UPOV

International Union for the Protection of New Varieties of Plants

E

Technical Working Party on Testing Methods and Techniques **TWM/3/6**

Third Session
Beijing, China, April 28 to May 1, 2025

Original: English
Date: March 26, 2025

INVITE GENOMIC PREDICTION FOR REFERENCE COLLECTION MANAGEMENT: WHEAT

Document prepared by experts from the United Kingdom

The annex to this document contains a copy of a presentation “INVITE Genomic prediction for reference collection management: Wheat”, to be made by experts from the United Kingdom, at the third session of the TWM.

[Annex follows]

Genomic prediction for variety collection management

Wheat

Adrian Roberts

BioSS NIAB

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

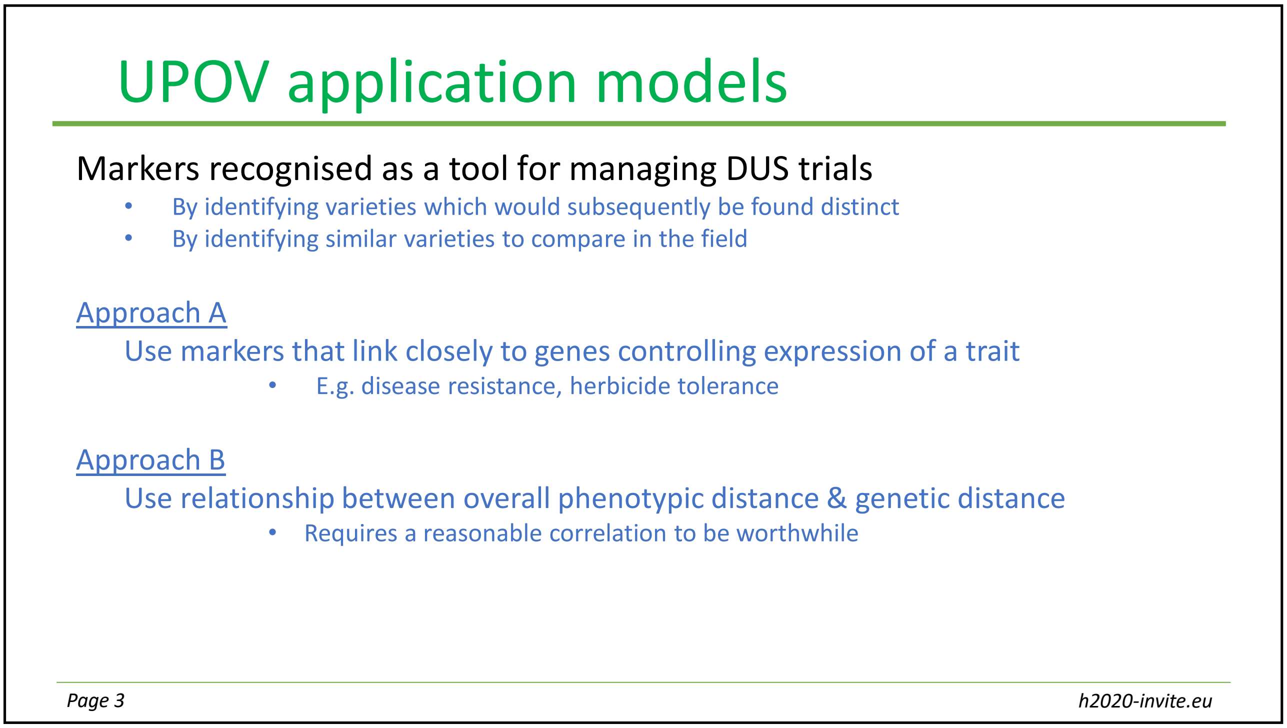
Outline

INVITE project defined a more tailored approach for using markers to manage DUS trial

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

Demonstrate with wheat historical data

In TWM/2/4 demonstrated with perennial ryegrass



UPOV application models

Markers recognised as a tool for managing DUS trials

- By identifying varieties which would subsequently be found distinct
- By identifying similar varieties to compare in the field

Approach A

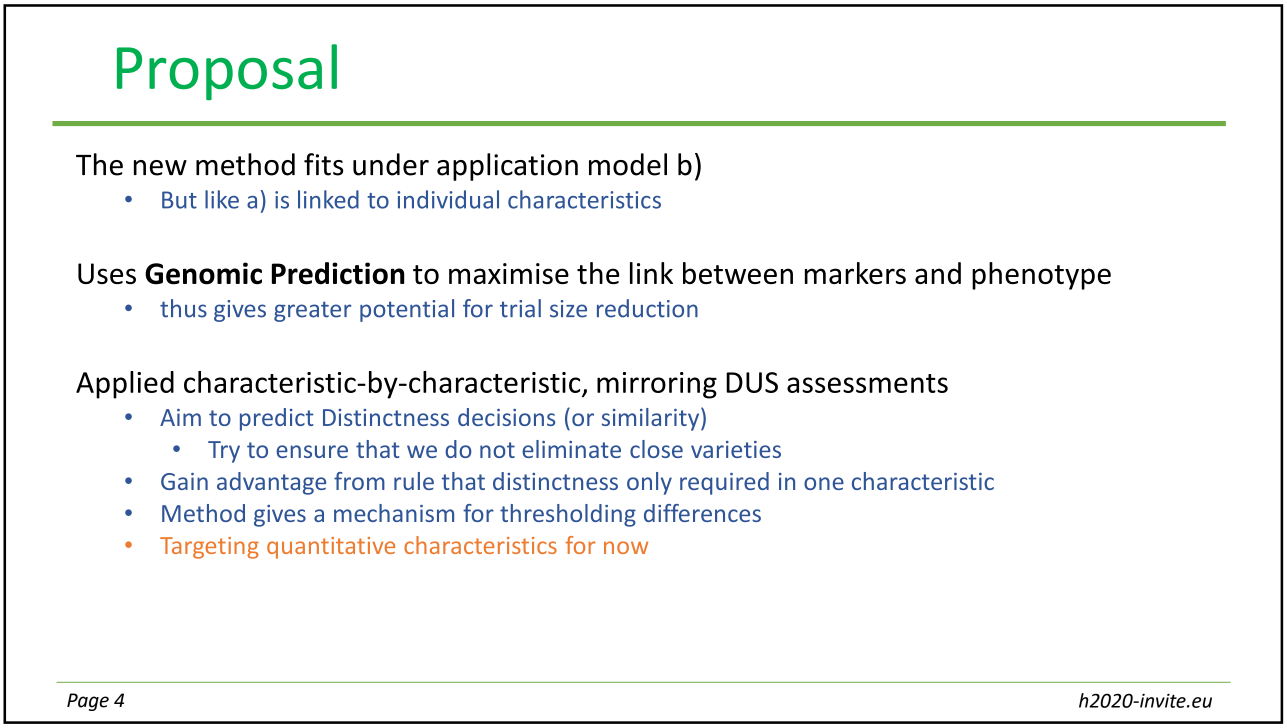
Use markers that link closely to genes controlling expression of a trait

  - E.g. disease resistance, herbicide tolerance

Approach B

Use relationship between overall phenotypic distance & genetic distance

  - Requires a reasonable correlation to be worthwhile



Proposal

The new method fits under application model b)

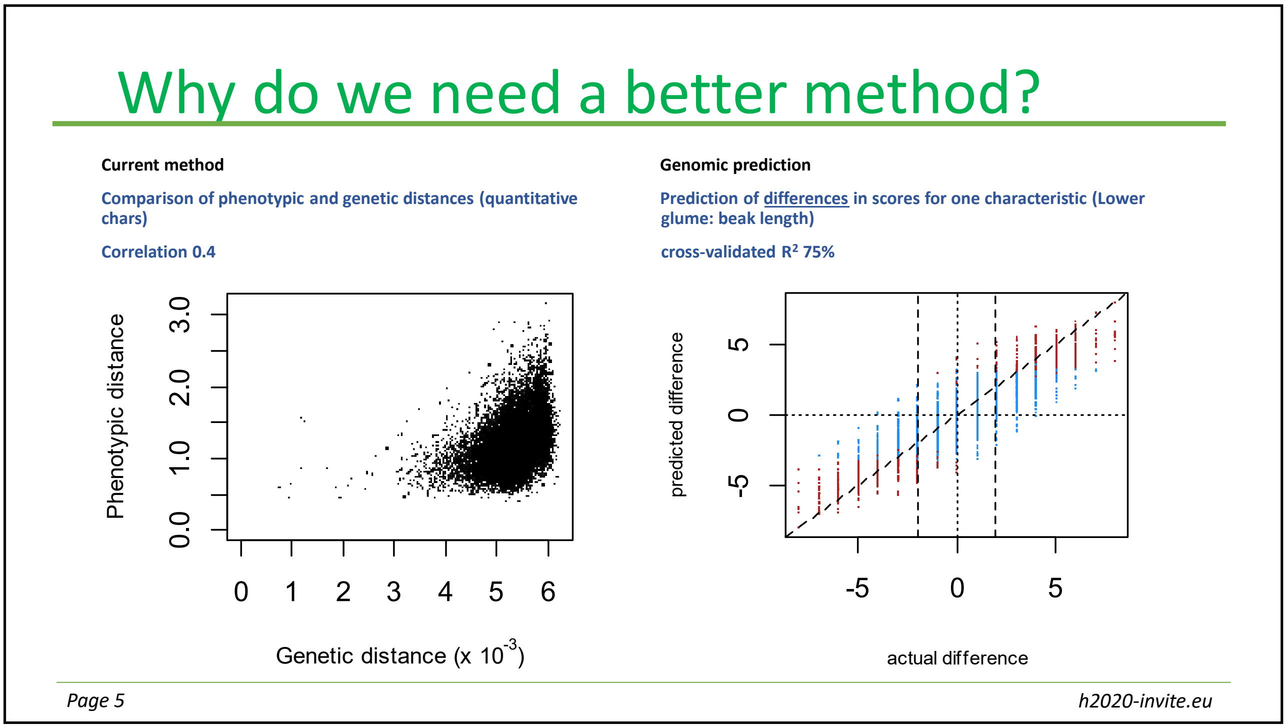
- But like a) is linked to individual characteristics

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 Distinctness decisions (or similarity)
  - Try to ensure that we do not eliminate close varieties
- Gain advantage from rule that distinctness only required in one characteristic
- Method gives a mechanism for thresholding differences
- Targeting quantitative characteristics for now



Why do we need a better method?

Current method

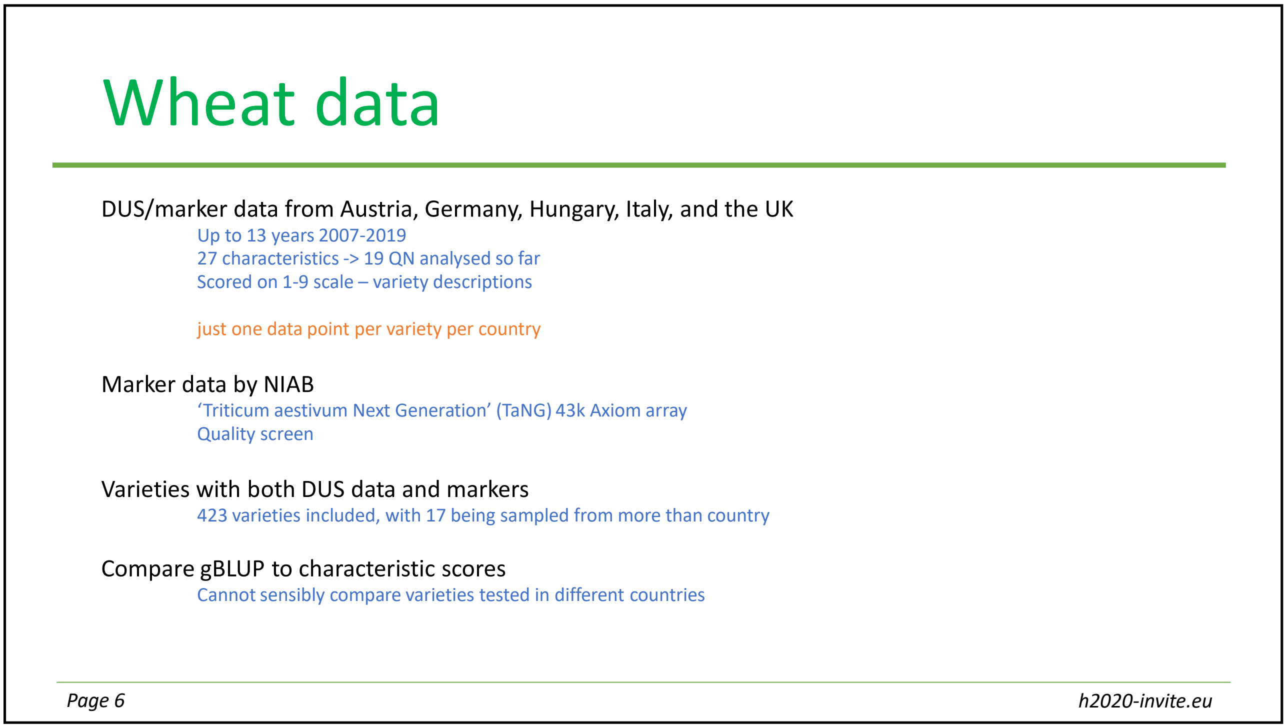
Comparison of phenotypic and genetic distances (quantitative chars)

Correlation 0.4

Genomic prediction

Prediction of <u>differences</u> in scores for one characteristic (Lower glume: beak length)

cross-validated R^2 75%



Wheat data

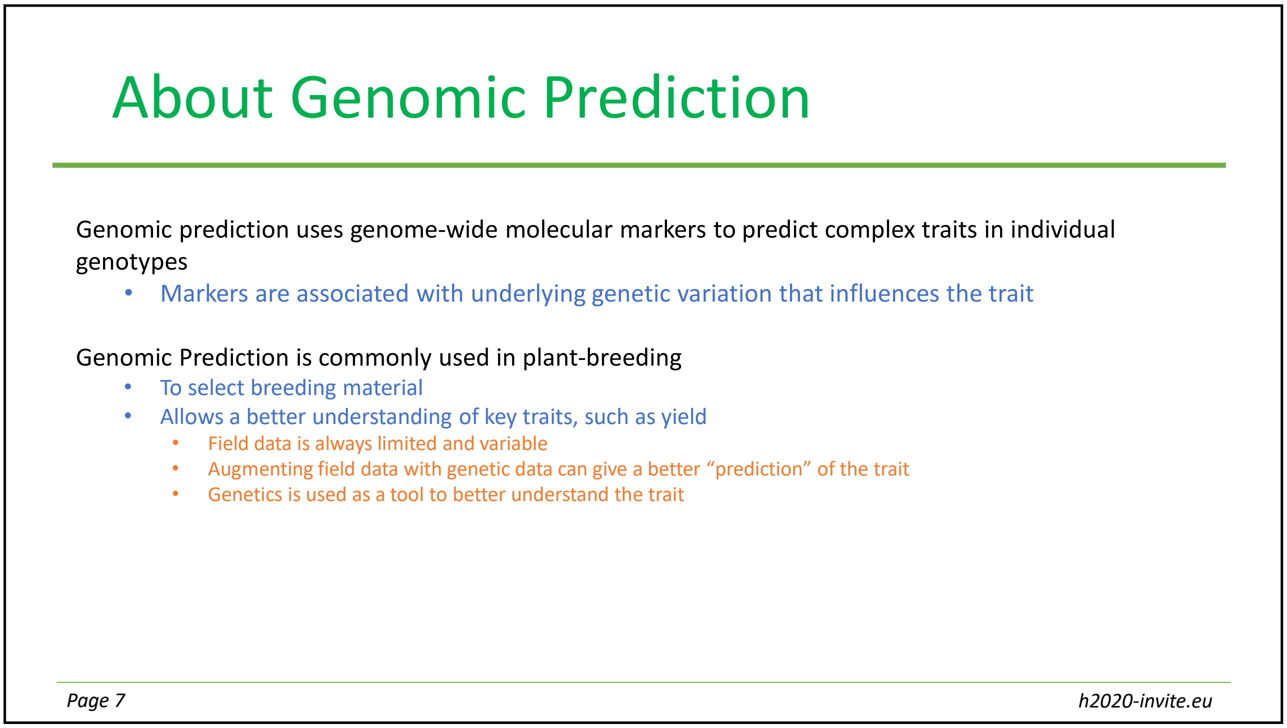
DUS/marker data from Austria, Germany, Hungary, Italy, and the UK
- Up to 13 years 2007-2019
- 27 characteristics -> 19 QN analysed so far
- Scored on 1-9 scale – variety descriptions

just one data point per variety per country

Marker data by NIAB
- ‘Triticum aestivum Next Generation’ (TaNG) 43k Axiom array
- Quality screen

Varieties with both DUS data and markers
- 423 varieties included, with 17 being sampled from more than country

Compare gBLUP to characteristic scores
- Cannot sensibly compare varieties tested in different countries



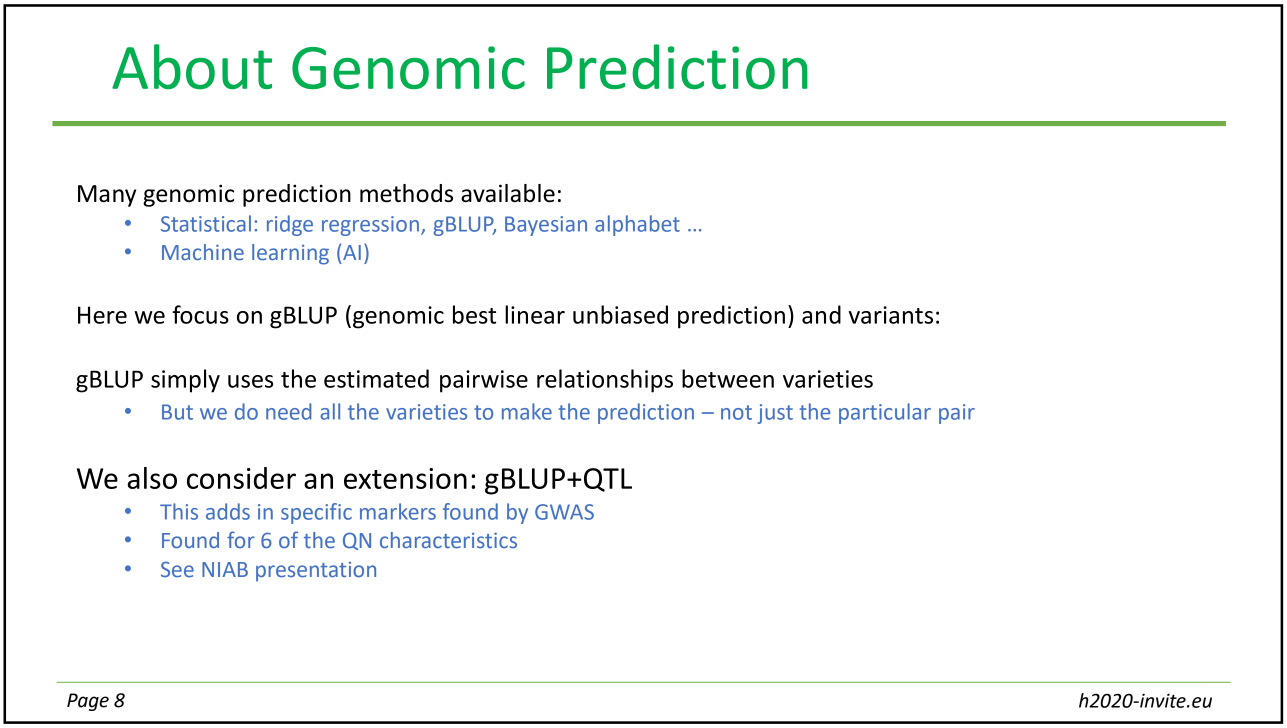
About Genomic Prediction

Genomic prediction uses genome-wide molecular markers to predict complex traits in individual genotypes

- Markers are associated with underlying genetic variation that influences the trait

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



About Genomic Prediction

Many genomic prediction methods available:

- Statistical: ridge regression, gBLUP, Bayesian alphabet …
- Machine learning (AI)

Here we focus on gBLUP (genomic best linear unbiased prediction) and variants:

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

We also consider an extension: gBLUP+QTL

- This adds in specific markers found by GWAS
- Found for 6 of the QN characteristics
- See NIAB presentation

Framework for using genomic prediction for trial management

Fit a genomic prediction model to existing data

Predict the difference between a candidate and another variety

- Varieties in collection 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%)
- For wheat, distinctness may be based on a fixed difference in UPOV notes (commonly 2 notes)

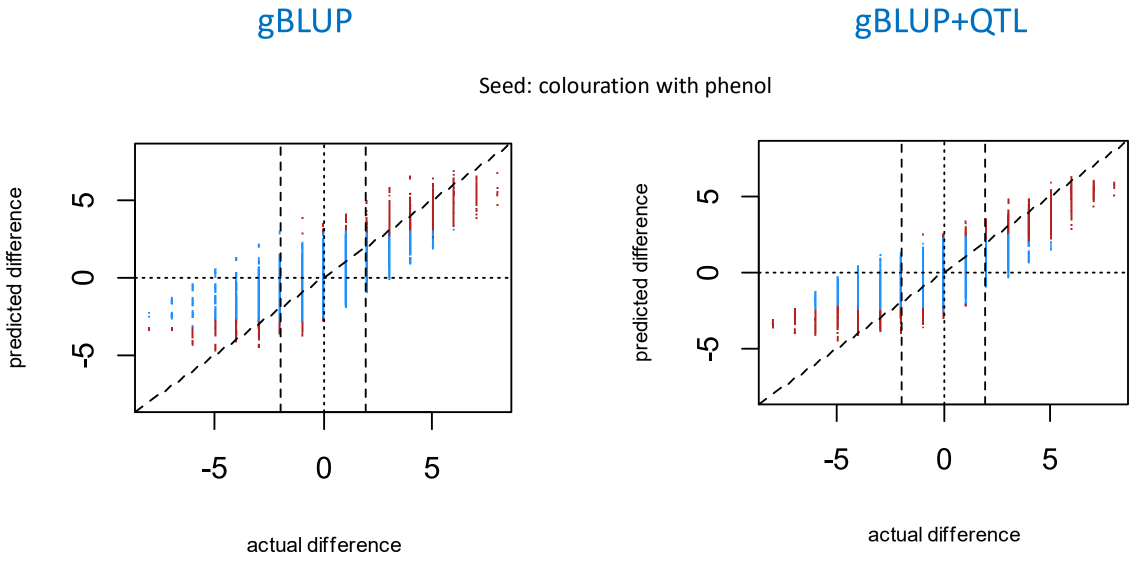
In some crops, including wheat, emphasis may be on finding similar varieties

Assessing how well GP works

Using historical 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 (Leave-One-Out)
3. Assess whether this predicted difference is significant at 1% (using gBLUP model)
4. Compare with actual differences in phenotypic means

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



black lines denote two notes difference
red points show sig dif (0.01) for gBLUP prediction

Wheat: Proportion of varieties distinguished by characteristic

Characteristic	gBLUP	gBLUP+QTL
Seed: colouration with phenol	5.8%	16.3%
Coleoptile: anthocyanin colouration	1.7%	5.9%
Growth habit	1.4%	
Frequency of plants with recurved flag leaves	7.1%	
Ear emergence	2.6%	
Flag leaf: glaucosity of sheath	1.0%	
Flag leaf: glaucosity of blade	3.8%	
Ear: glaucosity	1.6%	
Culm: glaucosity of neck	1.1%	
Plant: length	4.2%	
Ear density	2.3%	2.9%
Ear length	3.7%	
Awn or scur length	11.9%	23.3%
Area of hairiness on convex surface	1.1%	
Lower glume: shoulder width	1.3%	
Lower glume: shoulder shape	0.1%	
Lower glume: beak length	13.8%	19.8%
Lower glume: beak shape	0.9%	
Area of hairiness on internal surface	2.5%	4.5%

Proportion of varieties distinguished over all the QN characteristics

	Proportion distinct	
	GBLUP	GBUP+QTL
Wheat	38%	55%

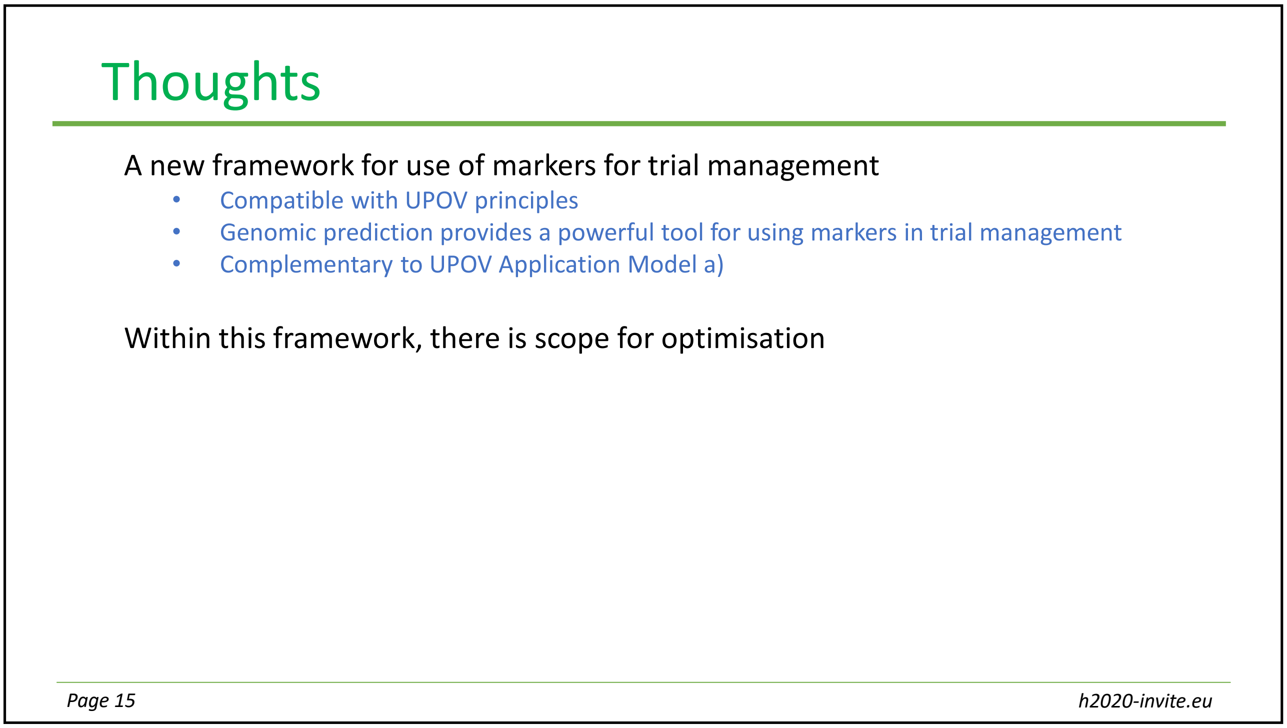
Summary

New method is clearly more effective than the current UPOV application model b)

For Wheat:

- Three QN characteristics well predicted by markers
- QL characteristics will add to overall discrimination – especially if model a) is possible

- Prediction of similarity is important for laying out first year trial
- Markers could also be used to improve information for laying out second year trial

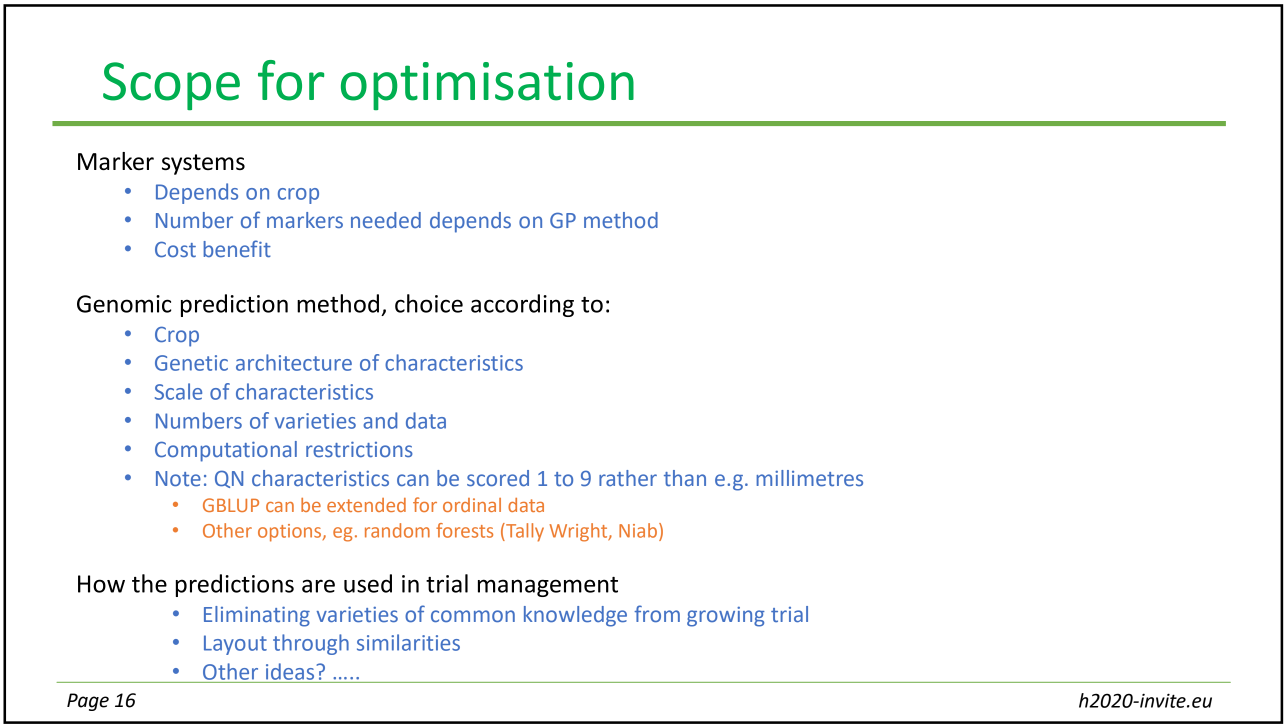


Thoughts

A new framework for use of markers for trial management

- Compatible with UPOV principles
- Genomic prediction provides a powerful tool for using markers in trial management
- Complementary to UPOV Application Model a)

Within this framework, there is scope for optimisation



Scope for optimisation

Marker systems

- Depends on crop
- Number of markers needed depends on GP method
- Cost benefit

Genomic prediction method, choice according to:

- Crop
- Genetic architecture of characteristics
- Scale of characteristics
- Numbers of varieties and data
- Computational restrictions
- Note: QN characteristics can be scored 1 to 9 rather than e.g. millimetres
  - GBLUP can be extended for ordinal data
  - Other options, eg. random forests (Tally Wright, Niab)

How the predictions are used in trial management

- Eliminating varieties of common knowledge from growing trial
- Layout through similarities
- Other ideas? …..

invite

Stay informed:

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

[End of Annex and of document]