E

BMT/13/6 Add.
ORIGINAL: English
DATE: December 8, 2011

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES, AND DNA-PROFILING IN PARTICULAR

Thirteenth Session
Brasilia, November 22 to 24, 2011

ADDENDUM

A POTENTIAL UPOV OPTION 2 APPROACH FOR BARLEY USING HIGH DENSITY SNP GENOTYPING

Document prepared by experts from the United Kingdom

Plant Science into practice

NIAB
National Institute of Agricultural Botany

Presenter Name Carol Norris Date November 2011

NIAB

A Potential UPOV Option 2 Approach for Barley Using High Density SNP Genotyping

A project co-funded by the Community Plant Variety Office (CPVO)

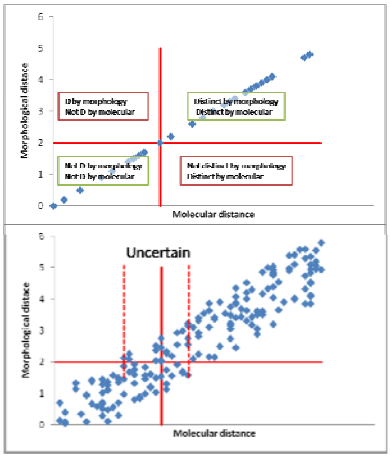
- UPOV Model 2 ""Calibration of threshold levels for molecular characteristics against the minimum distance in traditional characteristics".
- Aim is to ensure the same decisions are made under molecular or morphological testing system.
- Costs of genotyping have decreased, costs of running field trials constant or increasing and better molecular data are available

Presenter Name Carol Norris Date November 2011

NIAB

UPOV Model 2

- Ideal: strong positive correlation, no 'noise'
- Identical decisions made by either method
- Reality: uncertainty if the correlation is not perfect
- How much uncertainty is acceptable?
- Can we use Model 2 taking into account this uncertainty?



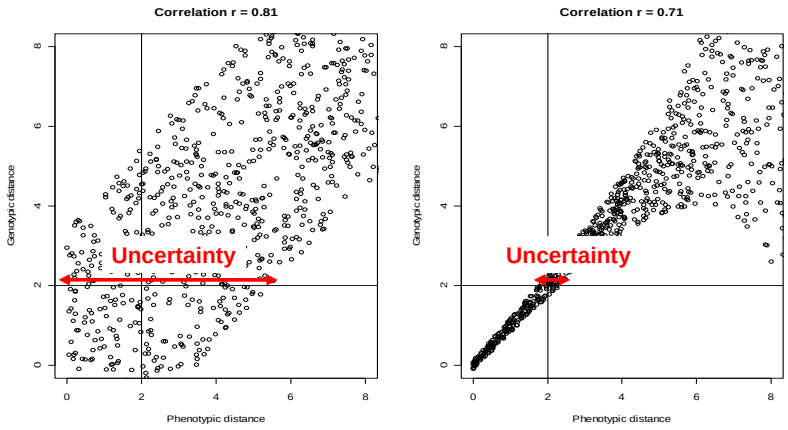
Presenter Name Carol Norris Date November 2011

NIAB

Hypotheses

- Genotypic and phenotypic distance measures for a set of varieties will have a strong positive correlation to each other.
- Varieties shown as 'similar' using phenotypic distances will also be shown as 'similar' using genotypic distances.

Correlation isn't everything



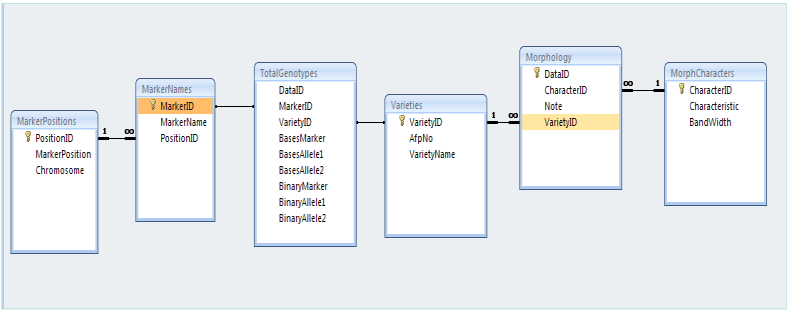
Presenter Name Carol Norris Date November 2011

Project resources

- Genotype data from AGOUEB project - Association mapping used to detect associations between SNPs and DUS characteristics
  - 3072 SNP marker loci
  - 500 UK barley varieties
- Phenotypic data from UK DUS trials
  - 579 winter and spring barley varieties
  - 33 characteristics (some UK)
  - 28 CPVO characteristics
- 431 varieties with both phenotypic and genotypic data
  - Some varieties rejected – too much missing data

Presenter Name Carol Norris Date November 2011

Database structures used to store and manage data



Presenter Name Carol Norris Date November 2011

Dealing with missing phenotypic data

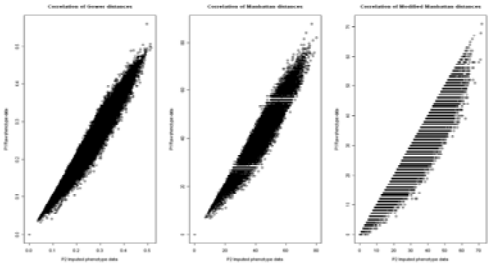
- For each characteristic, missing data were replaced by values drawn at random from the existing data (imputation)
- Missing data were replaced by imputed values to generate complete datasets
- Multiple sets of phenotype data were generated in this way and distance matrices calculated for each of them.
- Distance matrices were pooled and compared to ensure that results were defensible

Presenter Name Carol Norris Date November 2011

NIAB

Validation of phenotypic datasets

- Two datasets used to calculate phenotypic distances
- Raw phenotypic data (P1)
- Data where missing values were replaced by imputation (P2)
- The dataset with missing data (P2) was validated by correlation with the raw phenotype data (P1)



Presenter Name Carol Norris Date November 2011

NIAB

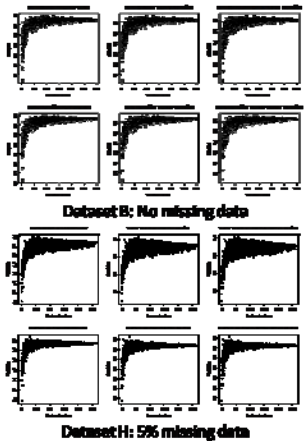
Statistical toolkit

- Phenotypic distances
  - Manhattan distance
  - Modified Manhattan distance
  - Gower distance
- Genetic distance
  - Manhattan distance
  - Euclidean distance

NIAB

Do we have enough markers?

- Initially correlations increase with the number of markers used
- As the number of markers increases further, the correlation values plateau
- Once the correlation has reached a plateau, the scatter of correlations around a central value reduces with increasing marker numbers
- Genotypic data divided into subsets
- 300-400 markers needed from Dataset B (no missing data)
- 800-1000 markers needed from Dataset H (5% missing data)



Presenter Name Carol Norris Date November 2011

NIAB

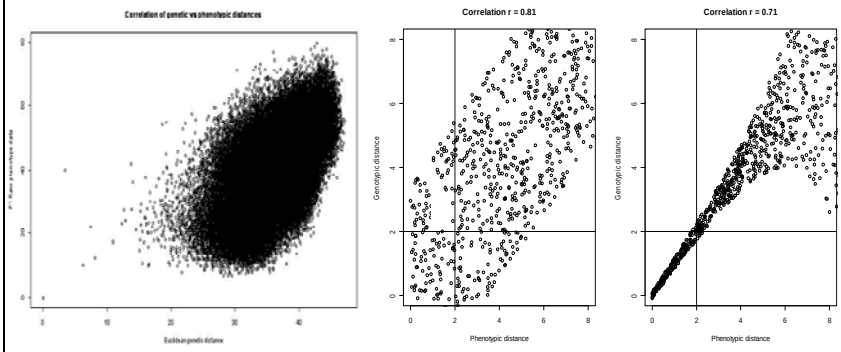
Results

- Correlations between phenotypic and genotypic distances were all positive
- Gower phenotypic distances gave better correlations than Manhattan and Modified Manhattan
- Most correlations were between 0.62 to 0.66 when using Gower's Distance and 0.61 to 0.63 when using Manhattan Distance
- Modified Manhattan Distance gave correlations of 0.58 to 0.60
- Correlations using P2 data were greater than those using P1 data

Presenter Name Carol Norris Date November 2011

NIAB

Typical scatter of genetic vs phenotypic distances



Presenter Name Carol Norris Date November 2011

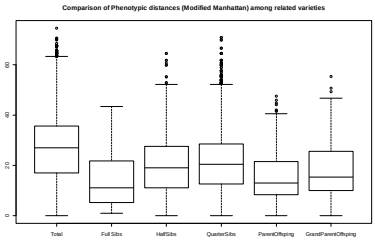
NIAB

Looking at pedigree relationships

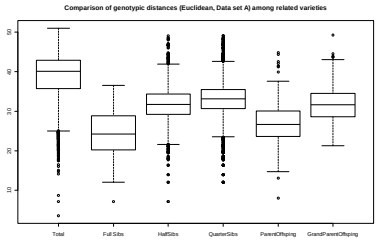
- Pedigree data extracted from Technical Questionnaires and tabulated
- All possible full, half and quarter siblings identified
- Parent-offspring pairs, grandparent-off-spring pairs identified
- Phenotypic and marker-based estimates of relationships were examined

Kinship among the varieties

Morphological distances



Genetic distances



Both methods show the same trends and ranking for relatedness of varieties

Next steps

- Genomic prediction of morphology
- Optimisation of marker selection using spaced markers may give better correlations
- Testing of decision making – will the same decisions be made with markers as those made with current systems?
- Report to CPVO in December
- Too early to make conclusions

NIAB

Acknowledgements

- CPVO
- Huw Jones
- James Cockram
- David Smith
- Ian Mackay

[End of document]