UPOV
International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/2/4 Add.**

Second Session
Virtual meeting, April 8 to 11, 2024

Original: English
Date: April 8, 2024

ADDENDUM TO:
REFERENCE COLLECTION MANAGEMENT USING MOLECULAR MARKERS: A NEW APPROACH BASED ON GENOMIC PREDICTION

Document prepared by experts from the United Kingdom

Disclaimer: this document does not represent UPOV policies or guidance

The annex to this document contains a copy of a presentation “Reference collection management using molecular markers: a new approach based on genomic prediction”, made by an expert from the United Kingdom, at the second session of the Technical Working Party on Testing Methods and Techniques (TWM).

[Annex follows]

invite

Genomic prediction for reference collection management (TWM/2/4)

BioSS

cagasc

nak tuinbouw

UNIVERSITY OF HOHENHEIM

This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under Grant Agreement No 817970.

1

Outline

Define a more tailored approach for using markers to reduce trial sizes

- Based on modelling the association between markers and characteristics
- Ties more closely with approach used for distinctness assessment

Assess in 4 crops based on real DUS data

- Maize ✗
- Perennial ryegrass ✓
- Soybean ✗
- Wheat ✗

Introduction

Currently: 2 main approaches for use of markers in DUS:

a) Characteristic-Specific Molecular Markers

Identify markers that are very closely linked to QTLs determining characteristics:

- Mainly disease resistance

b) Combining Phenotypic and Molecular Distances in the Management of Variety Collections

- Compare overall genetic vs phenotypic distances
  - Identify threshold for genetic distance that is likely "guarantee" phenotypic distinctness ←
- Identification of similar varieties to the candidate

Proposal

Our proposal fits under application model b)

Uses Genomic Prediction to maximise the link between markers and phenotype

- thus gives greater potential for trial size reduction

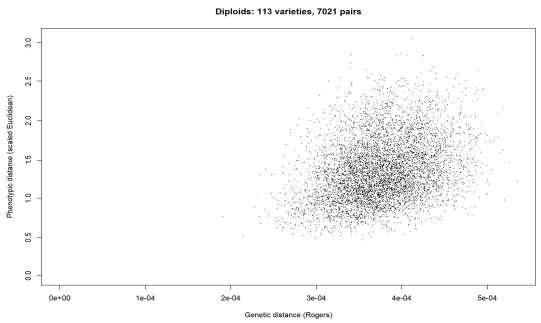
Applied characteristic-by-characteristic, mirroring DUS assessments

- Aim to predict D decisions
- Try to ensure that we do not eliminate close varieties
- Gain advantage from rule that distinctness only required in one characteristic

Why? (perennial ryegrass example)

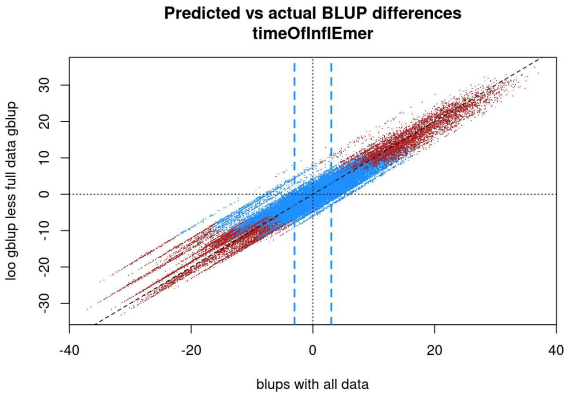
Current method

Low correlation between phenotypic and genetic distances



Genomic prediction

One character: cross-validated



Genomic Prediction

Genomic Prediction is commonly used in plant-breeding

- To select breeding material
- Allows a better understanding of key traits, such as yield
  - Field data is always limited and variable
  - Augmenting field data with genetic data can give a better "prediction" of the trait
  - Genetics is used as a tool to better understand the trait

Many different methods of genomic prediction

- Some work better for traits driven by a few important genes
  - Bayesian Lasso is an example
- Others work better for traits driven by many small genes
  - GBLUP is an example

Genomic Prediction

Today we focus on GBLUP (genomic best linear unbiased prediction)

- Straightforward to estimate uncertainty in predictions
- More complex models do not work as well with our ryegrass data

GBLUP simply uses the estimated pairwise relationships between varieties

- But we do need all the varieties to make the prediction – not just the particular pair
- Maybe of advantage re confidentiality?

We also consider an extension: GBLUP+QTL

- This adds in specific markers found by GWAS

Framework for using GP

Fit a GP model to existing data

Predict the difference between a candidate and the reference variety

- Reference varieties have genetic data + historic phenotypic data
- Candidate variety has genetic data only

Assess whether the difference is significant

- Can use same probability values as COYD (eg 1%)

Perennial Ryegrass data

DUS data from Naktuinbouw
- Up to 13 years 2007-2019
- 21 characteristics

Markers: ~187k SNPs with allelic frequencies, after quality screen

Varieties with both DUS data and markers
- Diploids: 119
- Tetraploid: 149
- Combine for analysis (268), taking into account ploidy, then separate for decisions

Compare to long-term COYD at 1%

Note:
- fraction of the National List due to lack of breeder permission
  - Reduces data set size, but also genetic variability
    - Both will affect success here
- only registered varieties, not failed applications
  - Cannot assess chance that we find a failed candidate incorrectly distinct or uniform

Assessing how well GP works

Using historical PRG data set

1. Take each variety in turn and treat as candidate
   - try to predict its over-year mean (like COYD)

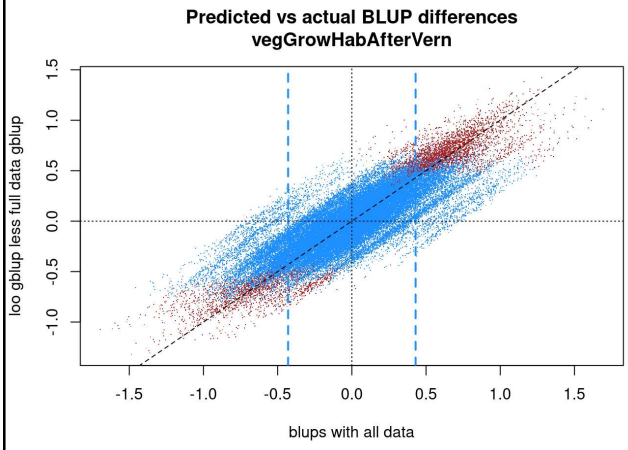
2. Predict differences between "candidate" and reference variety
   - *Reference* based on phenotype data (and genetic data)
   - *Candidate* based on genetic data (no phenotypic data) – LOO

3. Assess whether this predicted difference is significant at 1% (using GBLUP model)

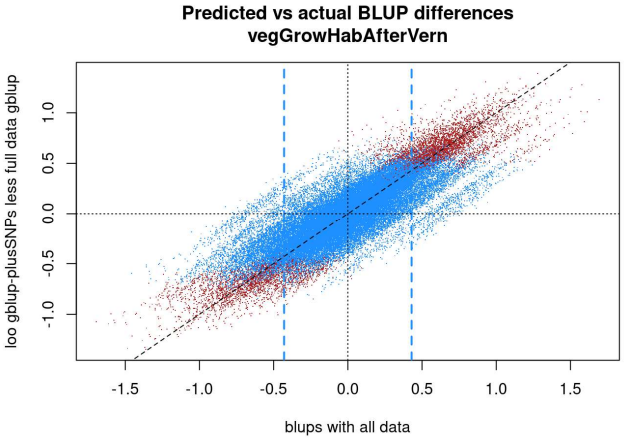
4. Compare with actual differences in phenotypic means (based on long-term BLUP) and long-term COYD distinctness decisions at 1%

Note: we do not have failed candidates in the data set

GBLUP



GBLUP+QTL

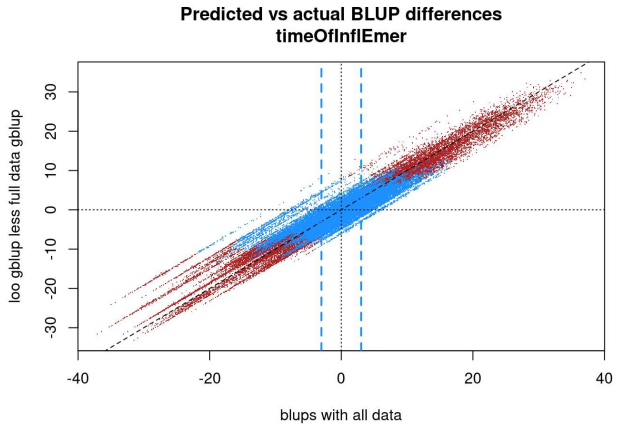


blue lines denote long-term COYD threshold based on BLUP analysis
red points show sig dif (0.01) for GBLUP prediction

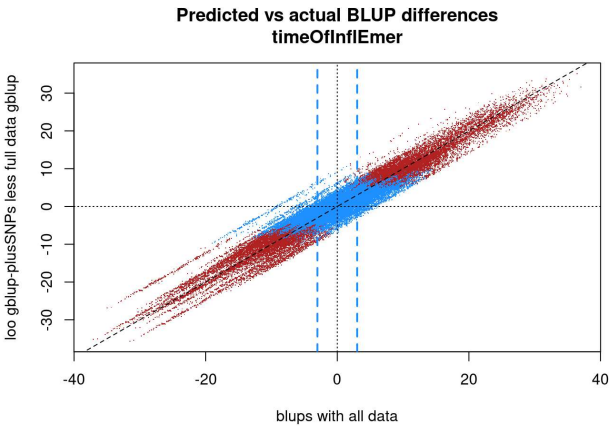
Page 11 *h2020-invite.eu*

11

GBLUP



GBLUP+QTL



blue lines denote long-term COYD threshold
red points show sig dif (0.01) for GBLUP prediction

Page 12 *h2020-invite.eu*

12

Proportion of candidates distinguished based on SNPs

	diploid		tetraploid	
Characteristic	gblup	gblup_QTL	gblup	gblup_QTL
vegGrowHab	2.4%		0.4%	
intGreenCol	0.3%		0.1%	
plantWidth	0.2%	1.0%	0.7%	2.9%
vegGrowHabAfterVern	7.2%	9.4%	6.5%	9.0%
plantHeight	11.0%	13.2%	6.5%	8.8%
intGreenColAfterVern	1.2%		4.1%	
timeOfInflEmer	24.4%	38.6%	27.0%	39.0%
natHeightAtInflEmer	5.3%		4.8%	
growHabAtInflEmer	2.9%		12.5%	
flagLeafLength	0.7%		1.0%	
flagLeafWidth	6.5%		9.6%	
flagLeafLengthWidthRatio	3.7%	5.1%	1.1%	2.4%
lengthOfLongestStem	7.7%		9.2%	
lengthOfUpperInternode	1.1%		0.6%	
inflLength	0.7%		4.8%	
NumberOfSpikelets	2.6%		1.4%	
inflDensity	2.1%		5.6%	
lenghtOfOuterGlumeOnBasalSpikelet	2.0%		0.1%	
lengthOfBasalSpikelet	2.2%		4.1%	
spikeletProtuberance	1.6%		3.8%	
glumeSpan	4.4%		2.0%	

Over characteristics – success rates

		Proportion Distinct		GP false positives
		Blup	GP	
Diploid	GBLUP	95%	41%	4%
	GBUP+QTL		52%	6%
Tetraploid	GBLUP	89%	41%	5%
	GBUP+QTL		51%	9%

How does this translate into trial size reduction?

Assume plant whole collection unless we know distinct from candidate
Separate by ploidy

Issue: all candidates have to be distinct from a reference variety to eliminate that reference variety from the trial
Simulation based on this data set – mean reduction

Expected trial size reduction

Number of candidates	Diploids	Tetraploids
1	52%	52%
2	31%	31%
5	12%	13%
10	6%	7%

Summary

New method is clearly more effective than the current UPOV model for using markers

For ryegrass:

- Only one characteristic is well predicted by markers based on this analysis
- If there was a full reference collection, it is likely that distinctness rates would improve
  - Greater genetic variability
  - Just more varieties
- Uncertain at the moment whether this approach would be more effective than cyclic planting when >2 candidates
- Number of markers can be much lower than in practice for this method:
  - Estimation of genomic relationship matrix needs far fewer markers (needs reference!)
  - Will need extra markers selected for QTLs
    - a dense map to start with helps, then can select

Are there alternative ways to design trials using this method?

Future work

Perennial ryegrass

- To add in dominance and epistasis effects into GBLUP

Three more crops

- Wheat, maize, soybean
- GWAS currently being undertaken by NIAB and Hohenheim
- Higher heritability and better linkage disequilibrium in these crops
  - ➔ will work better than PRG?

Discuss with end users in INVITE

- *Can we plan trials better using this approach?*

A note: GBLUP only requires a matrix of pairwise "relationships" between varieties

- May provide a way for breeder to keep control of fingerprints
  - See BMT/14/5
  - But may not work for other GP methods

invite

Stay informed:

Website: www.h2020-invite.eu
Email: a.roberts@bioss.ac.uk

[End of Annex and of document]