Including molecular marker information in the analysis of multi-environment trial data helps differentiate superior genotypes from promising parents: a wheat example

M.G. Borgognone, D. Butler, F. Ogbonnaya, M.F. Dreccer

The dataset

MET data

- 37 international wheat trials

- 4 years: 2009 – 2012

- 6 countries: Australia, Ethiopia, Lebanon, Mexico, Morocco, Syria

- 17 locations

Designs: RCBD with 2 – 4 replicates & 47 – 200 entries

Several traits measured on each plot of each trial

Genotypes and genetic information

Entries or genotypes: 243 in total

Genotype concurrence generally robust: 46 – 200 in common

Pedigrees for most entries: incomplete or inaccurate

DArT markers: available for 197 of the genotypes } 2487 polymorphic markers, call rate > 80%, minor allele frequency > 5%

Objective of the analysis

To assess which entries

- performed better at which environments and
- which ones would be the best parents for further crossing and for which environments

Statistical methods to analyse the data

$$\boldsymbol{y} = \boldsymbol{X\tau} + \boldsymbol{Z_g u_g} + \boldsymbol{Z_p u_p} + \boldsymbol{e}$$

Smith et al (2001)

$\boldsymbol{y}$ = vector of combined yields for individual trials

$\boldsymbol{\tau}$ = vector of fixed effects

$\boldsymbol{u_g}$ = vector of random genetic effects

$\boldsymbol{u_p}$ = vector of random non-genetic effects

$\boldsymbol{e}$ = vector of combined residuals for individual trials

GxE effects accounted for through a multiplicative factor analytic (FA) model

$\boldsymbol{u_g}, \boldsymbol{u_p}, \boldsymbol{e}$ pairwise independent

normally distributed with zero means

$$var(\boldsymbol{u_p}) = \boldsymbol{G_p} = \bigoplus_{l=1}^{b} \boldsymbol{G_{p_l}}$$

$$var(\boldsymbol{e}) = \boldsymbol{R} = \bigoplus_{i=1}^{t} \boldsymbol{R}_i$$

Genetic effects

When we have pedigree information the total genetic effects can be partitioned into additive and residual non-additive genetic effects

$$\boldsymbol{u_g = u_a + u_e}$$

$\boldsymbol{u_a}$ additive genetic effects (or breeding values) are useful when selecting parents for crossing

$\boldsymbol{u_e}$ non-additive genetic effects can be responsible for specific adaptation

$\boldsymbol{u_g}$ total genetic effects predict the overall performance of genotypes in the target environments.

Used for selecting genotypes for promotion in the breeding program and eventual release to industry as varieties.

$$\boldsymbol{u_g = u_a + u_e}$$

Oakey et al (2006, 2007)

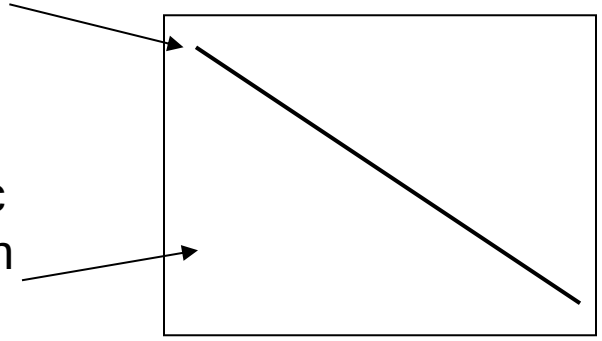
$$var(\boldsymbol{u_g}) = \boldsymbol{G_a \otimes A + G_e \otimes I_m}$$

$\boldsymbol{A}$: - additive relationship matrix
- relationship matrix derived from the pedigrees of the genotypes
- reflects the expected degree of co-ancestry among the genotypes

$\boldsymbol{G_a}$ and $\boldsymbol{G_e}$: additive and non-additive genetic variance matrices across trials

Diagonal: genetic variances for individual trials

Off-diagonal: genetic covariances between pairs of trials



txt

FA model to estimate $\boldsymbol{G_a}$ and $\boldsymbol{G_e}$

$\boldsymbol{\mathcal{G}}$: - genomic relationship matrix

- relationship matrix derived from the molecular markers of the genotypes
- its elements reflect the actual proportion of the genome that is identical by state between pairs of individuals

$$var(\boldsymbol{u_g}) = \boldsymbol{G_a} \otimes \boldsymbol{\mathcal{G}} + \boldsymbol{G_e} \otimes \boldsymbol{I_m}$$

Form of $\boldsymbol{\mathcal{G}}$ which ensures that interpretation of the predicted genetic effects remains unchanged – compared to when using $\boldsymbol{A}$

VanRaden (2008)

$$\boldsymbol{\mathcal{G}} = \frac{\boldsymbol{\Pi\Pi'}}{2\sum p_k(1-p_k)}$$

$\boldsymbol{\Pi}$ is the zero centred matrix of allele effects

Divisor equal to twice the sum of the variances of the markers scales $\boldsymbol{\Pi\Pi'}$

A.mat function in the R package rrBLUP calculates $\boldsymbol{G}$ (Endelman 2011)

- Two methods for imputation of missing marker values:
  marker mean or EM algorithm

The zero-centred $\boldsymbol{G}$ matrix is singular and therefore non-invertible but ASReml 4 can use it when fitting the linear mixed model

Cluster analysis was applied to the additive and total genetic correlation matrices to identify groups of environments where genotypes performed similarly.

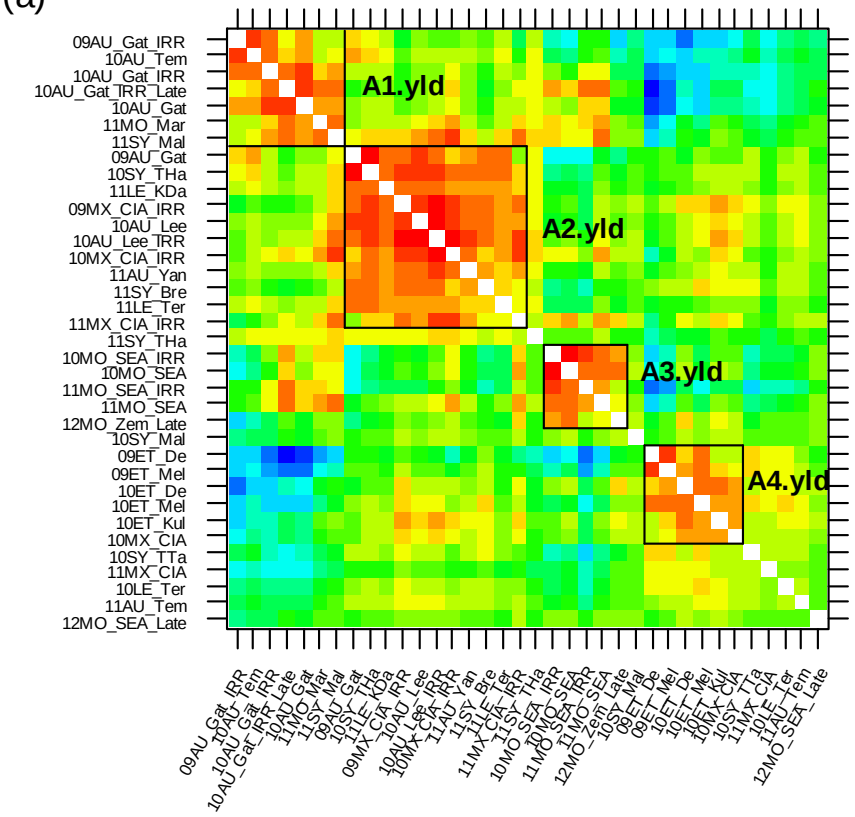
Heatmaps for graphical representation.

E-BLUPS for additive and total effects were obtained from the model.

Practical uses of the results (yield)

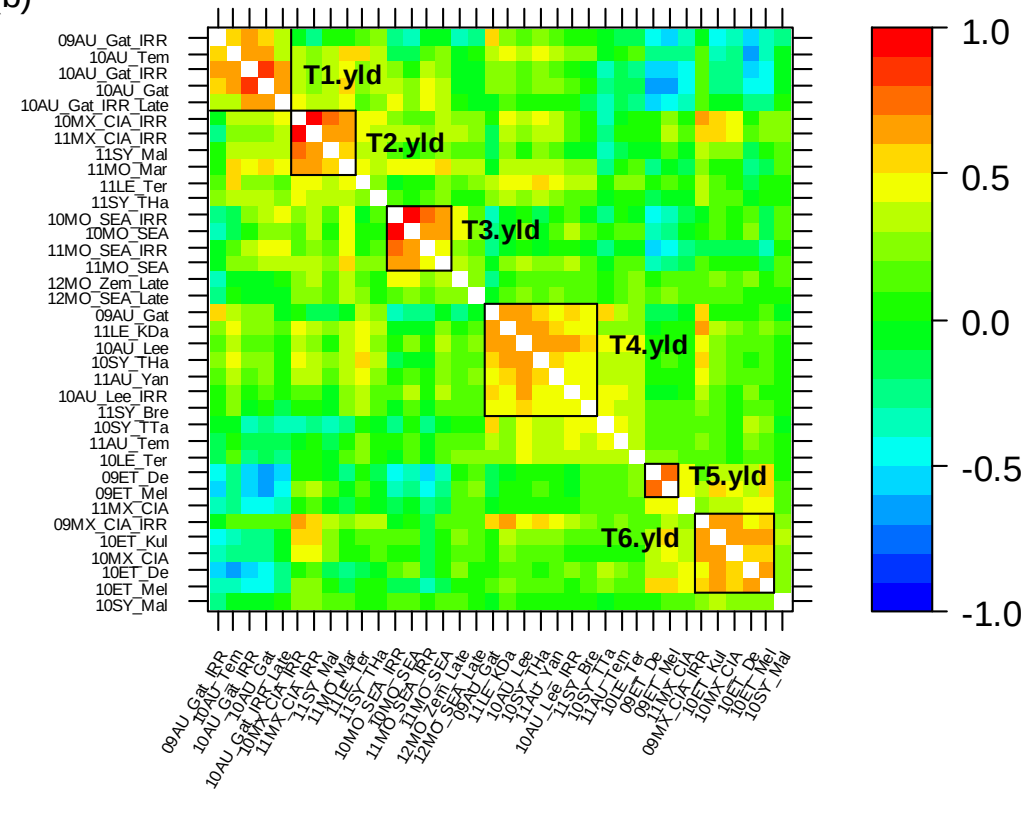
Estimated **additive** gen. correl. matrix

(a)



Estimated **total** gen. correl. matrix

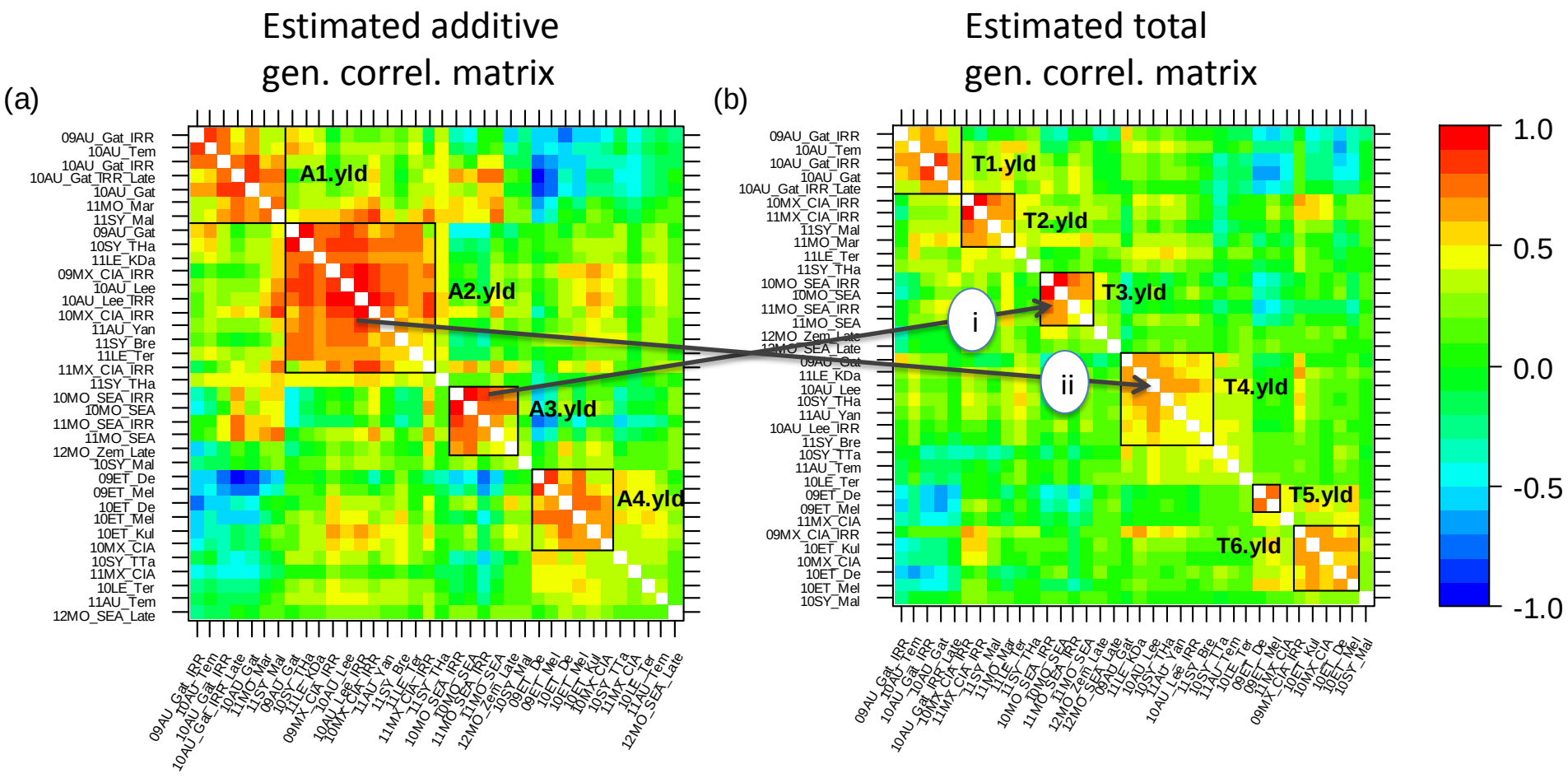
(b)



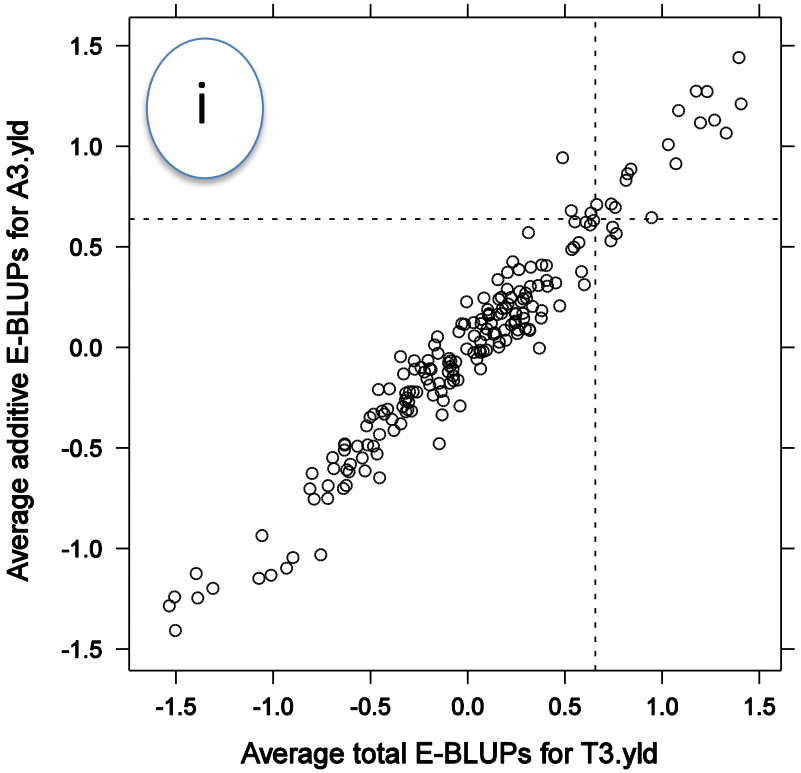
Trial	Total genetic variance	% additive gen. var.	%vaf FA3 additive	% non-add. gen. var.	%vaf FA2 non-additive
1					
2					
3					
4					
…					
…					
35					
36					
37					

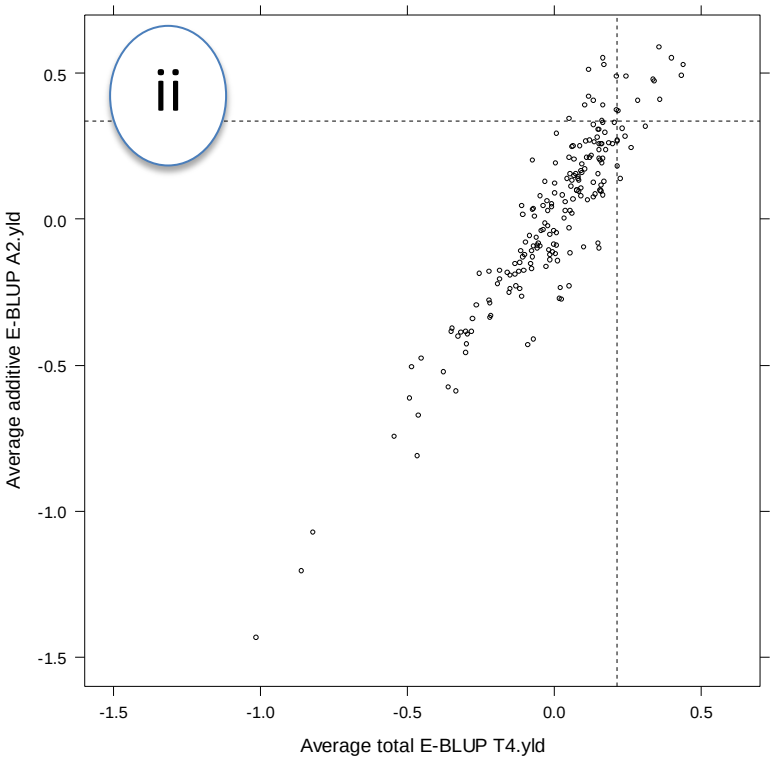
Trial	Total genetic variance	% additive gen. var.	%vaf FA3 additive	% non-add. gen. var.	%vaf FA2 non-additive
1	i	HIGH	HIGH	LOW	HIGH OR LOW
2					
3					
4	ii	MOD.	HIGH	MOD.	LOW
…					
…					
35					
36					
37					

Additive genetic effects dominate the genotype rankings



When additive effects dominate the genotype rankings, trials belong to the same environment groups for additive and total effects.





Trials with:

- high (68-97%) additive genetic variance well accounted for by the model
- low non-additive genetic variance

17 out of the top 20 selections coincide

Trials with:

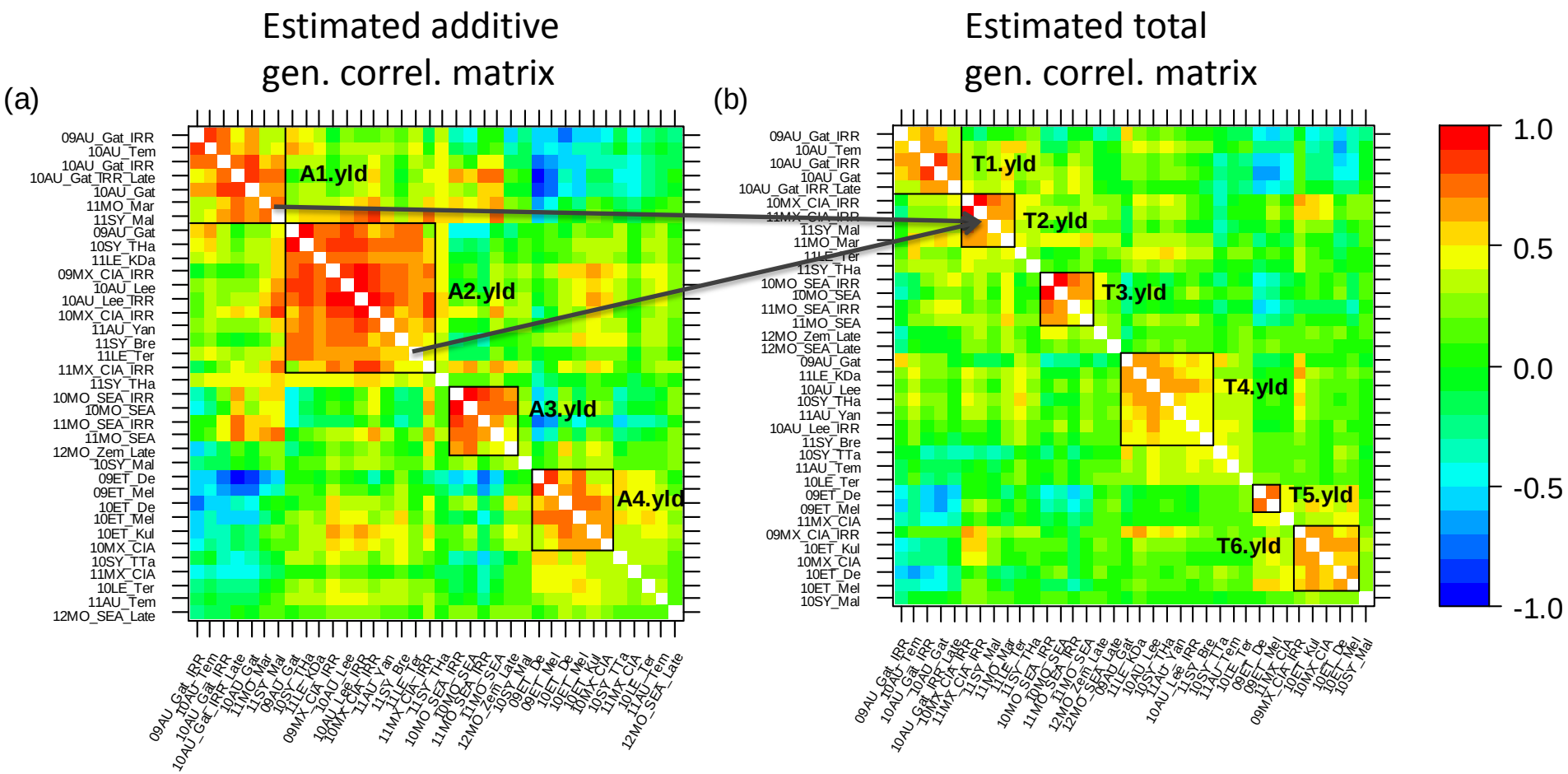
- moderate (32-75%) additive genetic variance well accounted for by the model
- moderate (25-68%) non-additive genetic variance only partially accounted for by the model

10 out of the top 20 selections coincide

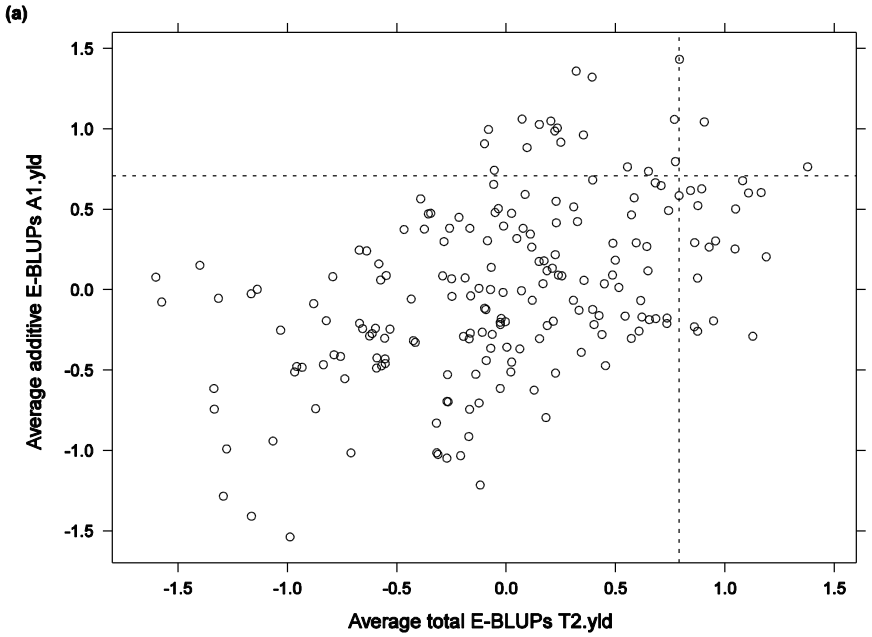
Trial	Total genetic variance	% additive gen. var.	%vaf FA3 additive	% non-add. gen. var.	%vaf FA2 non-additive
1	i	HIGH	HIGH	LOW	HIGH OR LOW
2					
3					
4	ii	MOD.	HIGH	MOD.	LOW
…					
…					
35	iii	MOD.	RELATIVELY HIGH	MOD.	HIGH
36					
37					

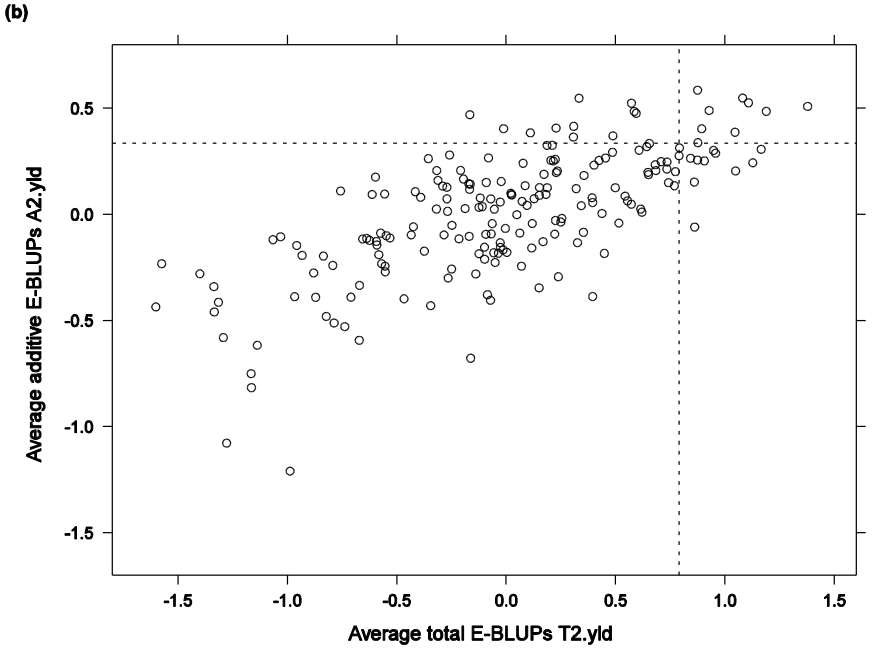
Additive genetic effects dominate the genotype rankings

Non-additive genetic effects change the pattern of genotype rankings



When non-additive effects change the genotype rankings, trials belong to different environment groups for additive and total effects.





Trials with:

- moderate (22-60%) additive genetic variance well accounted for by the model
- moderate (40-78%) non-additive genetic variance well accounted for by the model

2 and 8 genotypes identified as best parents in A1.yld and A2.yld coincide with the top 20 yielders in T2.yld.

Best potential parents and best yielders generally do not overlap

Conclusions

- In the presence of strong GxE the best is to select parents from environment groups for additive effects, and best yielding individuals from environment groups for total effects.
- Including $\boldsymbol{G}$ in the linear mixed model accounts for the relationships amongst the genotypes under consideration and allows partitioning the genetic effects in a way familiar to plant breeders.
- This method of analysis could be readily implemented by any keen statistician.

Acknowledgements

Co-authors:
David Butler (QLD Dept. of Agriculture and Fisheries)
Francis Ogbonnaya (GRDC and ICARDA)
M. Fernanda Dreccer (CSIRO)

Arthur Gilmour for his invaluable help with coding in ASReml 4

Funding:
Generation Challenge Program (Mexico)
Grains Research and Development Corporation (Australia)
International Centre for Agricultural Research in the Dry Areas (ICARDA)

Contributions to the project and running field trials:
Farid Makdis (ICARDA), Hassan Ouabbou (INRA-CRRA), Solomon Gelacha and Firdissa Eticha (EIAR), Tony Condon (CSIRO) and Matthew Reynolds (CIMMYT)

References

Smith A, Cullis B, Thompson R (2001) Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138-1147

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modeling of additive and non-additive genetic line effects in single field trials. TAG 113:809-819

Oakey H, Verbyla AP, Cullis BR, Wei XM, Pitchford WS (2007) Joint modeling of additive and non-additive (genetic line) effects in multi-environment trials. TAG 114:1319-1332

VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science 91:4414-4423

Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4:250-255