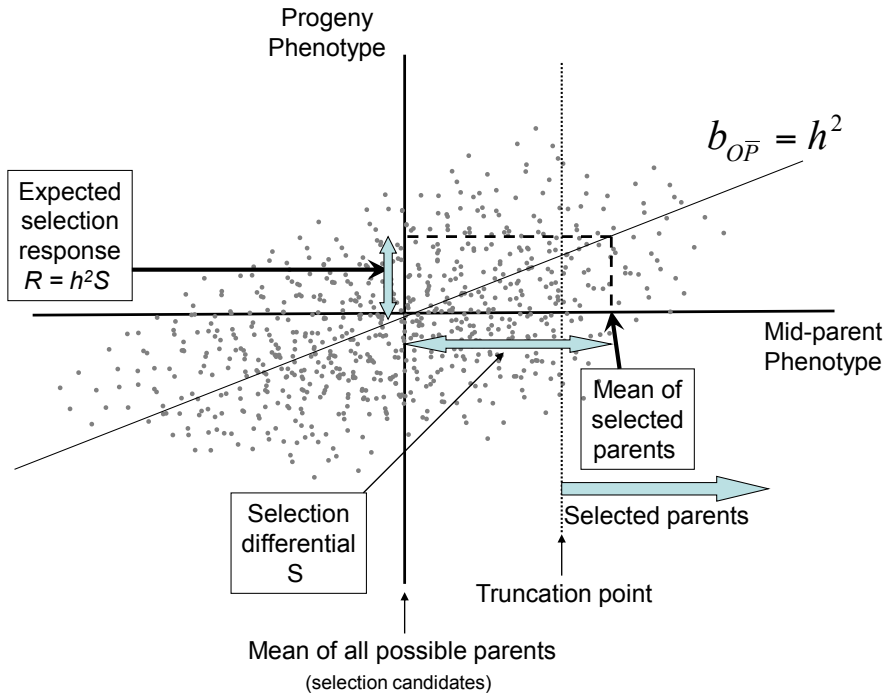
Wisser *et al.* (2011) Plant Gen Res 9(2):281–283

Quantitative trait locus mapping…



Zhu, C.G. et al., *The Plant Genome*. **1**, 5 (2008)

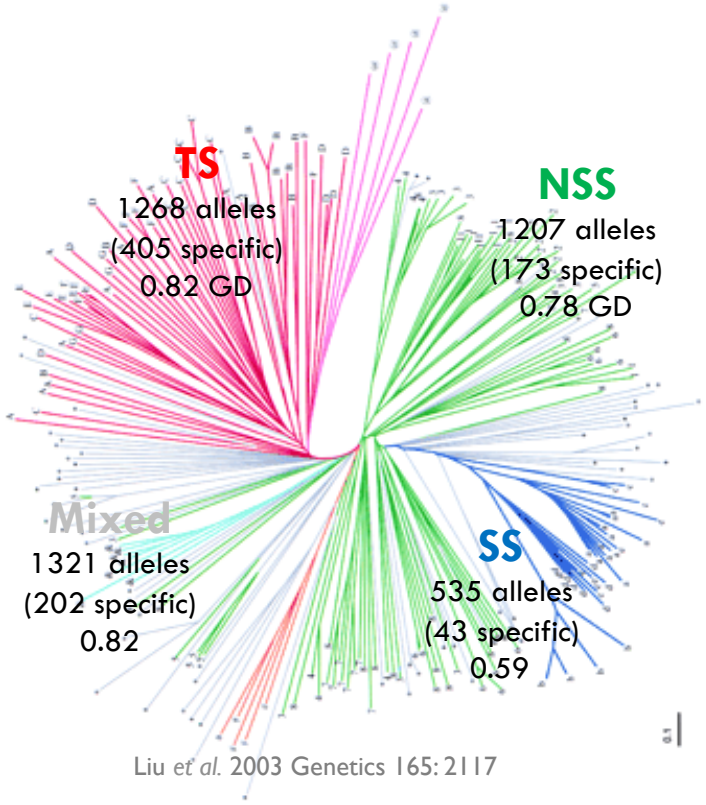
Understanding the genetic basis of selection response



http://www.public.iastate.edu/~jjannink/Teaching/2004/Lecture_points_11.doc

Recurrent selection increases mean trait value while maintaining genetic variability such that gains due to selection are sustained

Genetic relatedness in association studies



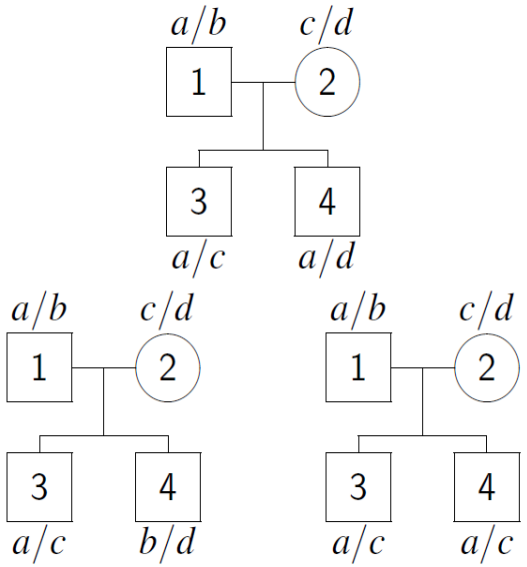
Liu *et al.* 2003 Genetics 165: 2117

Individuals in association panels are not genetically "independent!"

Not accounting for genetic relatedness leads to biased allele effect estimates and a high degree of false positive associations

Coancestry

One measure of coancestry is the probability of *identity by descent* (IBD).

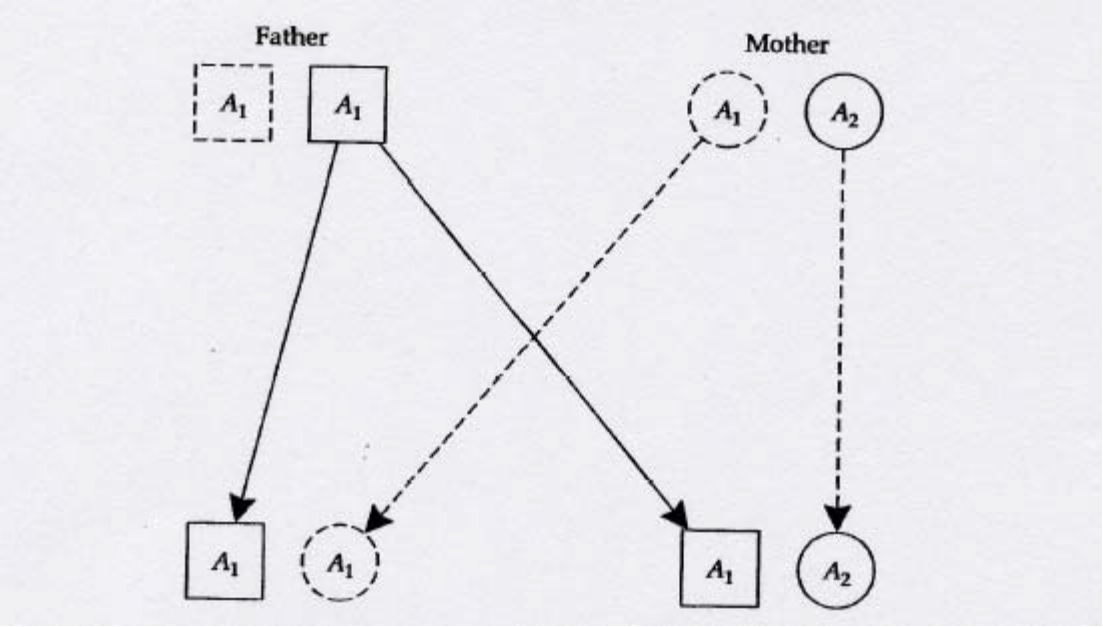


- Subjects 3 and 4 share one gene IBD (the paternal allele, a)
- But, in these families, they share respectively 0 and 2 genes IBD

Florence: IBD, 2005

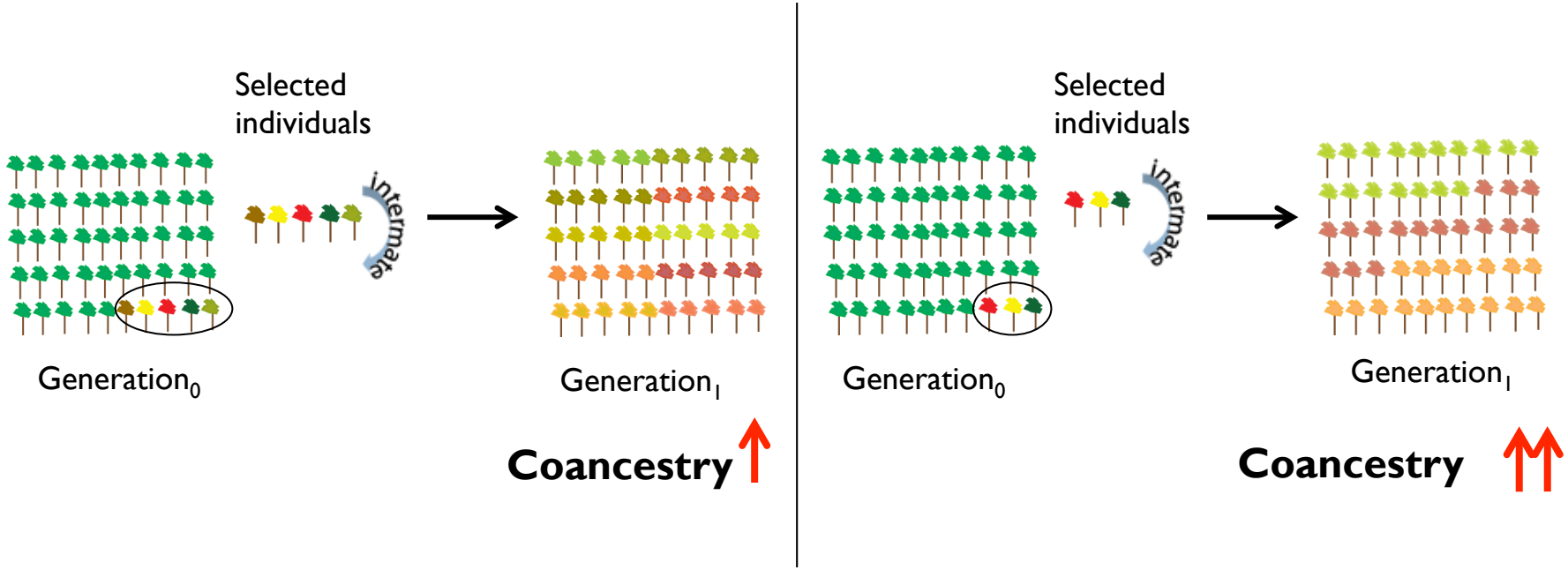
Identity by descent vs. identity by state: implications for estimating coancestry

Molecularly identical alleles are IBS, leading to upwardly biased estimates of coancestry



Lynch M, Walsh B(1998) Genetics and analysis of quantitative traits.

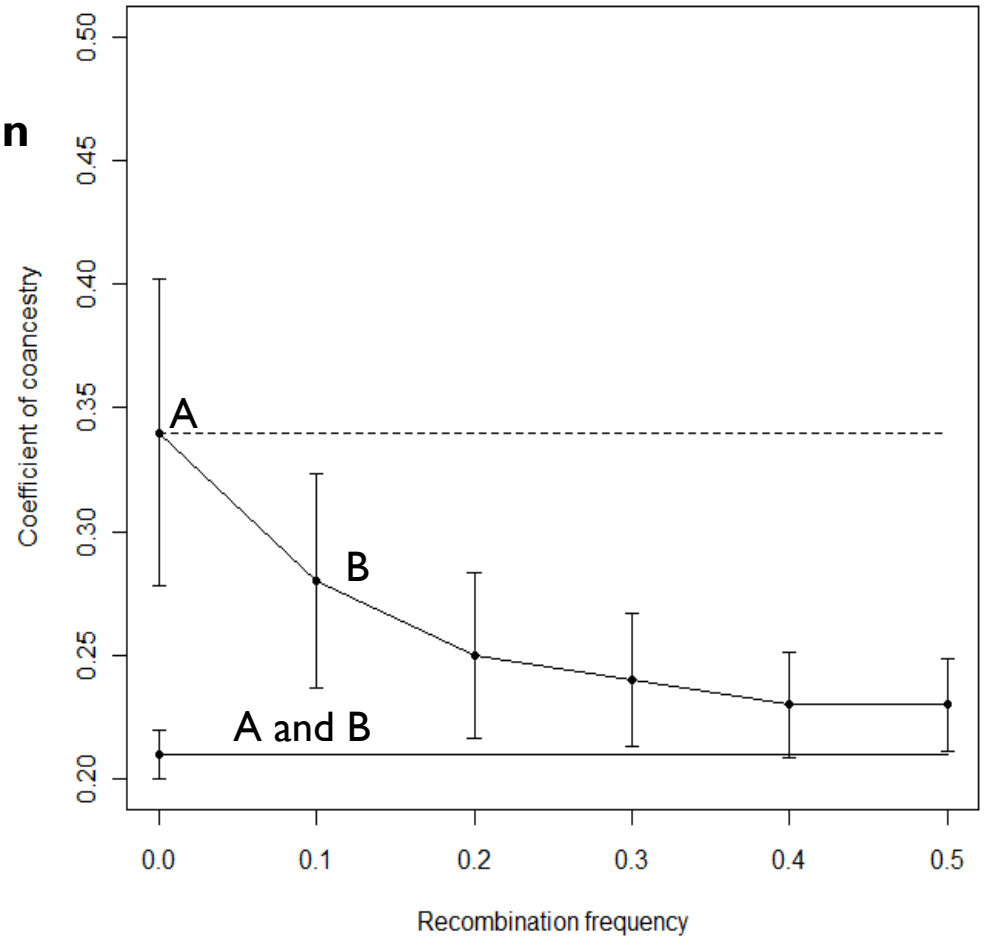
Coancestry in populations under recurrent selection…



The smaller the ratio of the number of individuals selected in G_0 to the number of individuals grown in G_1, the greater the extent of coancestry in G_1

Impact of selection on coancestry

2-locus simulation
Generations: 3
Intensity: 10%
Pop size: 500



Selected locus (A)

Linked locus (B)

A and B under pure drift

Thesis hypotheses and objectives

- A new weighted estimator for coancestry called the quantitative identity by state (QIBS) is a precise and unbiased estimator of coancestry among individuals in recurrently selected populations.

- To develop a whole genome computer simulation program to examine the statistical properties (accuracy and precision) of the QIBS estimator relative to other estimators (IBS and Loiselle) when applied to population samples subjected to recurrent selection.

Thesis hypotheses and objectives

- Recurrently selected populations are suitable for association mapping.

- To use the simulation program developed for objective 1 to examine how different estimators of coancestry affect the statistical power and type I error rate of the test statistic for association mapping in recurrently selected populations.

INTRODUCTION

THEORY AND APPROACHES

COANCESTRY IN RECURRENTLY SELECTED POPULATIONS

ASSOCIATION MAPPING IN RECURRENTLY SELECTED POPULATIONS

Estimation of coancestry – existing estimators

- Conventional estimators of coancestry:
  - Assumes population is generated from unrelated founders.
  - Assumes randomly mating populations in Hardy-Weinberg equilibrium
  - Produces estimates outside the [0,1] probabilistic range, requiring rescaling or truncation while model fitting
  - *E.g. Lynch and Ritland (1999), Thompson's MLE (1974), Bernardo's estimator (1993).*
- Maenhout et al. (2009) devised an estimator that was positive-semi-definite and and applicable to population samples from a hybrid breeding program under artificial selection.

Estimation of coancestry - QIBS

We have developed a new weighted estimator for coancestry called the **quantitative identity by state** estimator (QIBS), which:

- accounts for relatedness among the founders
- is positive-semi-definite
- attempts to correct the inflation in coancestry due to IBS

Quantitative IBS (QIBS) estimator

$$QIBS = \frac{XWX'}{4l} + Q$$

Given *m* diploid individuals in a population and *p* alleles observed in a collection of genotypes across *l* loci,
X is an incidence matrix of size *m x p* indicating the presence/absence of an allele at a locus in an individual; each row takes 1 for a heterozygous allele and 2 for a homozygous allele
Q is the diagonal matrix of inbreeding coefficients

Calculation of weight matrix W

- The weights for QIBS are estimated as:

$$\omega_z = \frac{\sum_{a=1}^{n}\sum_{b=1}^{n} R_{z(a,b)}}{\sum_{a=1}^{n}\sum_{b=1}^{n} M_{z(a,b)}}$$

where,

$$R_z = Q \circ M_z$$

This is the element-by-element (or Hadamard) product of matrices Q and M_z

Q is a $2n$ x$2n$ IBS matrix (n= no. of founders),
M_z (z=1, 2,… p) is an allele-specific relationship matrix, i.e.
$M_{z(a,b)}$ takes values 1 or 0 based on whether how many founders are IBS for allele z (z = 1,2, …, p).

Whole genome simulation based approach

Start with "unrelated" ancestors

Define chromosomal number and lengths, and marker spacing(in cM)

Create haplotype blocks

Define SNP effect distributions, and trait heritabilities, create phenotypes

Forward simulation

Select "related" founders as base of recurrently selection

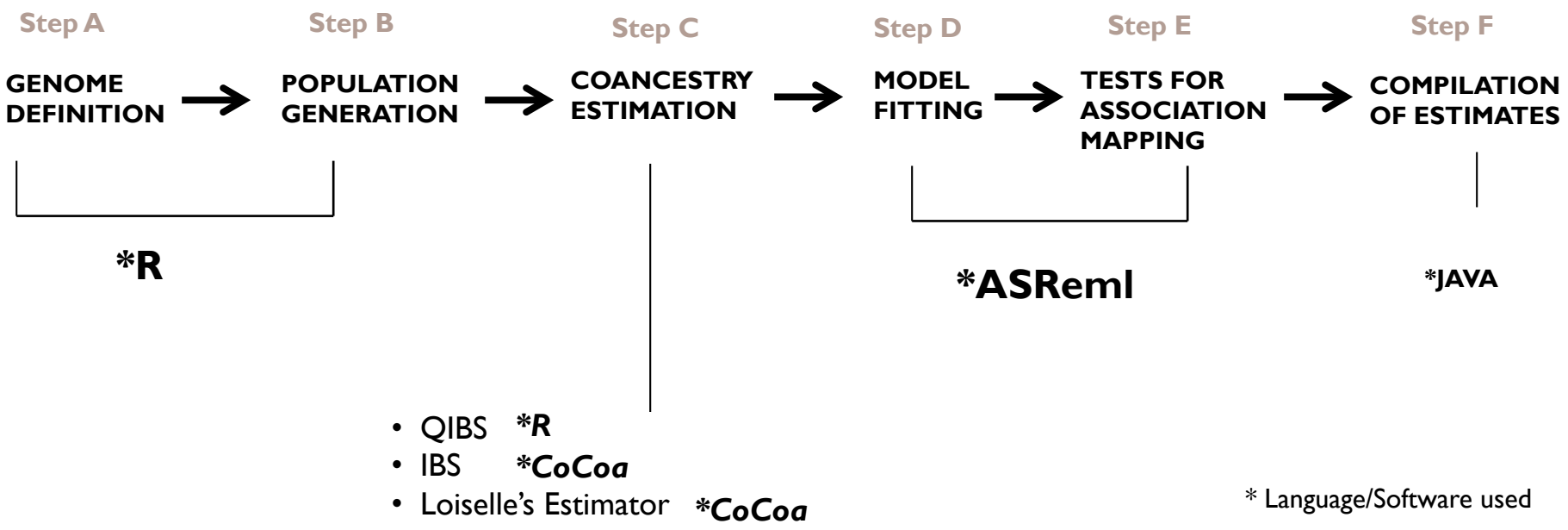
Start recurrent selection

Define **selection intensities**, number of **generations of selection and inter-mating**, and **sample sizes**

N generations of selection and inter-mating…

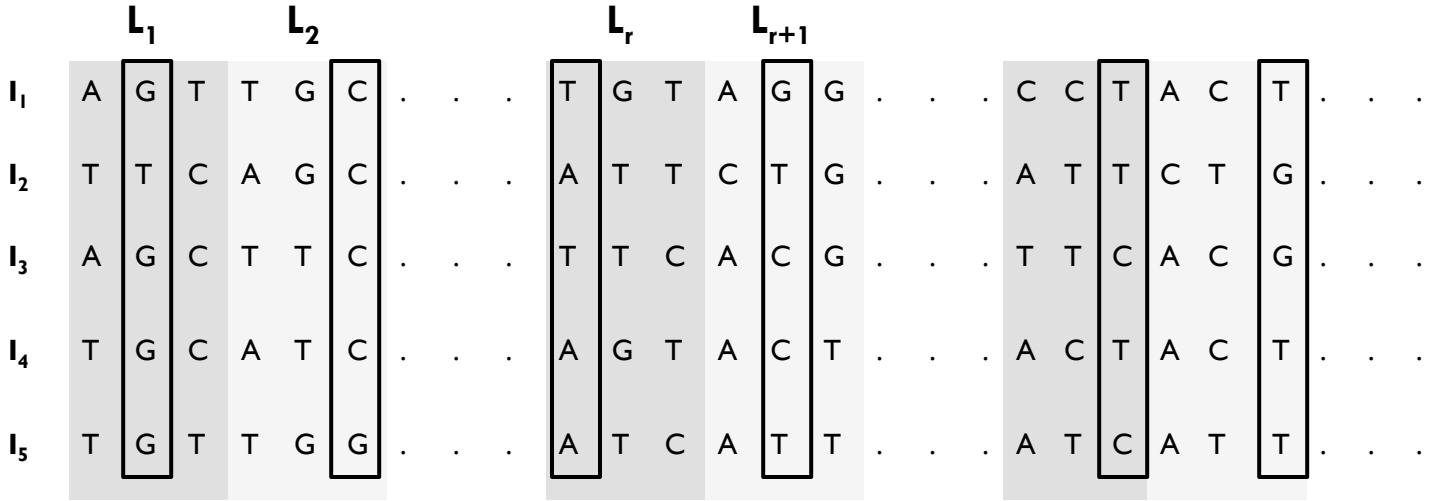
Final population for association mapping!

Whole genome simulation based approach

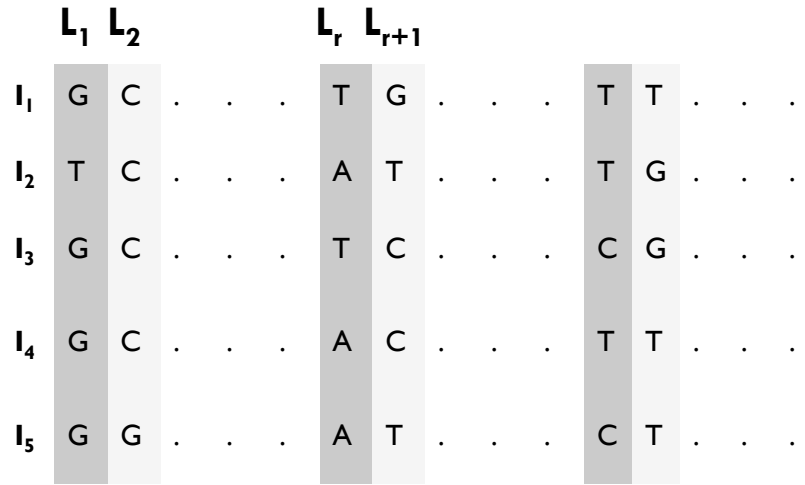


Allele convergence

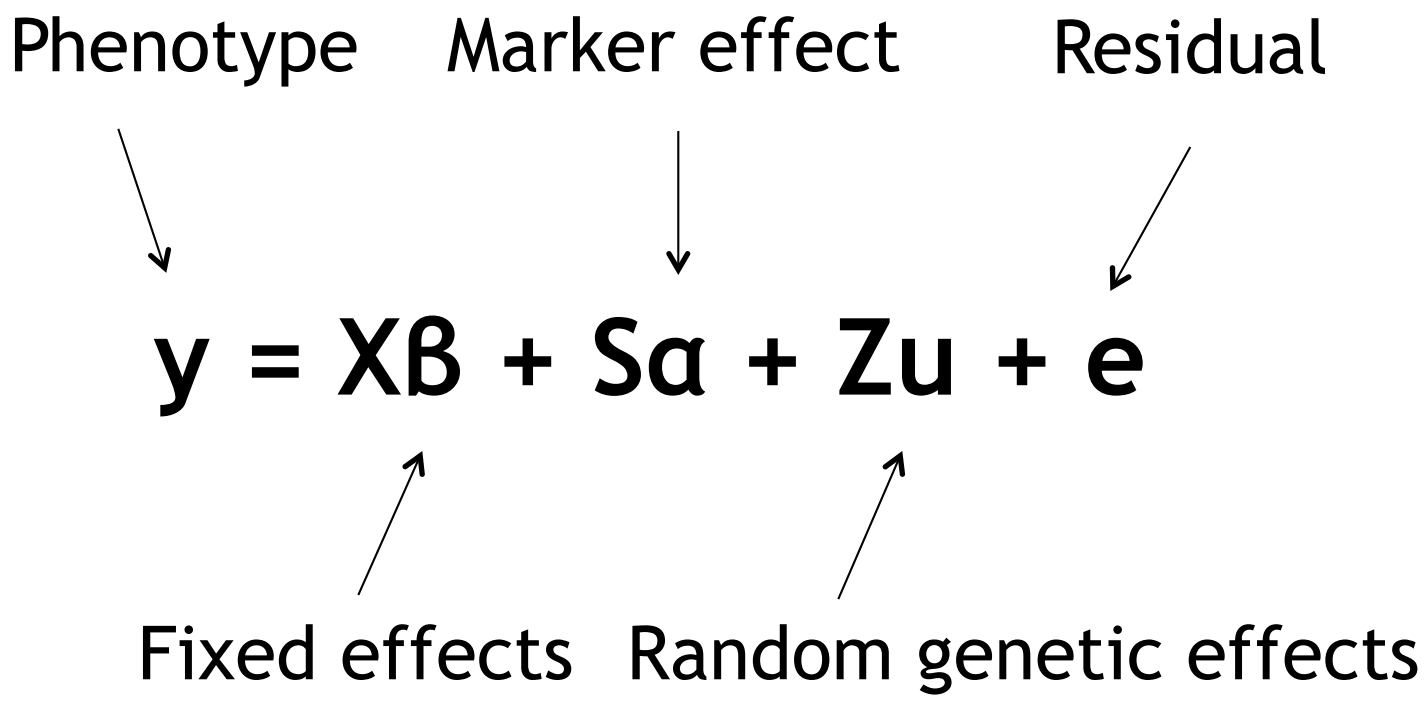
Multi-allelic model



Bi-allelic model



A mixed model for association mapping

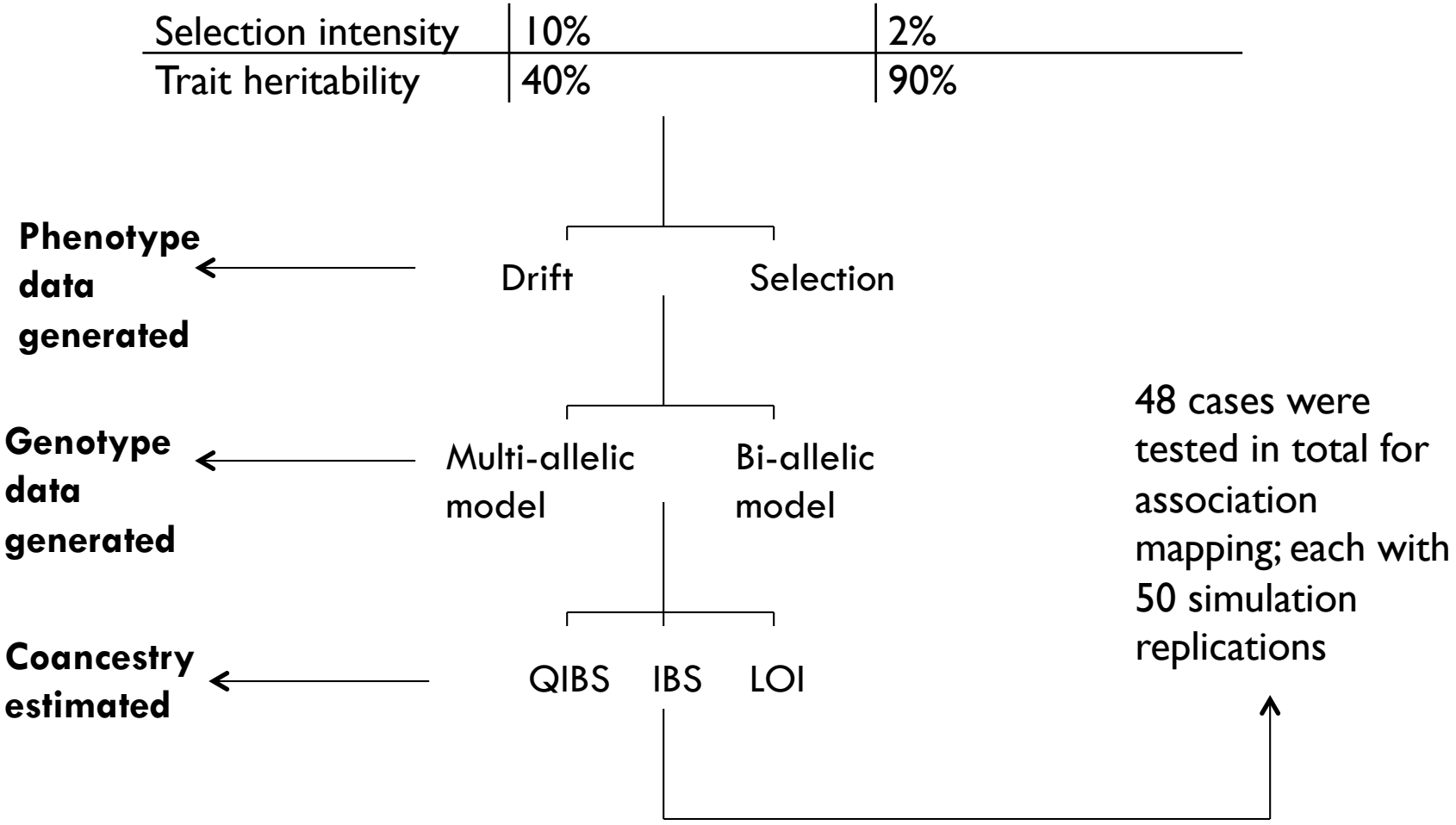


$\mathbf{Var(u) = 2KV_A}$ K=QIBS

V_A = Additive genetic variance

K= Coancestry matrix

Simulation cases



INTRODUCTION

THEORY AND APPROACHES

COANCESTRY IN RECURRENTLY SELECTED POPULATIONS

ASSOCIATION MAPPING IN RECURRENTLY SELECTED POPULATIONS

Coancestry comparisons

- Coancestry comparisons made between QIBS, IBS, and LOI using simulated genotype data
- Impact of the following factors was examined on coancestry
  - drift and selection
  - selection intensity and trait heritability
  - allele convergence

RMSE with QIBS at 10% selection

Selection intensity =10%			Multi-allelic model		Bi-allelic model	
			Drift	Selection	Drift	Selection
Heritability	40	IBS	0.000[a]	0.000[a]	0.233	0.224
		QIBS	0.015	0.025	0.178	0.160
	90	IBS	0.000[a]	0.000[a]	0.231	0.217
		QIBS	0.018	0.033	0.174	0.146

[a]Expected correlation with multi-allelic IBS

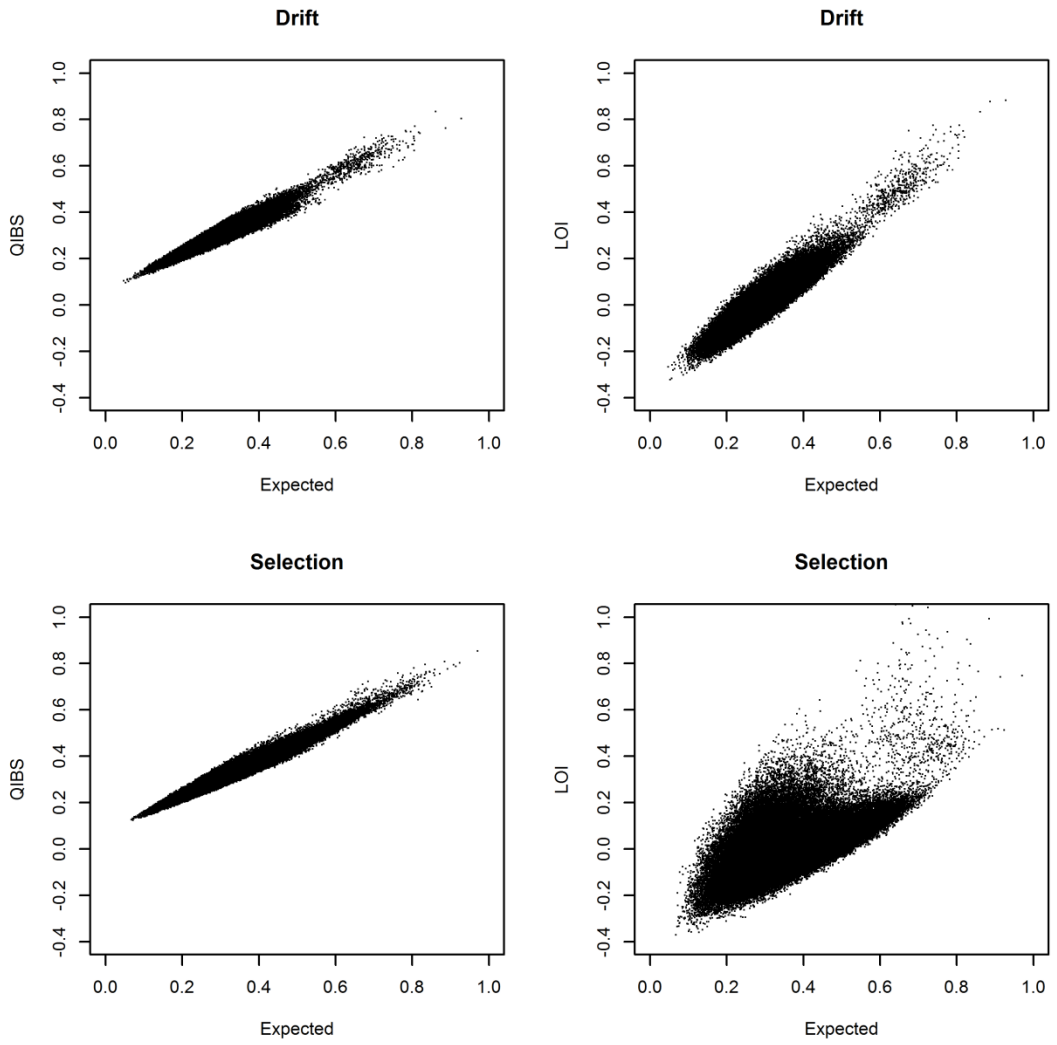
Correlations between IBS, QIBS, and LOI with the expected

Selection intensity =10%			Multi-allelic model		Bi-allelic model	
			Drift	Selection	Drift	Selection
Heritability	40	IBS	1.000±0.000[a]	1.000±0.000[a]	0.858±0.002	0.908±0.002
		QIBS	0.955±0.001	0.969±0.001	0.869±0.002	0.914±0.002
		LOI	0.861±0.005	0.724±0.005	0.758±0.004	0.647±0.004
	90	IBS	1.000±0.000[a]	1.000±0.000[a]	0.876±0.003	0.937±0.001
		QIBS	0.959±0.001	0.979±0.000	0.884±0.003	0.940±0.001
		LOI	0.822±0.006	0.614±0.004	0.728±0.005	0.555±0.003

[a]Expected correlation with multi-allelic IBS

Comparison of QIBS and LOI with the expected

h^2=90%
i=10%
(multi-allelic model)



Summary

- Selection increased coancestry
- Heritability and selection intensity also increased coancestry
- QIBS gave biased estimates under allele convergence
- However, the bias was lower than IBS

Summary

- LOI had very low correlation with the expected, especially under selection; is not an appropriate estimator of coancestry for selected populations
- For real populations (where IBS always tends to inflate coancestry), QIBS is the best estimator among the three estimators considered here; provided founders are known

INTRODUCTION

THEORY AND APPROACHES

COANCESTRY IN RECURRENTLY SELECTED POPULATIONS

ASSOCIATION MAPPING IN RECURRENTLY SELECTED POPULATIONS

Comparisons of the association mapping framework

- Comparisons between the three estimators under drift and selection at different intensities and heritabilities were based on:

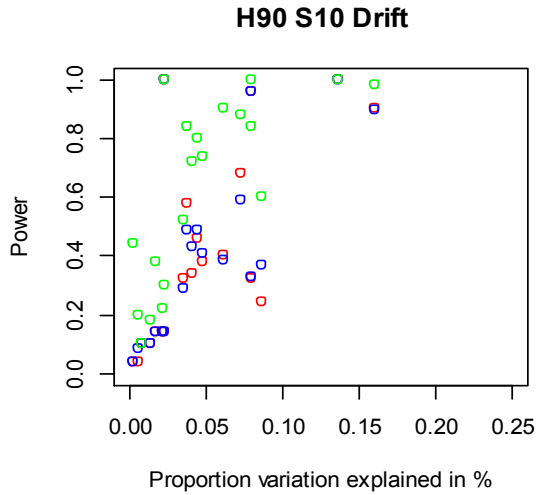
- Model fit
- Power of detecting QTL
- Type-I errors for neutral markers

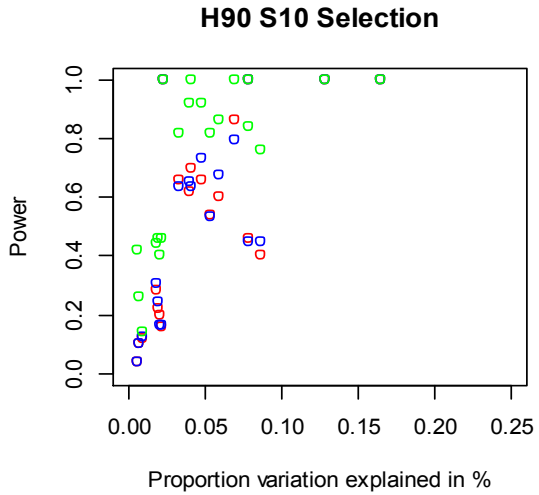
Comparisons of model fit

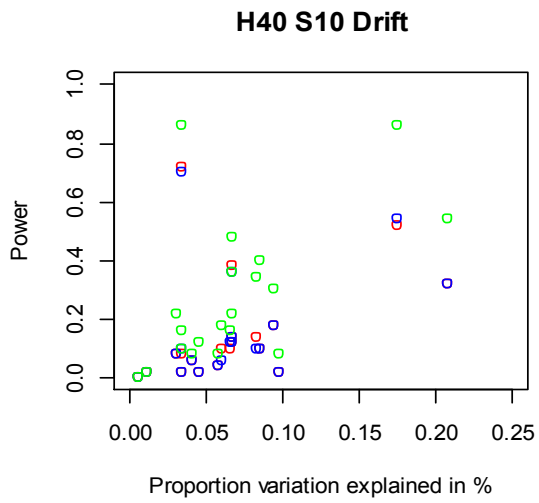
Selection intensity=10%			Multi-allelic model		Bi-allelic model	
			Drift	Selection	Drift	Selection
Heritability	40	IBS	-1283.2±0.24	-1226.5±0.21	-1286.0±0.25	-1228.9±0.22
		QIBS	-1283.8±0.24	-1227.2±0.21	-1285.1±0.25	-1228.1±0.24
		LOI	-1297.4±0.23	-1243.3±0.21	-1299.8±0.23	-1245.1±0.22
	90	IBS	-778.5±0.19	-696.1±0.18	-769.1±0.19	-692.1±0.21
		QIBS	-779.7±0.19	-697.1±0.18	-769.6±0.20	-692.4±0.21
		LOI	-879.7±0.18	-805.3±0.17	-863.5±0.18	-796.5±0.18

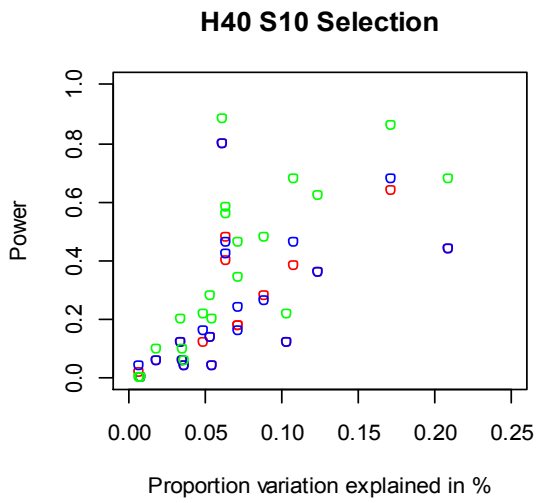
Comparisons of power

i=10%
(multi-allelic model)



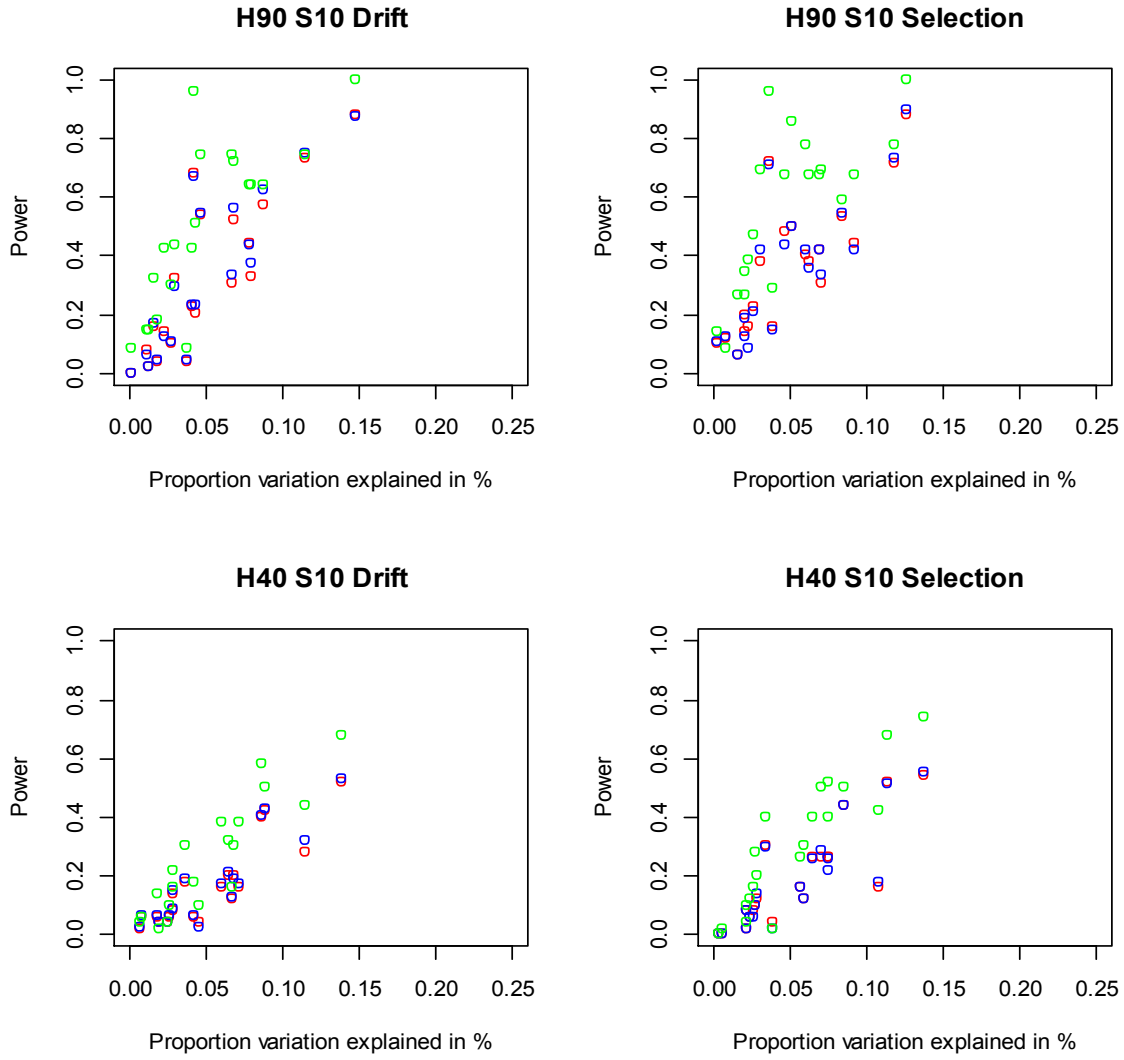






Comparisons of power

i=10%
(bi-allelic model)

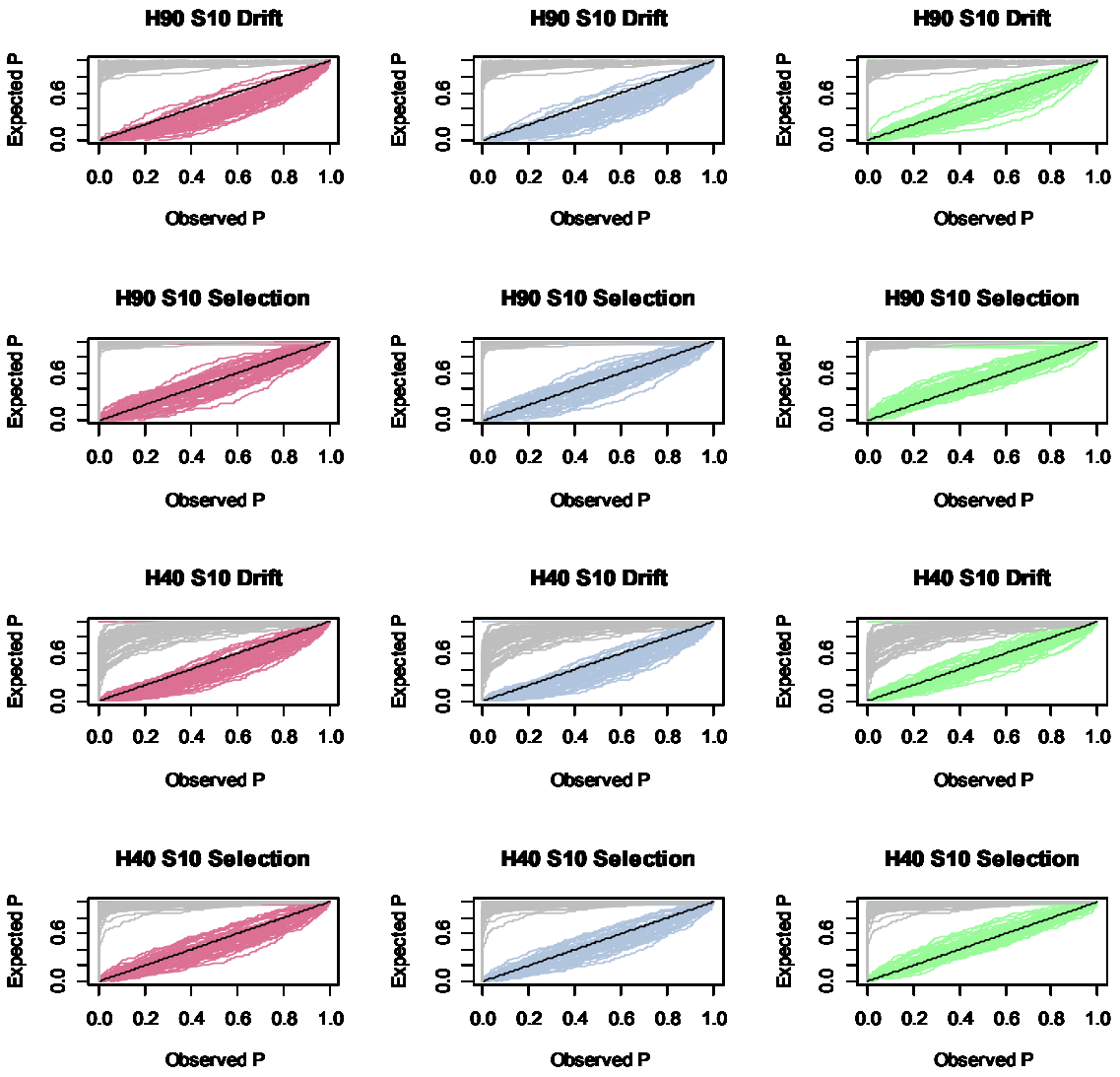


Comparisons of power : Proportion of QTL detected

Selection intensity=10%			Multi-allelic model		Bi-allelic model	
			Drift	Selection	Drift	Selection
Heritability	40	IBS	16%	24%	16%	17%
		QIBS	16%	25%	16%	16%
		LOI	28%	38%	25%	30%
	90	IBS	41%	53%	29%	34%
		QIBS	42%	54%	30%	35%
		LOI	63%	73%	49%	54%

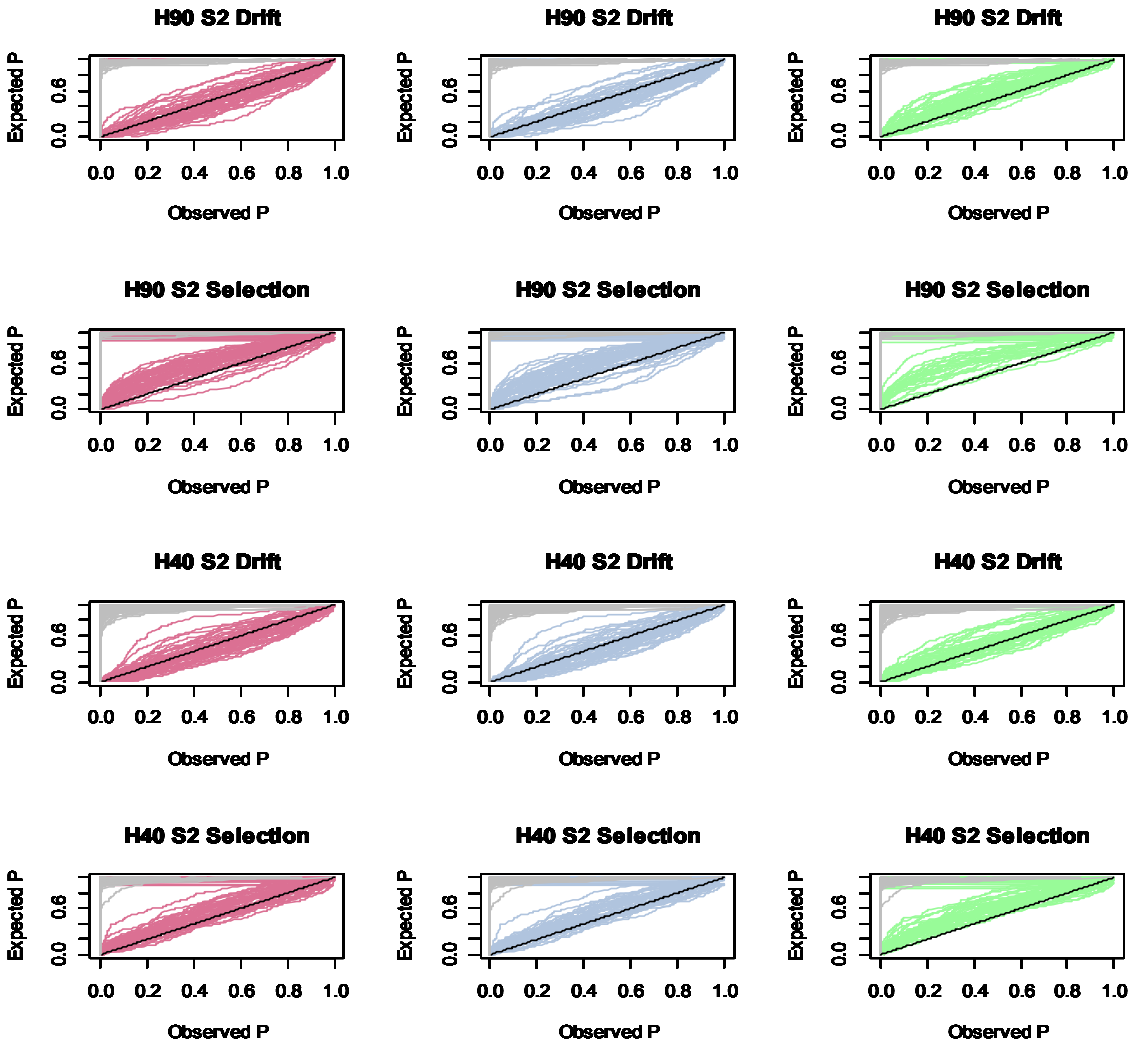
Comparisons of type-I errors

S.I.=10%
(multi-allelic model)



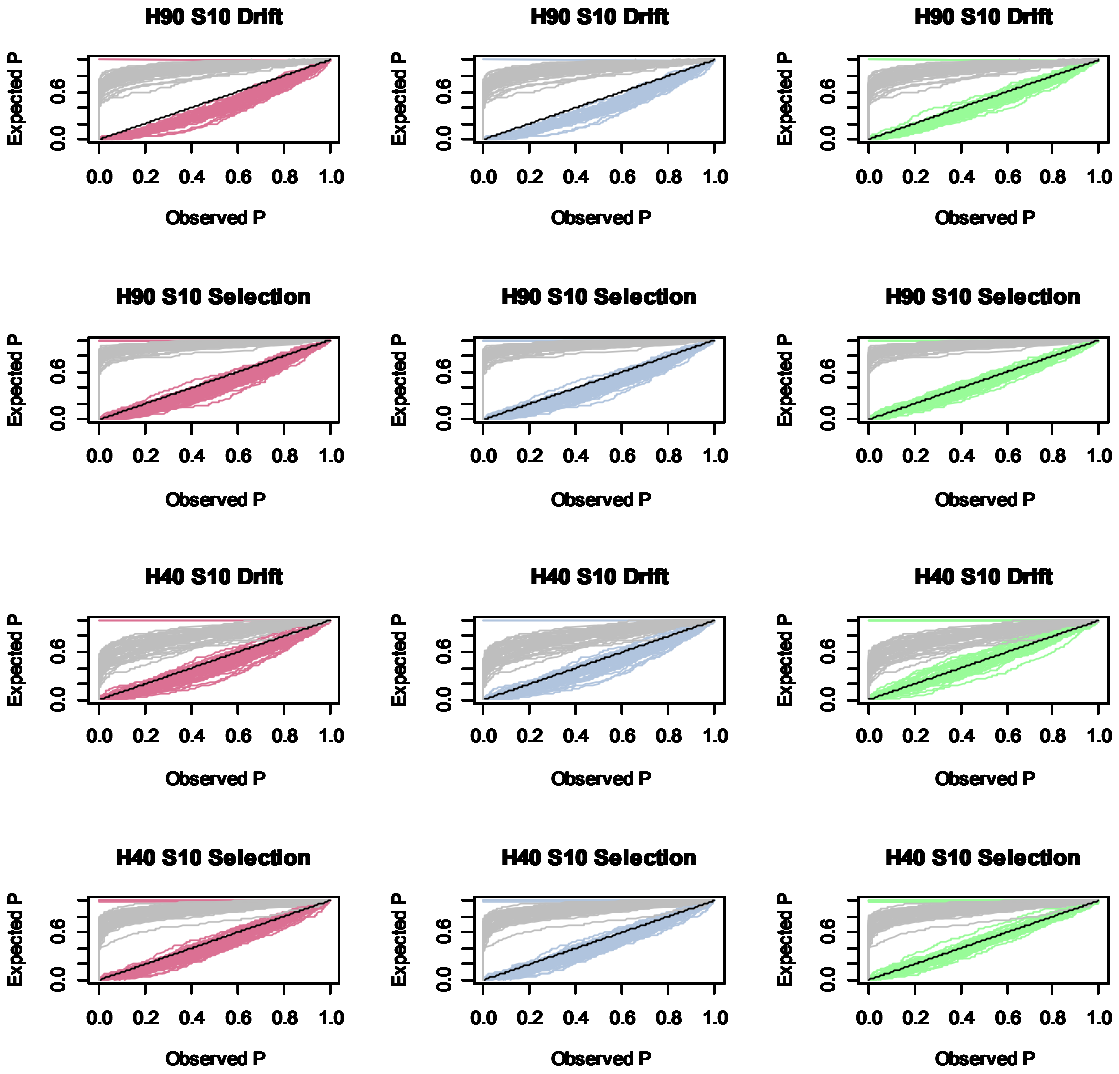
Comparisons of type-I errors

S.I.=2%
(multi-allelic model)



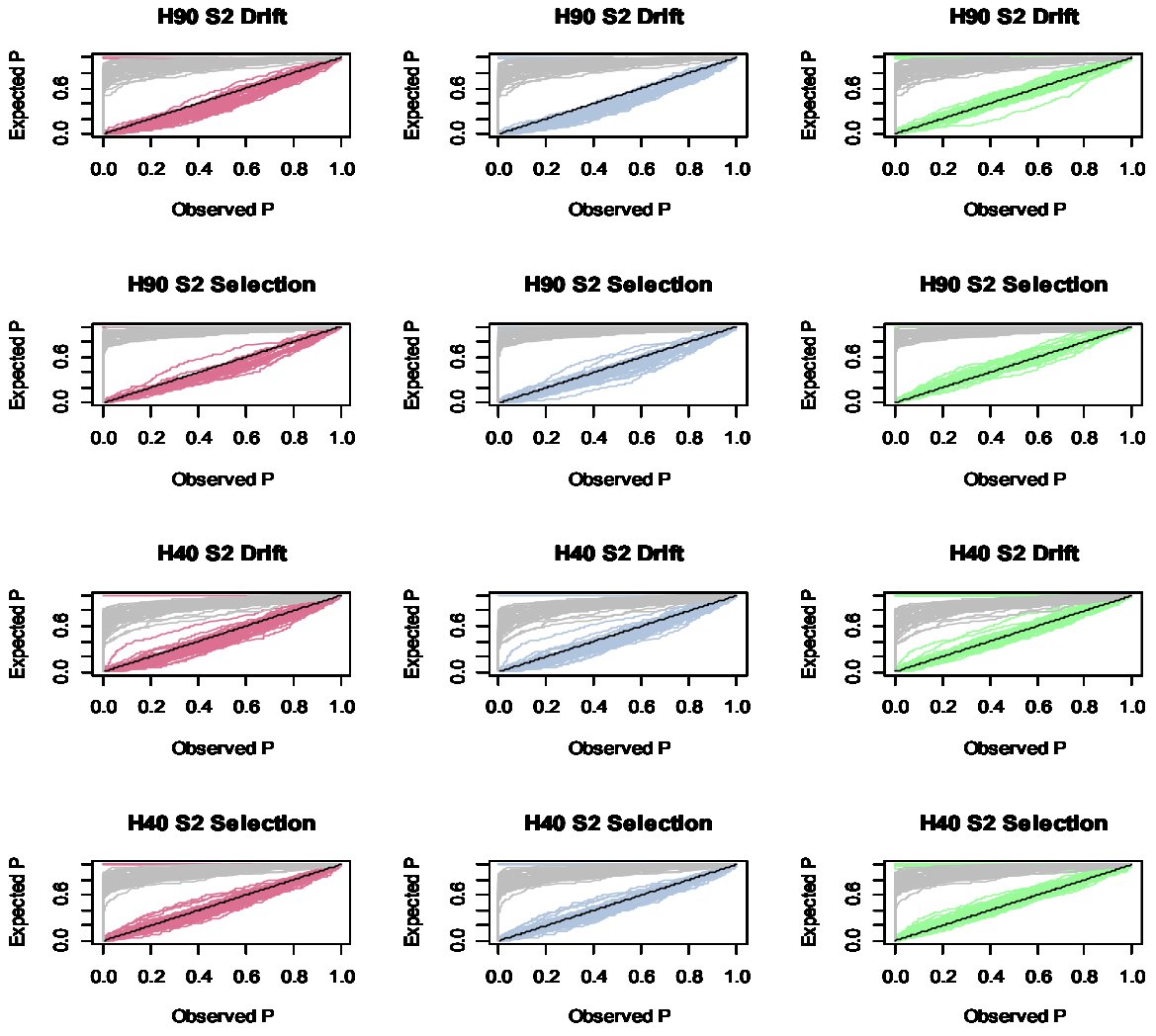
Comparisons of type-I errors

S.I.=10%
(bi-allelic model)



Comparisons of type-I errors

S.I.=2%
(bi-allelic model)



Summary

- Association mapping can be applied to closed-breeding populations under recurrent selection.
- Any estimator of coancestry that more accurately captured the variation in response resulted in better model fit; however since most variation in the response is accounted for, the power was worse
- Poorer estimates of coancestry gave better power; however this came at the expense of higher rate of false positives
- Heritability had a huge impact on power, much higher than selection intensity

Future directions

- For coancestry estimation, alternative weighting schemes can be employed to further correct for the inflation observed with QIBS
- Since the whole genome simulator enabled us to catalogue all the datasets generated; systematic impact of different evolutionary forces at different stages of population improvement can be examined on coancestry and LD
- Selection dynamics, which are under the control of experimenters can be optimized in future studies to better understand selection response via association mapping

Thanks!