



QTLxE, and MAS?

An eco-physiological - statistical framework for the analysis of GxE and QTLxE as occurring in abiotic stress trials, with applications to the CIMMYT drought stress programs in tropical maize and bread wheat

Fred van Eeuwijk

GCP #4, Benoni, South Africa, September 2007



People involved

- Post-Docs
 - Ky Mathews (CSIRO) & Marcos Malosetti (WUR)
- PI
 - Fred van Eeuwijk (WUR)
- Co-PI
 - Scott Chapman (CSIRO)
 - Matthew Reynolds (CIMMYT) & Jean-Marcel Ribaut (GCP)
- Further Collaborators
 - José Crossa (CIMMYT)
 - Mateo Vargas (Univ. Chapingo, Mx.)
 - Sergio Ceretta (INIA, Uruguay)
 - Martin Boer (WUR)

Objectives

- The development of an eco-physiological statistical framework for the simultaneous analysis of **GxE and QTLxE** in data from abiotic stress breeding programs, adding value to existing data sets
- The statistical modeling should account for
 - plot heterogeneity
 - multiple environments (GxE, QTLxE)
 - multiple traits
 - **multiple crosses**
- Course material making the developed methodology accessible to GCP members and others interested, including documentation and free software (Genstat Discovery / R).



Products/ deliverables 1

- Mixed model methodology for multi-trait multi-environment QTL analyses has been developed and tested in applications to CIMMYT maize and wheat data
- Papers explaining the methodology and its application have been published/ are in press/ are about to be submitted (next slide)
- Course material has been developed and tested in four courses (Spain, 3x South America)
- Genstat Discovery programs are available (GCP-GDP)



Publications

(for pdf's, marcos.malosetti@wur.nl)



- Malosetti M, Boer MP, Bink MCAM, van Eeuwijk FA (2006) Multi-trait QTL analysis based on mixed models with parsimonious covariance matrices. In: Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, August 13-18, Belo Horizonte, MG, Brasil. <http://www.wcgalp8.org.br/wcgalp8/> Article 25-04
- Malosetti M, Ribaut JM, Vargas M, Crossa J, Boer MP, van Eeuwijk FA (2007) Multi-trait multi-environment QTL modelling for drought-stress adaptation in maize. In: JHJ Spiertz, Struik PC, van Laar HH (Eds.), Scale and Complexity in Plant Systems Research. Gene-Plant-Crop Relations, pp. 25-36. Springer, Dordrecht, The Netherlands.
- van Eeuwijk FA, Malosetti M, Boer MP (2007) Modelling the genetic basis of response curves underlying genotype x environment interaction. In: JHJ Spiertz, Struik PC, van Laar HH (Eds.), Scale and Complexity in Plant Systems Research. Gene-Plant-Crop Relations, pp. 115-126. Springer, Dordrecht, The Netherlands.
- Malosetti M, Ribaut JM, Vargas M, Crossa J, van Eeuwijk FA (2007) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (*Zea mays* L.). *Euphytica* (in press)
- Martin Boer, Deanne Wright, Lizhi Feng, Dean Podlich, Lang Luo, Mark Cooper, Fred van Eeuwijk (2008) A mixed model QTL analysis for multiple environment trial data using environmental covariables for QTLx E , with an example in maize. *Genetics* (in press).
- Mathews KL, Malosetti M, Chapman SC, McIntyre L, Reynolds MP, Shorter R, van Eeuwijk FA Multi-environment QTL mixed models for drought stress adaptation in wheat. Submitted.

Courses (all with CD)

(for more information, marcos.malosetti@wur.nl)



- **Course:** “Design and analysis of multi-environment trials: conventional and QTL-based methods”
Date: 12-23 September, 2005
Place: Zaragoza, Spain
web: <http://www.generationcp.org/capcorner.php?da=0531820>
- **Course:** “Phenotypic and marker based analysis of multi-environment data: Identifying the genetic basis of GxE and responses to stress”.
Date: September 11th, 2006
Place: São Paulo, Brazil
web: <http://www.generationcp.org/capcorner.php?da=0531820>
- **Course:** “Phenotypic and marker based analysis of multi-environment data: Identifying the genetic basis of GxE and responses to stress”.
Date: 12-13 July, 2006
Place: Colonia, Uruguay
web: <http://www.inia.org.uy/online/site/32454811.php>
- **Course:** “Phenotypic and marker based analysis of multi-environment data: Identifying the genetic basis of GxE and responses to stress”.
Date: 17-20 July, 2006
Place: Piracicaba (SP), Brazil
web: <http://www.genetica.esalq.usp.br/en/qtls/>

Research and education



- A general approach to the analysis of GxE and QTLxE was developed over the last two years by continuous interaction with
 - Relevant data
 - CIMMYT maize
 - CIMMYT / CSIRO wheat
 - Interested students
- An overview of the approach for use in research as well as in teaching is given on the next 5 slides



Modeling GxE and QTLxE: 1

- **Analysis of single traits in single environments:
Producing reliable genotypic means + weights**
 - Pay attention to the **design** of the experiment
 - replicates, rows, columns, incomplete blocks
 - Take care to model **plot heterogeneity** and trends
 - Spatial modeling of error variation (AR x AR)
 - Post blocking
 - Check **model assumptions** & identify outliers (**data quality**)
 - Produce **summary statistics**
 - BLUEs (genotypic means)
 - Standard Errors
 - Evaluate the efficiency of the design on the basis of the estimates for error variance components, spatial correlations and SED's

Modeling GxE and QTLxE: 2



- **Analysis of single traits in multiple environments: Modeling GxE**
 - Construct **GxE table of means** and corresponding table of SEs (used in weights in subsequent GxE and QTLxE analyses)
 - Fit **mixed model**: $P = E + G + GE = E + GGE$
 - Take **E fixed** and **GGE random**
 - Assess **best model for VCOV of GGE** using AIC/BIC
 - Model heterogeneity of genetic variances across environments
 - Model heterogeneity of genetic correlations between environments
 - Assess which environments are high/low variance, calculate some form of heritability
 - Assess which pairs of environments are (dis)similar

Modeling GxE and QTLxE: 3



- **Single trait QTL mapping across multiple environments: Modeling QTLxE**
 - Inspect **genotypic score matrix**
 - Dimensions: genotypes x markers
 - Create **genetic predictors**, i.e., markers + virtual markers from the genotypic score matrix
 - Dimensions: genotypes x genomic evaluation points
 - **Scan genome** for QTLs by fitting mixed model
 - **$P = E + QQE + GGE$**
 - with $QQE = \text{QTL main effect} + \text{QTLxE}$
 - QQE represents QTL with effects that depend on environment
 - QTL effects are estimated with respect to genetic predictors
 - E and QQE are fixed, GGE is random according to model assessed in modeling step 2

Modeling GxE and QTLxE: 4



● Refinements

- Constructing **multiple QTL models**
 - One or more rounds of CIM
 - Backward selection from set of already identified QTLs
- Testing for **QTLxE** to see whether QGE can be reduced to just QTL main effect
- **Regressing QTLxE** on environmental characterizations
 - This step is equivalent to fitting **QTLs for response curves** across environments (c.f. Tardieu)
- **Validating** effects of **candidate genes** and QTLs
- Including **genotypic covariables** to correct for disturbing factors
 - Example in wheat: correction for disease resistance in QTLxE analysis of yield
- Including **dominance** and **epistasis**

Modeling GxE and QTLxE: 5



- **Further extensions**

- Instead of working with GxE summary table of means for MET analysis, **work with the plot observations** and include the error models for individual trials in the MET models (phenotypic and QTL)
 - Advantage: better weights for individual genotypes in individual trials
 - Disadvantage: logistically and numerically harder
- **Multi-trait multi-environment** QTL mapping
 - Model genetic correlations in their dependence across environments (in addition to genetic variances)
 - Distinguish pleiotropy from linkage as basis for genetic correlation
 - Correct yield and yield components for phenology
- **Introducing multiple crosses (QTL x background)**



Wheat: some typical output

Multi-environment QTL analysis with additional genotypic covariables

Seri x Babax, 6 environments

Yield QTLs depend strongly on the environment
(QQE effects)



Environment ^a	Rye. Env ^b	1D-a	1D-b	4B-b	5A-a	6B-a	6D-a	7A-a	7B-a
Main							-6.11		
BILO05	-14.83	11.71	-14.68	6.51	-5.15	-12.18		6.20	7.71
LUND05	-22.74	-3.00	-2.85	5.63	-9.09	-4.40		8.90	7.08
GATD02	-14.15	-1.04	-1.96	-2.38	-1.00	-7.52		8.87	-3.00
GATD04	-1.42	-0.72	-4.73	-0.95	2.55	-2.34		3.95	-3.92
GATD05	-13.27	0.33	-1.03	6.81	-0.54	-2.13		2.25	3.75
GATI06	-20.72	1.27	6.84	13.78	-0.44	0.87		18.92	-7.04
(avsed or s.e.)	(8.0)	(4.7)	(4.4)	(4.4)	(4.7)	(4.4)	(1.5)	(4.3)	(4.3)

Yield QTLs are small and depend strongly on the environment (% explained variance)



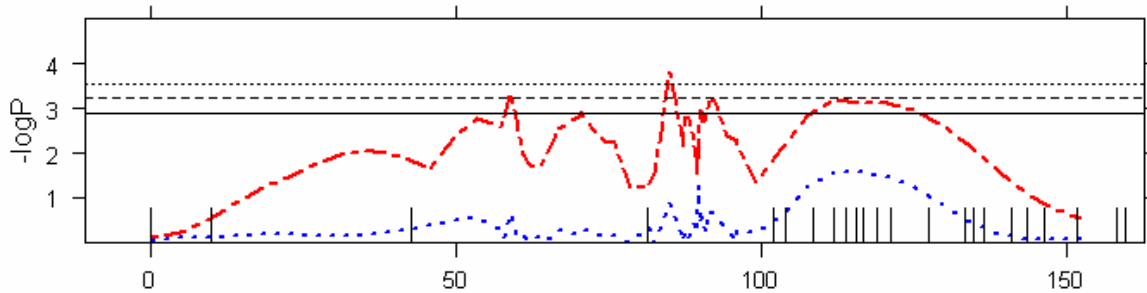
Table 1: Trial genetic variance for the GGE model and the percentage of variance explained by the multi-QTL model, S_Q . The percentage of variance explained for each QTL in each environment, expressed as a function of the GGE and final QQE models (S_q^{GGE} , S_q^{QQE}).

Environment ^a	genetic variance, GGE model	S_Q , % explained by QQE	Rye.Env ^b	1D-a	1D-b	4B-b	5A-a	6B-a	6D-a	7A-a	7B-a
BILO05	2090	28	0.5, 2.1	3.3, 6.6	6.9, 13.6	0.0, 2.0	1.4, 0.6	5.7, 9.0	0, 0	0.8, 1.7	1.2, 3.2
LUND05	1065	32	4.4, 13.4	0.2, 0.4	0.9, 1.0	1.3, 3.8	5.1, 9.6	0.4, 2.2	0.0, 0.5	4.5, 10.0	4.4, 6.3
GATD02	560	40	3.9, 10.1	0, 0	0.2, 0.0	2.6, 0.1	0, 0	6.5, 15.9	6.7, 13.2	8, 21.9	0.7, 1.3
GATD04	446	24	0, 0	0, 0	4.7, 5.8	0.8, 0.0	0.7, 0.7	0.4, 0.1	12.6, 15.7	1.4, 4.0	2.7, 2.5
GATD05	987	0	2, 2.2	0, 0	0, 0	3, 3.6	0, 0	0, 0	0, 0	0, 0	0, 0
GATI06	3623	11	1.3, 1.9	0, 0	0.8, 1.0	1.6, 4.6	0, 0	0, 0	0.5, 1.6	4.6, 8.5	0.0, 0.6

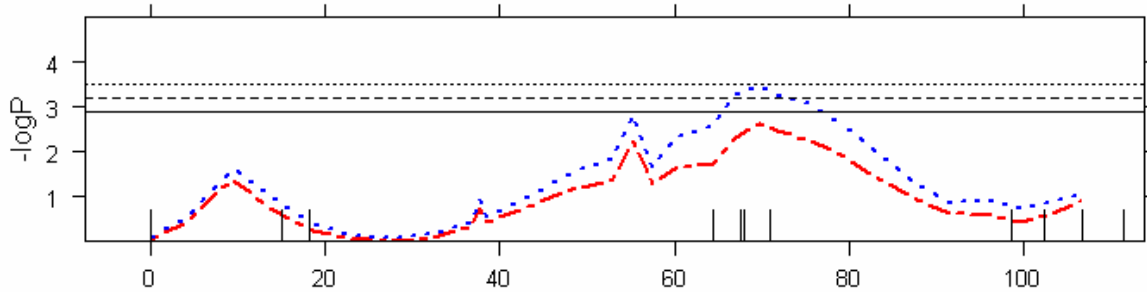
$-\log_{10}(p)$ profile for 1B-a, 5A-a and 7A-a illustrating the effect the factor, *rye* on yield. Significance levels; solid = 0.20, dashed = 0. 10, dotted = 0. 05.



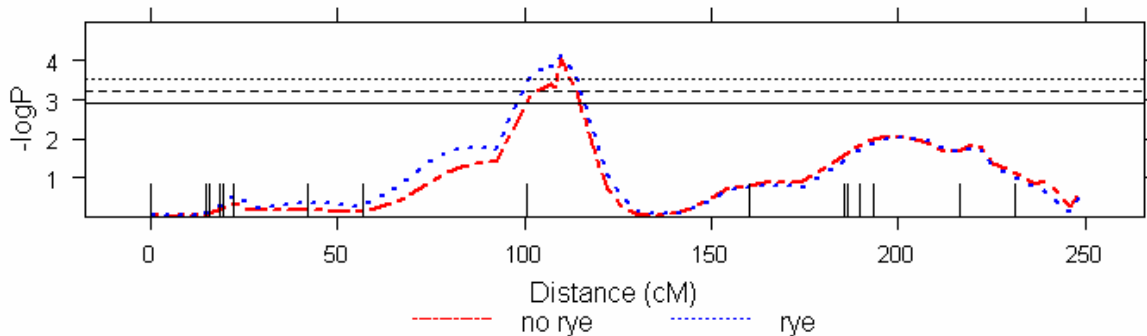
1B-a



5A-a



7A-a

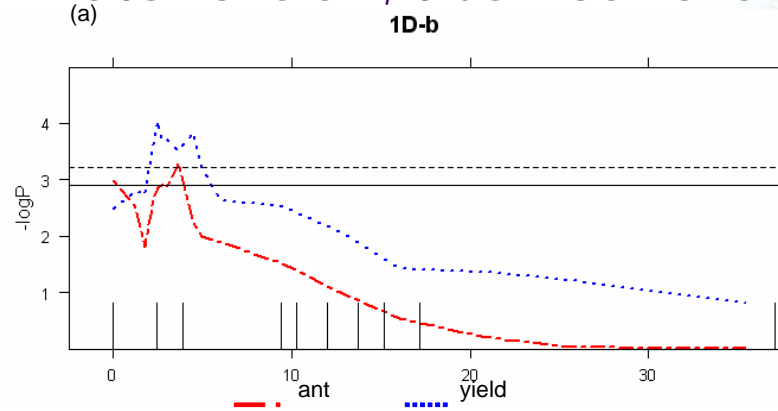


Introducing genotypic covariables to correct for disturbing factors in QTLx E analysis for yield

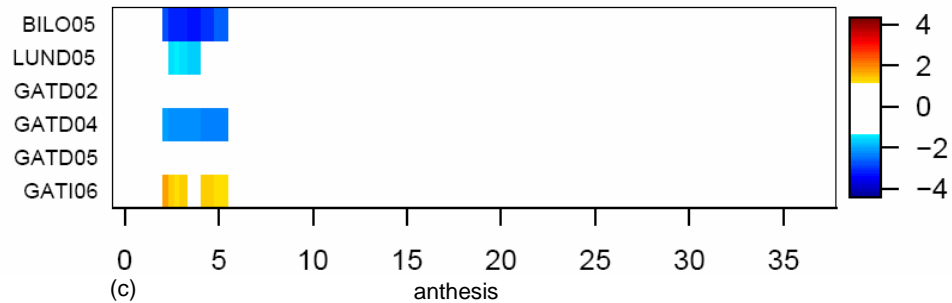
Co-incidence of yield and anthesis QTL and specific environment effects for linkage group 1D-b



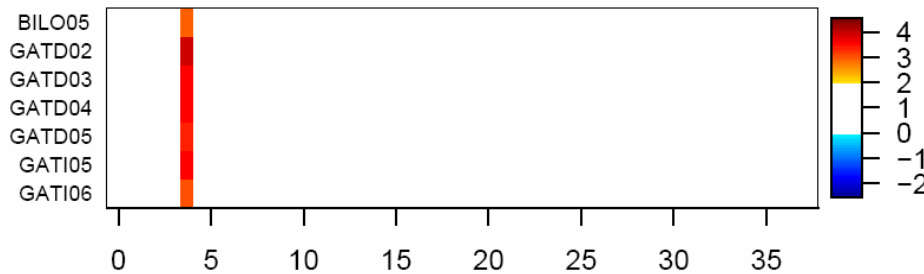
QQE image plots: red = Babax allele +, blue = Seri allele +



(b) yield



(c)



Cross over
QTLxE



Maize: some typical output

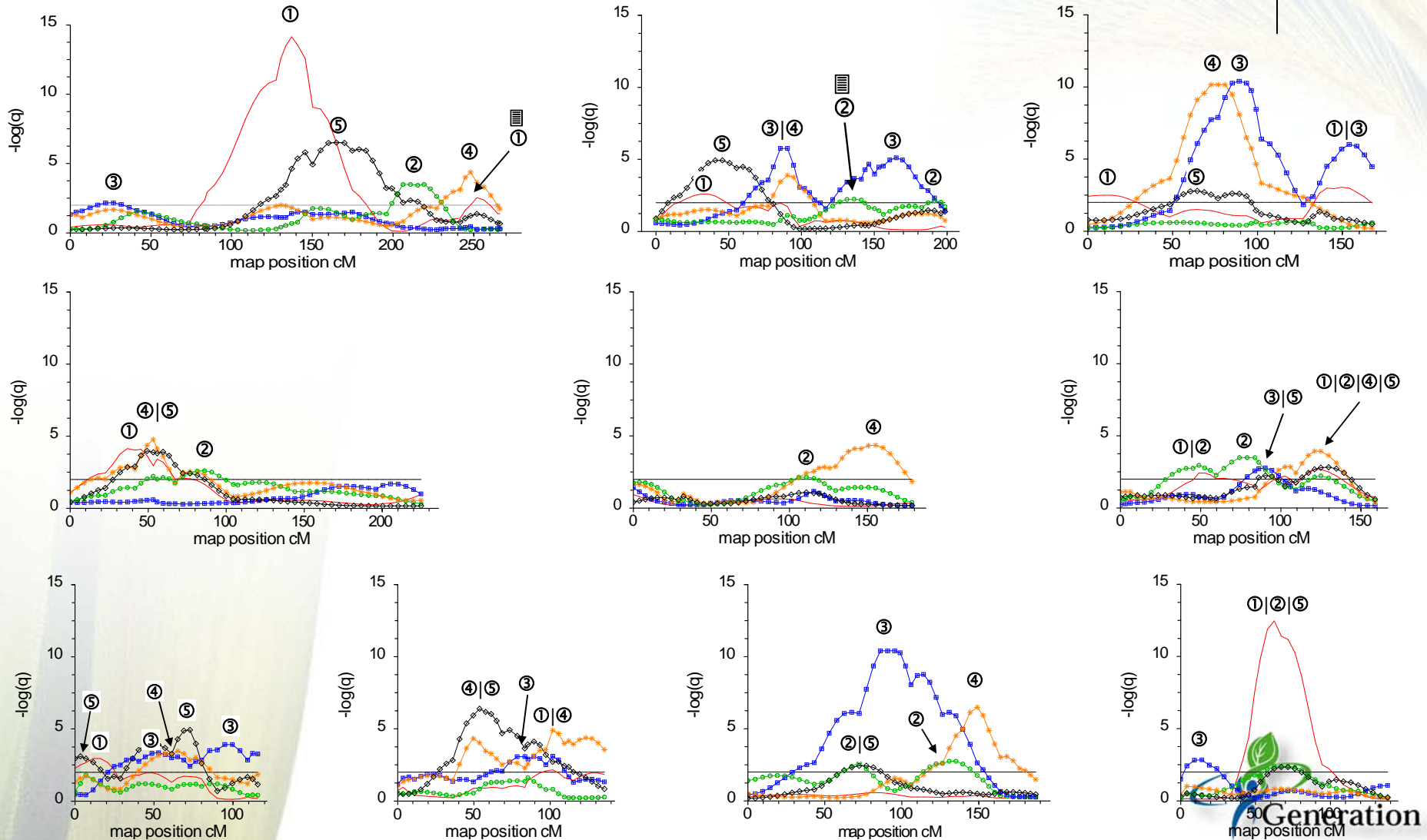
Multi-trait multi-environment QTL analysis with separate modeling of relations between environments and between traits

P1xP2, 8 environments, 5 traits (Yield, ASI, Male flowering, Nr of ears, Plant height)

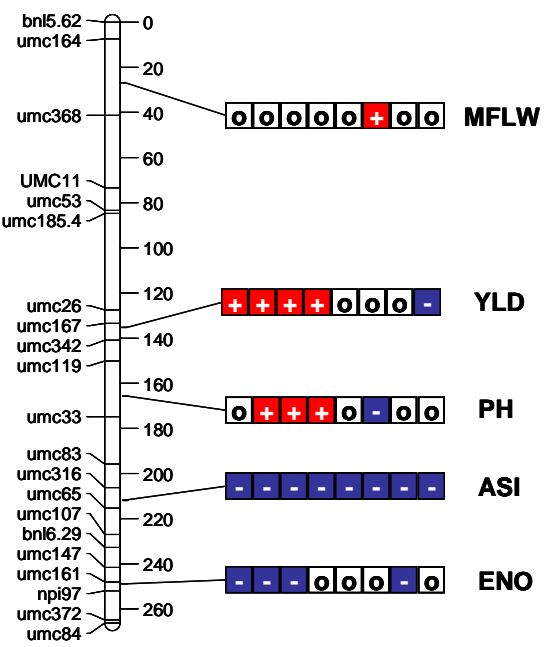


P1xP2, 8 environments, 5 traits

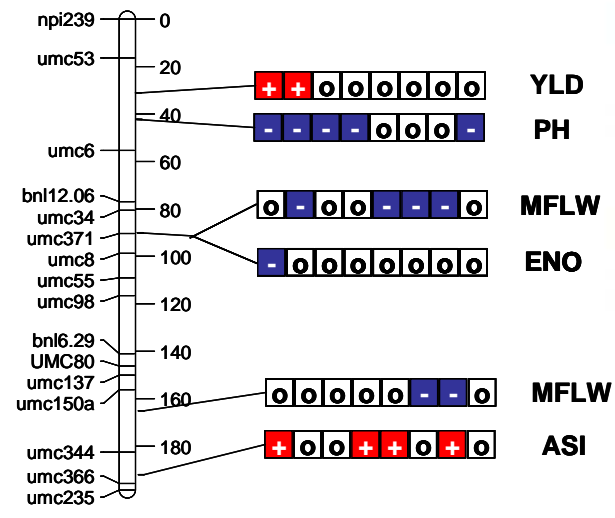
— YLD
—○— ASI
—□— MFLW
—*— ENO
—◇— PH



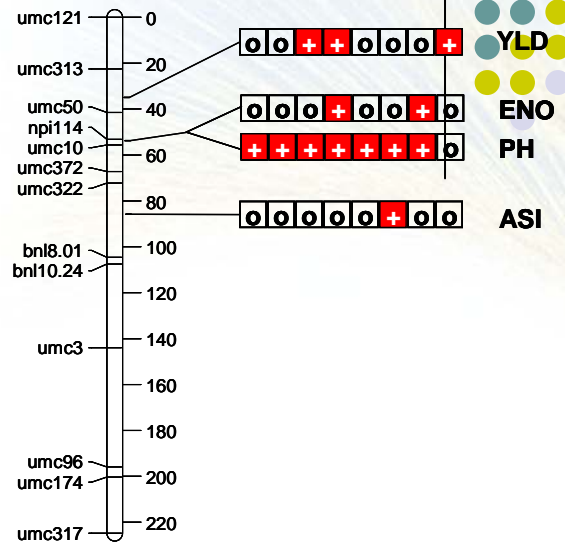
Chromosome 1 NS92a, IS92a, SS92a, IS94a, SS94a, LN96a, LN96b, HN96b



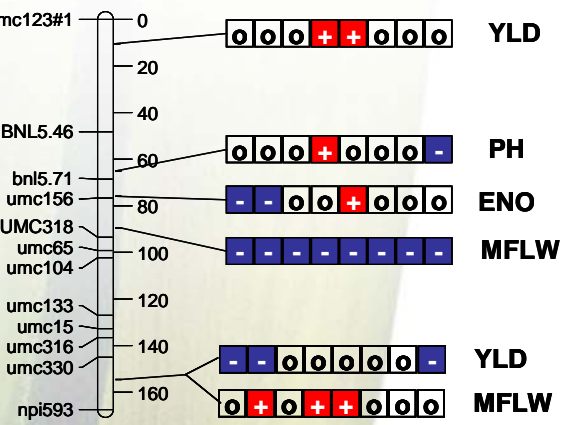
Chromosome 2



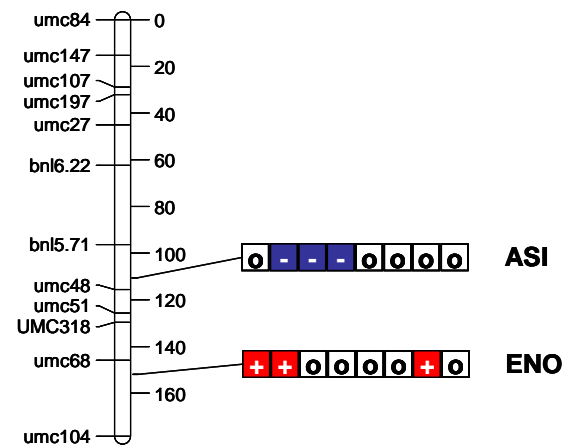
Chromosome 3



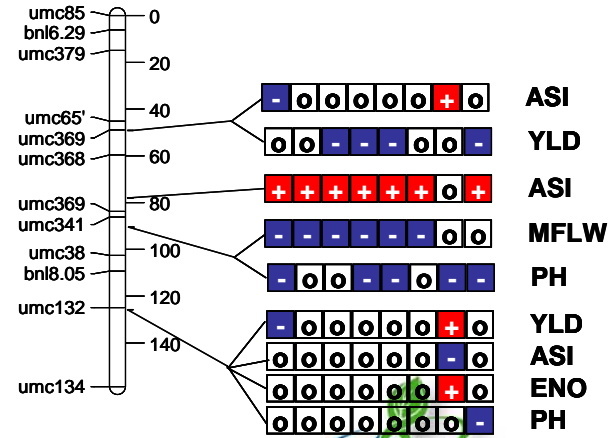
Chromosome 4



Chromosome 5

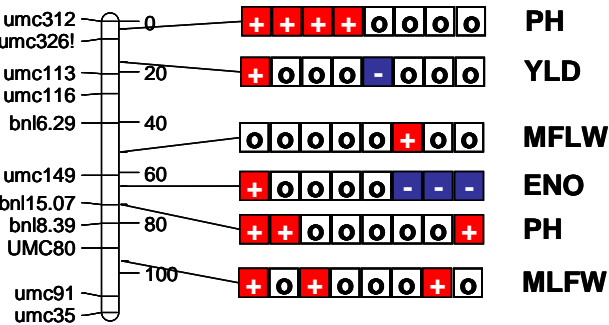


Chromosome 6

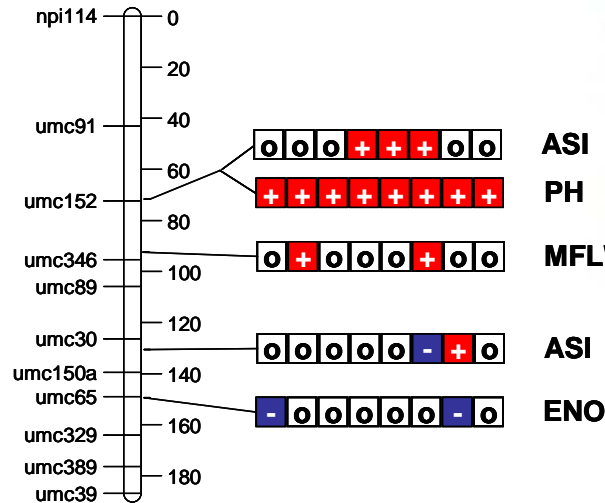




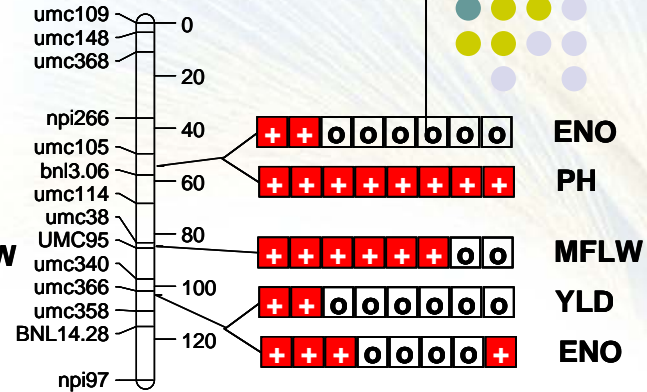
Chromosome 7



Chromosome 8



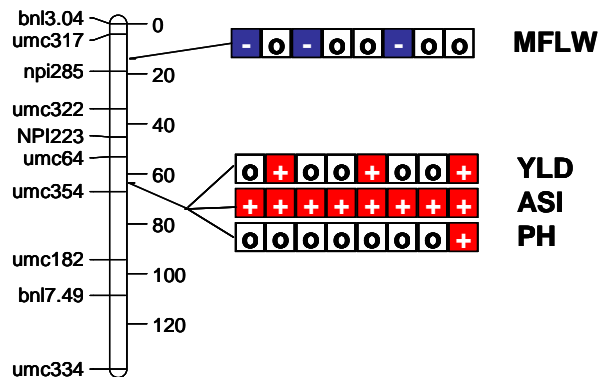
Chromosome 9



Strong inconsistencies

- For the same trait across environments, including cross overs (QTLx_E)
- For pairs of traits across environments (Changing r_G)

Chromosome 10



Deliverables in relation to the model crops: maize & wheat (to be ctd.)



- Identification and estimation of *QTL locations and effects* for productivity under drought
- Assessment of the *stability* of QTL expression across environments and genetic backgrounds
- Models for *QTLxE* in relation to *environmental characterizations*
- Insight in the *genetic basis of correlations* between traits
- Assessment of *pleiotropic / linkage* relations between QTLs

QTLxE and MAS

- QTLxE is the rule, a QTL main effect is an exception (small populations / homogeneous set of environments).
- Given the enormous inconsistency of QTLs across environments and traits, what kind of MAS can help us to improve yield?
- There is no decent theory for multi-trait MAS in the face of QTLxE and changing genetic correlations between traits across environments
- To assess the consequences of QTLxE work on selection response theory is needed

Planning

- Further development of the methodology
 - Comparison of multi-trait multi-environment QTLxE analyses between 3 biparental populations of maize (paper)
 - Multi-trait (including canopy temperature) multi-environment analysis QTLxE in wheat (paper)
- Further development of course material
 - Manual for QTLxE analysis (Chapter in GCP phenotyping book)
 - A set of Genstat Discovery procedures for QTLxE analysis (maybe also R)
 - One or two QTLxE courses (Africa /China)

Relation with other GCP work



- QTLxE courses in collaboration with SP5
- QSS (QTLxE Support System)
 - Whole crop modeling in maize (Francois Tardieu)
 - Aluminum tolerance in sorghum (Jurandir Magalhães)
 - Durum wheat (Roberto Tuberosa)
 - Others?