Smarter

SMAll RuminanTs breeding for Efficiency and Resilience

WP5: Genomic/genetic modelling and methods of selection for resilience and efficiency traits

Activities coordinated by UEDIN

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n°772787

WP Objectives

1. Develop genetic models to deal with resilience and efficiency under micro- and macro- environmental challenges

2. Infer (genomic) breeding values for R&E from longitudinal data on productive traits

3. Develop and assess methods for genomic prediction particularly suited to small ruminants

4. Develop and implement new methods for selected and under-utilised populations of sheep and goats while conserving genetic variability

Objectives and Tasks

Objectives	Tasks
Develop genetic models for R&E under micro- and macro-environmental challenges	Novel genetic models for efficiency and resilience traits under challenge
Infer (genomic) breeding values for R&E from longitudinal data on productive traits	Data mining of longitudinal performance data to identify and characterise events of environmental challenges
Develop and assess methods for genomic prediction suited to S&G	Enhancing performance of genomic prediction methodology
Develop and implement new methods for selected and under-utilised populations of S&G while conserving genetic variability	Incorporating genomic information to improve management of genetic diversity and to promote expression of heterosis

Deliverables and Milestones

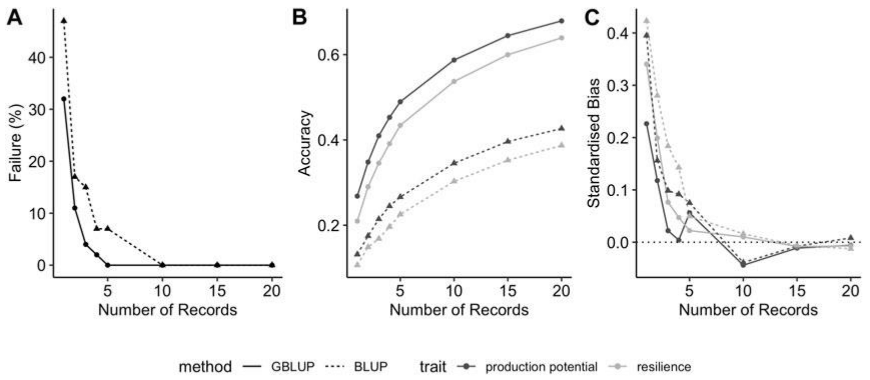
Number	Title	Lead beneficiary	Due date (M)	Due date
MS17	A computer software for genomic prediction taking into account the effect of environmental challenge on performance	UEDIN	18	30/04/2020
MS19	A computer software or method for assessing and avoiding potential bias due to design of cross-validation analysis	INRA	18	30/04/2020
MS20	A computer software to improve calculation of genomic relationship between purebred and crossbred populations	UEDIN	18	30/04/2020
D5.1	Method for identifying environmental challenge events and assessment of their value for selection for resilience	INRA	24	31/10/2020
D5.2	A report on novel methods and breeding practices to improve genomic prediction in sheep and goat populations	UEDIN	36	31/10/2021
MS18	A list of populationwise environmental challenges	INRA	24	31/10/2020
MS21	A computer software to optimise genetic contribution	UEDIN	24	31/10/2020
D5.3	Method for assessing potential bias due to design of crossvalidation analysis	INRA	36	31/10/2021
D5.4	A report for optimum contribution to manage diversity at critical regions and to assist mating design to maximise heterozygosity and expression of heterosis and evaluation of level inbreeding rate across genomic regions and their impact on performance	UEDIN	36	31/10/2021
D5.5	A manuscript on the development and testing of genomic evaluation tools to improve resilience and efficiency accounting for the impact of environmental challenge and using novel phenotypes	UEDIN	48	31/10/2022

Task 5.1: Novel genetic models for efficiency and resilience traits under challenge (M1-M48)

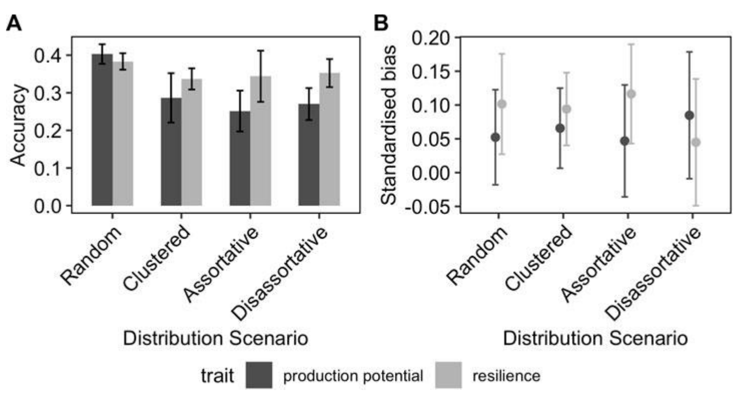
- **Objective of the task**
  - Develop genetic models to deal with resilience and efficiency under micro- and macro- environmental challenges

- **Work done**
  - **Simulation study**
    - **to assess the factors affecting the performance of RR-RN models**
    - **To quantify the potential benefit when selecting for resilience**

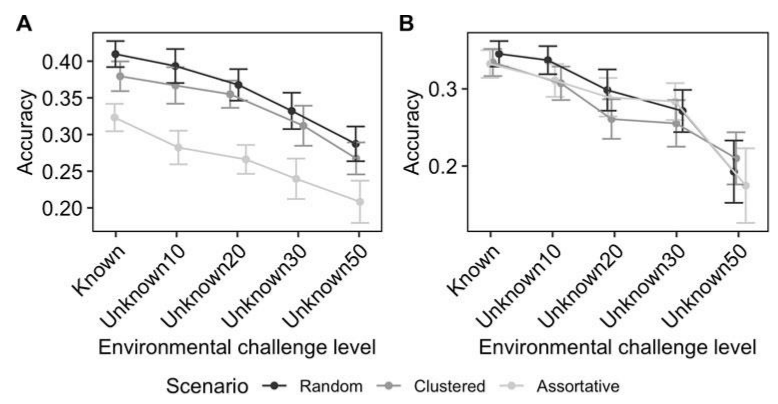
Genomic prediction with random regression/reaction norm models



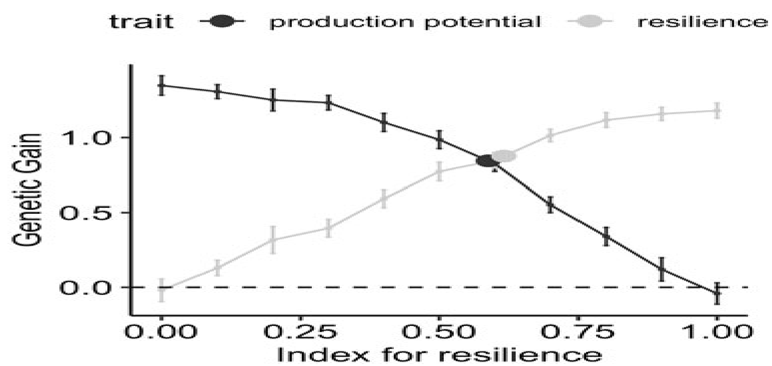
Effect of allocation of phenotype across environment

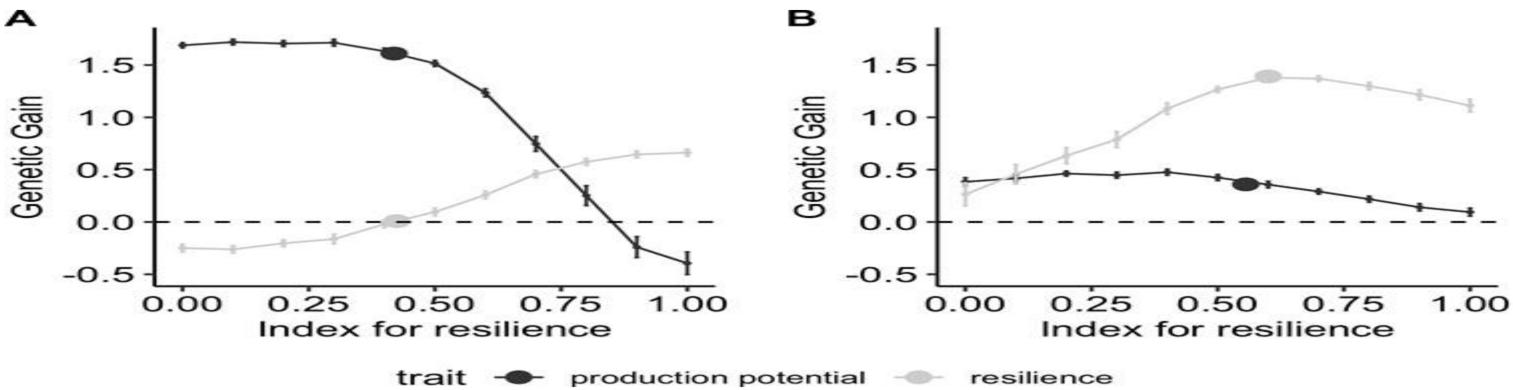


Effect of uncertainty of level of challenge



Genomic selection with random regression/reaction norm models





Testing in Good environment

Testing in bad environment

Task 5.2: Data mining of longitudinal performance data to identify and characterise events of environmental challenges (M1-M36)

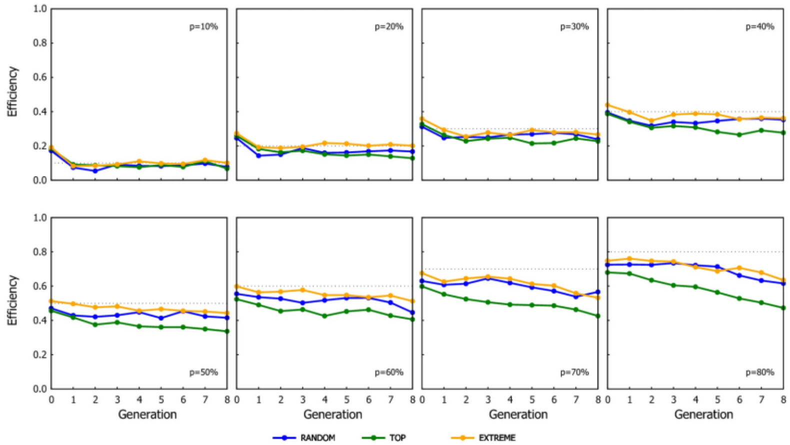
- **Objective of the task**
  - Infer (genomic) breeding values for R&E from longitudinal data on productive traits

- **Work done**
  - **UEDIN NOT INVOLVED IN THIS TASK**

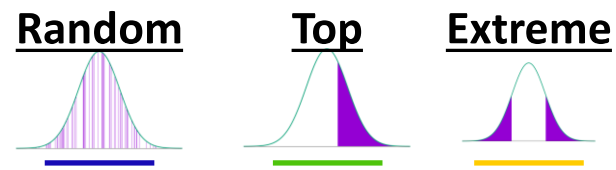
Task 5.3: Enhancing performance of genomic prediction methodology (M1-M36)

- **Objective of the task**
  - Develop and assess methods for genomic prediction particularly suited to small ruminants
- **Work done: Simulation study on:**
  - the proportion and genotyping strategies when selection based on ssGBLUP evaluation
  - The effect of including dominance effect in the model of analysis. (D5.2)
  - The accuracy of across-population prediction when accounting for divergency between population (D5.2)

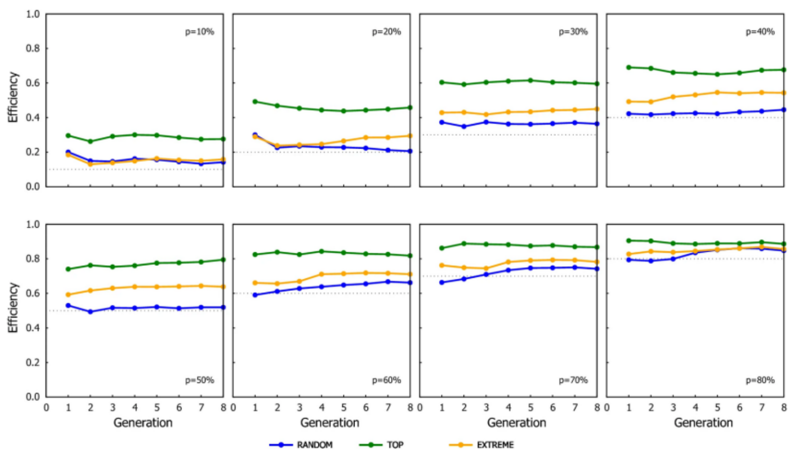
Impact of genotyping strategies of benefit of ssGBLUP



Efficiency of the ssGBLUP in term of their overall reliability (combining both genotyped and ungenotyped animals) using three genotyping strategies, when the proportion of the genotyped candidates was chosen based on phenotypes. The dotted line in each graph indicates the proportion of candidates that were genotyped in the corresponding scenario shown in the graph



Gain	**TOP ≥ EXTREME ≥ RANDOM**
r2	**EXTREME >> RANDOM > TOP**



Efficiency of the ssGBLUP scenarios in terms of their cumulative genetic response using three genotyping strategies, when the proportion of the genotyped candidates was chosen based on estimated breeding values. The dotted line in each graph indicates the proportion of candidates that were genotyped in the corresponding scenario shown in the graph

Optimum strategy cannot be determined by the accuracy of prediction only

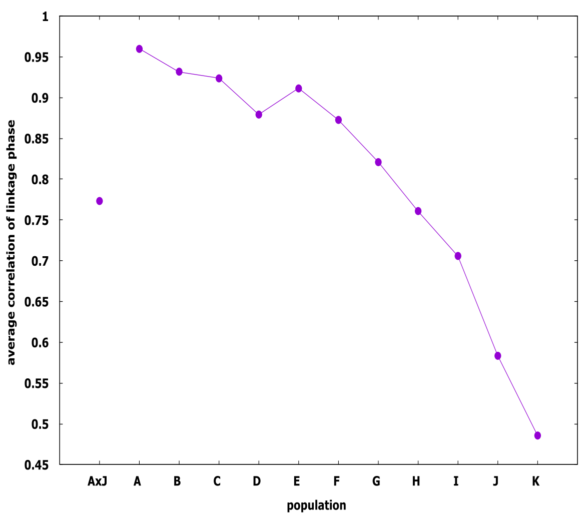
Accounting for dominance in model of analysis: Accuracy of GEBV

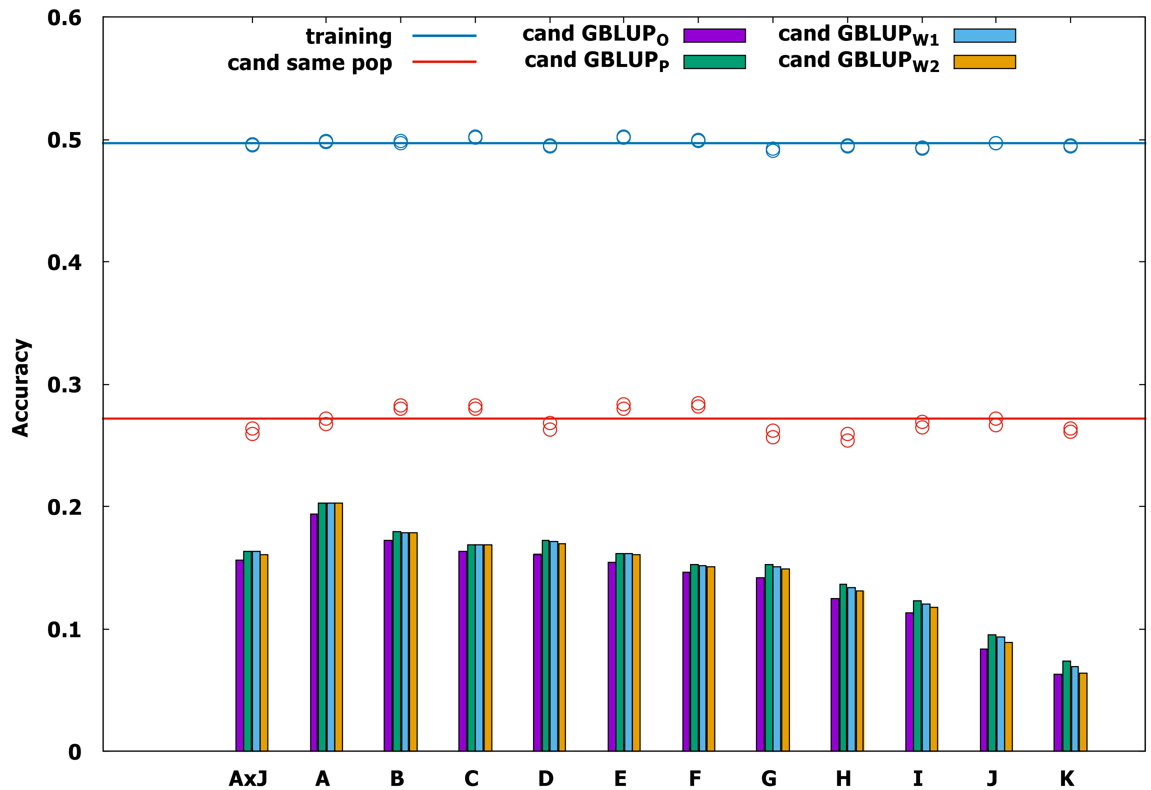
	Individuals with records			Individuals without records		
	Model of analysis			Model of analysis		
True σ_d^2	Additive only	Dominance breeding value parameterisation	Dominance genotypic parameterisation	Additive only	Dominance breeding value parameterisation	Dominance genotypic parameterisation
GBLUP using variance component estimated with REML analysis						
10	0.509	0.509	0.509	0.302	0.302	0.304
20	0.485	0.485	0.483	0.282	0.282	0.282
30	0.467	0.467	0.466	0.273	0.272	0.269
GBLUP using the true variance						
10	0.510	0.510	0.509	0.303	0.302	0.302
20	0.486	0.486	0.484	0.282	0.283	0.282
30	0.486	0.486	0.484	0.282	0.283	0.282

No improvement in accuracy when accounting for dominance

Accuracy of GEBV in reference and divergent populations

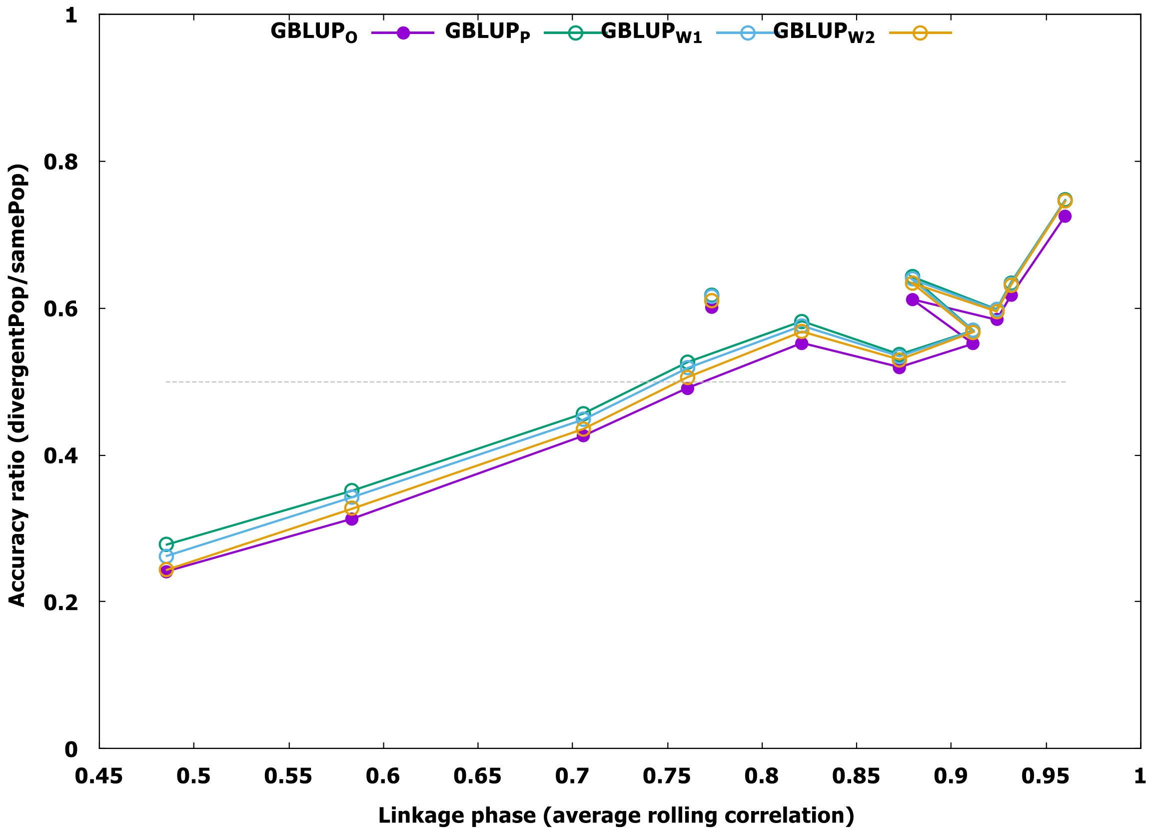
- Method of estimating GRM across population
- Standard GRM (GBLUP_o)
- Allele specific (GBLUP_p)
- SNP weighed by persistence of LD
- SNPweighed by persistence of LD





Very minor benefit when using specific allele frequency or Reweight for SNP for persistence of LD

Effect of persistence of LD



Persistence of LD affect accuracy, but minor effect when modifying GRM to account for persistence

Task 5.4: Incorporating genomic information to improve management of genetic diversity and to promote expression of heterosis (M1-M36)

- **Objective of the task**
  - Develop and implement new methods for selected and under-utilised populations of sheep and goats while conserving genetic variability
- **Work done**
  - Better formulation of the OCS problem.
  - Evaluating the different estimates of genomic inbreeding
  - Effect of using diffrenet GRM to manage genetic diversity

The OCS problems as a Mixed Integer programming (MIP)

Optimise $\mathbf{c}$ and $\tilde{\mathbf{c}}$

Minimise $h(\mathbf{c})$

st:

$\mathbf{s}'\mathbf{c} = 0.5$

$\mathbf{d}'\mathbf{c} = 0.5$

$\mathbf{c} \geq \underline{\boldsymbol{u}}\tilde{\mathbf{C}}$

$\mathbf{c} \leq \overline{\boldsymbol{u}}\tilde{\mathbf{C}}$

$\frac{\mathbf{c}'\mathbf{G}_j\mathbf{c}}{\mathbf{2}} \leq \mathrm{F}_{\boldsymbol{j}}^{*}, j = 1, p$

$\tilde{\mathrm{c}}_i * (\tilde{\mathrm{c}}_i - 1) = 0, i = 1, n$

Include an extra variable to optimise
$\tilde{\mathbf{c}}$ which is integer 0,1

Make it continuous [0,1] and introduce
Constraint $\tilde{\mathrm{c}}_i * (\tilde{\mathrm{c}}_i - 1) = 0$

$$\mathcal{L}(\mathbf{c}, \tilde{\mathbf{c}}, \lambda_s, \lambda_d, \boldsymbol{\lambda}_{\tilde{\mathrm{c}}}, \boldsymbol{\lambda}_{\underline{\boldsymbol{u}}}, \boldsymbol{\lambda}_{\overline{\boldsymbol{u}}}, \lambda_j) = h(\mathbf{c}) - \lambda_s(\mathbf{s}'\mathbf{c} - 0.5) - \lambda_d(\mathbf{d}'\mathbf{c} - 0.5) - \tilde{\mathbf{c}}'(\tilde{\mathbf{C}} - \mathbf{I})\boldsymbol{\lambda}_{\tilde{\mathrm{c}}} - \boldsymbol{\lambda}_{\underline{\boldsymbol{u}}}'(\mathbf{c} - \underline{\boldsymbol{u}}\tilde{\mathbf{C}}) + \boldsymbol{\lambda}_{\overline{\boldsymbol{u}}}'(\mathbf{c} - \overline{\boldsymbol{u}}\tilde{\mathbf{C}}) + \sum_{j=1}^{p} \lambda_j(\mathbf{c}'\mathbf{G}_{\boldsymbol{j}}\mathbf{c}/2 - \mathrm{F}_{\boldsymbol{j}}^{*})$$

Better formulation of OCS problem but solving still a hard problem

A MIP OCS for maximising heterozygosity

Optimise $\tilde{c}_{i,j}$ mating status between male I and female j

Maximise $\sum \tilde{c}_{i,j} * \mathrm{He}_{i,j}$

st: $\sum_i \tilde{c}_{i,j} = n_i \;\; i = 1, m$

$\sum_j \tilde{c}_{i,j} = n_j \;\; j = 1, f$

$\tilde{c}_{ij} * (\tilde{c}_{ij} - 1) = 0, i = 1, m; j = 1, f$

The value of genomic relationship matrices for estimating inbreeding

Genomic Inbreeding (from diagonal of GRM)

$$F_{L\&H} = \frac{SF_{NEJ} - \sum_{k=1}^{S}\left[1 - 2p_{k(0)}\left(1 - p_{k(0)}\right)\right]}{S - \sum_{k=1}^{S}\left[1 - 2p_{k(0)}\left(1 - p_{k(0)}\right)\right]} \qquad]-\infty : 1\,]$$

$$F_{VR1} = \frac{\sum_{k=1}^{S}\left(x_k - 2p_{k(0)}\right)^2}{2\sum_{k=1}^{S} p_{k(0)}\left(1 - p_{k(0)}\right)} - 1 \qquad]-1 : \infty\,]$$

$$F_{YAN} = \frac{1}{S}\sum_{k=1}^{S}\frac{x_k^2 - \left(1 + 2p_{k(0)}\right)x_{k_i} + 2p_{k(0)}^2}{2p_{k(0)}\left(1 - p_{k(0)}\right)} \qquad]-1 : \infty\,]$$

The value of genomic relationship matrices for estimating inbreeding

Expected genomic Inbreeding from different GRM
(based in change in allele frequencies)

F > 1

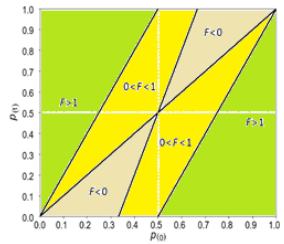
> 100% initial diversity lost

0 <= F <= 1

Diversity lost

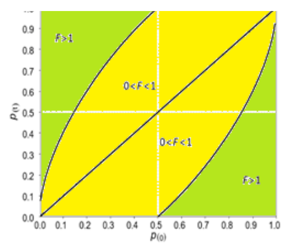
F < 0

Diversity gained



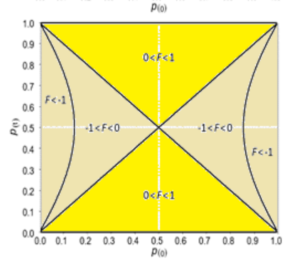
VR1/VR2

- Losing more variance than the initial value
- Loss of variance when it's been gained
- Gain of variance when it's been lost



Yang

- Losing more variance than the initial value
- Expect not gain in variance (in average)



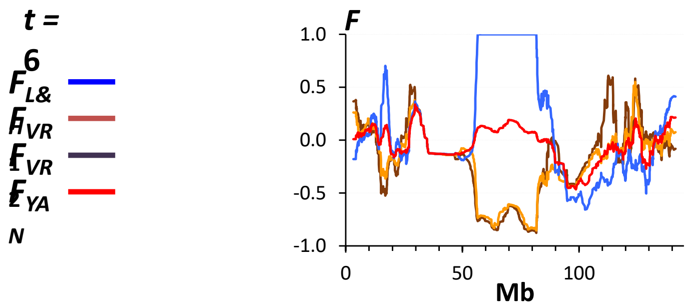
L&H

- Loss/gain in variance in right direction
- Allow for gain in variance

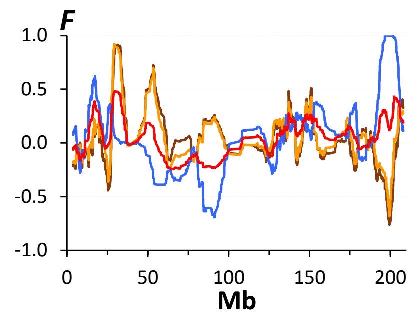
T5.4
D5.4

Patterns of $F_{L\&H}$, F_{VR1}, F_{VR2} and F_{YAN} – Cohort 6

SSC14 - Whole region fixed



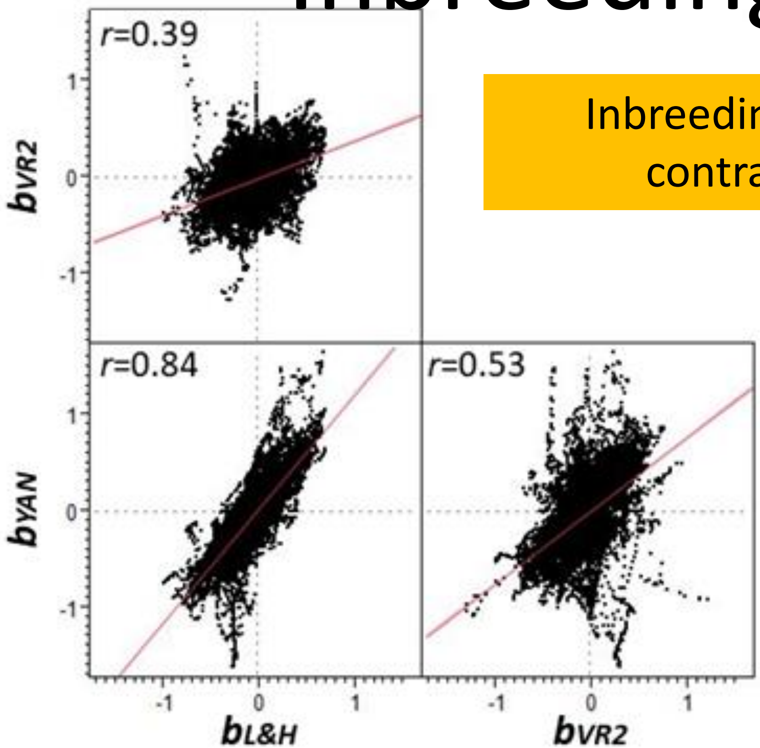
SSC13 - Region where homoz decreased



- $F_{L\&H} = 1$ ➡ All variability lost
- F_{VR1}, $F_{VR2} < 0$ ➡ Gain in variability
- $F_{YAN} > 0$ ➡ Some variability lost

- $F_{L\&H} < 0$ ➡ Gain in variability
- F_{VR1}, $F_{VR2} > 0$ ➡ Some variability lost
- $F_{YAN} \sim 0$

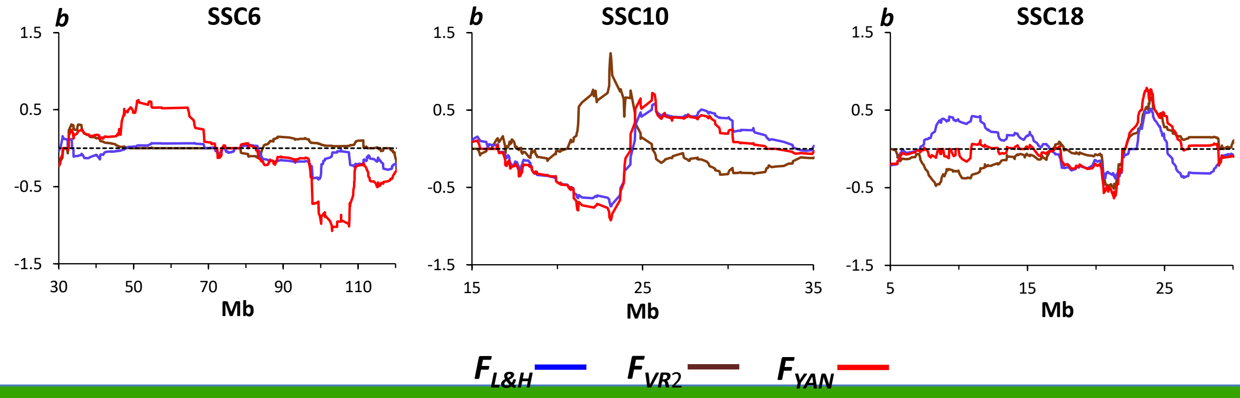
Inbreeding depression



Inbreeding depression estimates can have contrary direction amount methods

% of *b* values with opposite signs

- **40% considering $F_{L\&H}$ and F_{VR2}**
- **27% considering F_{VR2} and F_{YAN}**
- **15% considering $F_{L\&H}$ and F_{YAN}**



Effect of using these matrices in OCS

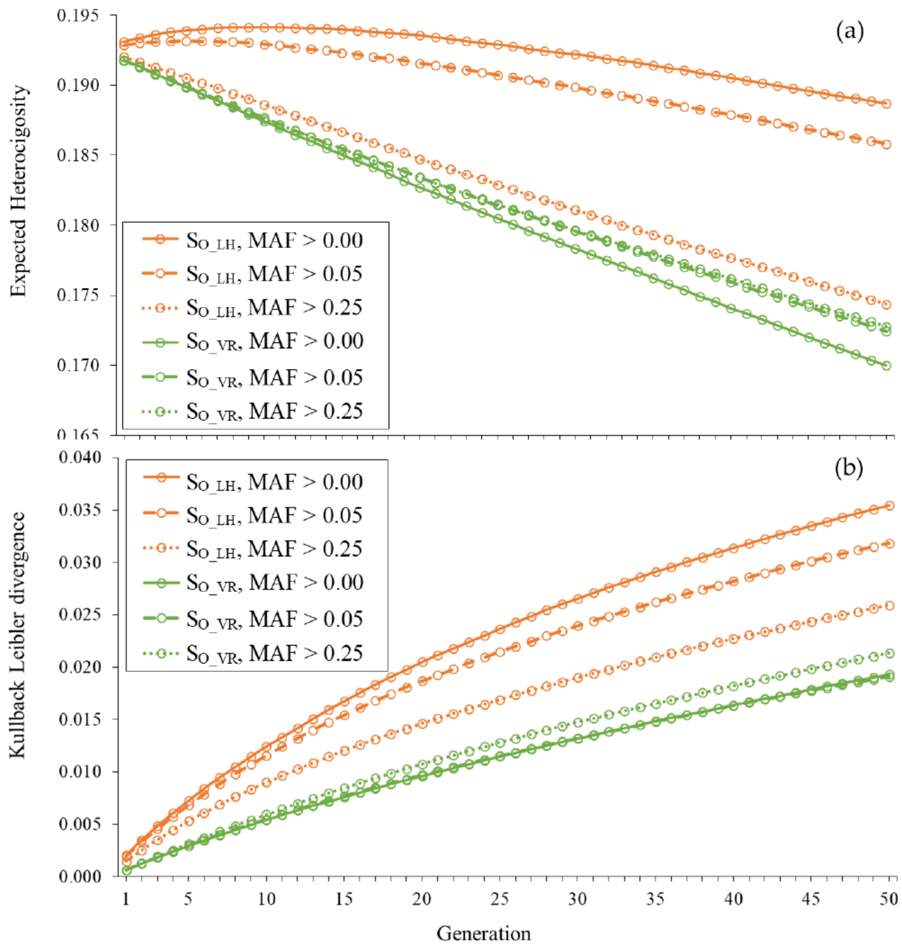
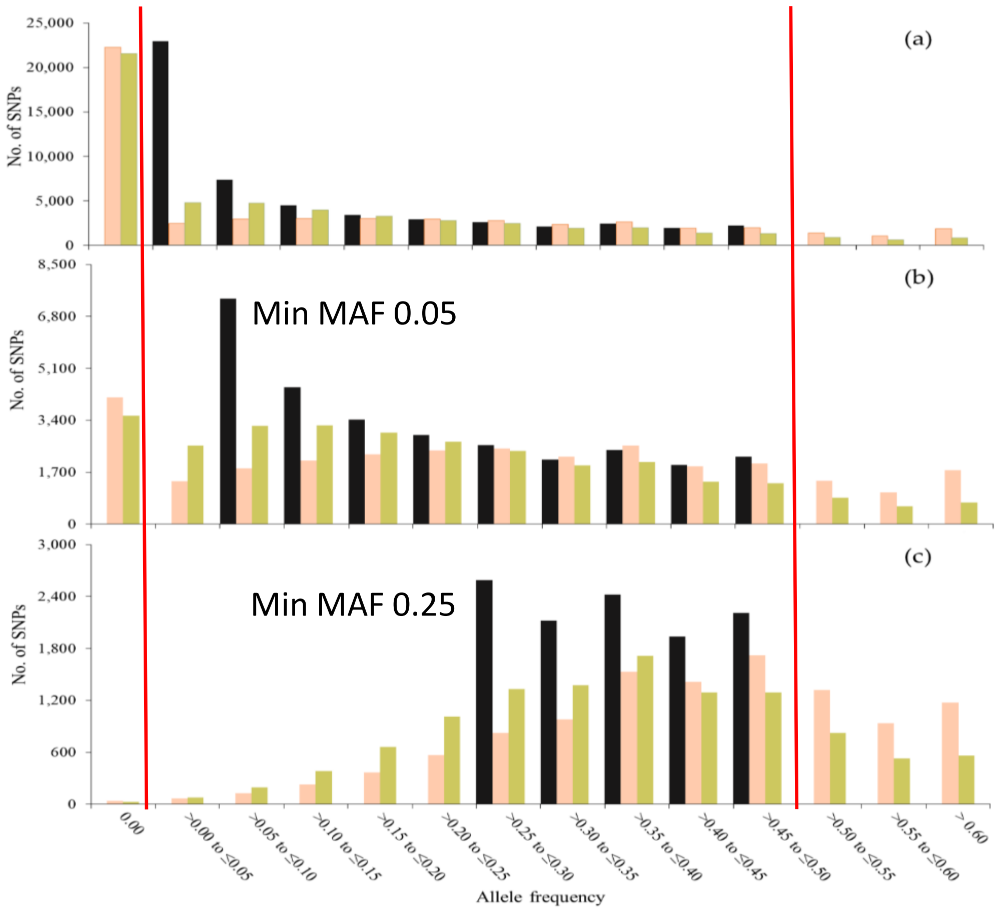


Figure 3. Expected heterozygosity (**a**) and Kullback–Leibler divergence (**b**) for unobserved loci across generations when contributions are optimized using Li and Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices computed with SNPs with MAF > 0.00, MAF > 0.05 and MAF > 0.25 in a population of 100 individuals.

L&H

- retain more heterozygosity
- Greater deviation from initial freq

Change in frequency in observed SNP



G0

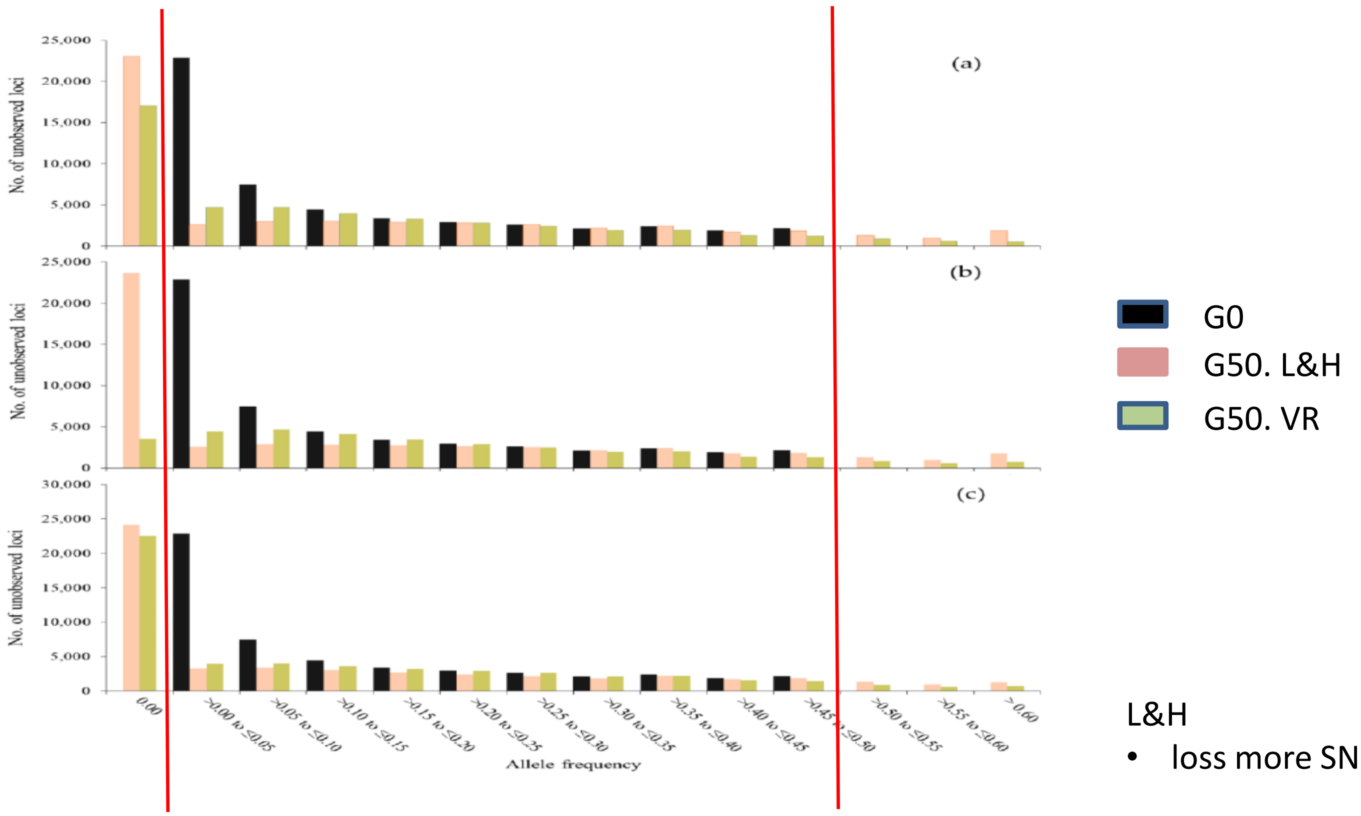
G50. L&H

G50. VR

L&H

- moves freq towards 0.5
- loss more SNP to minor allele

Change in frequency in unobserved SNP



L&H

- loss more SNP to minor allele

Choice of GRM in OCS depend of objective:
Increasing variability: L&H
Mantaining the original gene frequency: VR

Ne

Table 4. Effective population size (N_e) across generations (t) when contributions are equalized (S_E) and when they are optimized using Li and Horvitz (S_{O_LH}) and VanRaden (S_{O_VR}) coancestry matrices in populations of different sizes (N).

	N = 100			N = 20		
t	S_E	S_{O_LH}	S_{O_VR}	S_E	S_{O_LH}	S_{O_VR}
1	188.21	−111.90	195.55	36.92	42.27	40.40
5	199.07	−855.78	197.46	36.78	41.24	34.31
10	191.56	−5777.32	193.05	38.54	40.81	41.77
15	203.50	1855.71	194.54	36.65	45.41	43.18
20	202.62	1033.03	201.52	40.61	47.25	40.02
25	190.44	636.00	209.85	40.20	47.08	42.02
30	193.58	670.07	209.79	36.45	53.03	38.57
35	193.30	524.97	206.03	33.41	50.28	44.62
40	204.95	601.67	212.53	36.94	47.91	49.68
45	207.44	703.31	205.00	37.52	48.50	40.09
50	206.86	481.08	213.02	41.99	46.20	38.53

Ne several folds greater with L&H than VR With n=100

Ne=1/2 ΔHe

Smarter