High througput phenotyping and its role in a strategy for improving complex traits

Fred van Eeuwijk & EU-SPICY consortium

PhenoDays, Wageningen, October 2011

WAGENINGEN UNIVERSITY
WAGENINGEN UR

spicy

Smart tools for Prediction and Improvement of Crop Yield

New genotype-to-phenotype models
at the intersection of
genetics, physiology and statistics

WAGENINGEN UNIVERSITY
WAGENINGEN UR

spicy

Aim EU project SPICY (KBBE-2008-211347)

- Predictive models for phenotypic variation in complex traits
  - Take into account
    - a range of environmental conditions
    - multiple traits
    - developmental time
  - Use as input
    - molecular marker variation
      - DNA polymorphisms
    - environmental inputs
      - temperature, humidity, light, nutrients, …
    - genomic information
      - gene expression, metabolites, proteins, …
    - intermediate phenotypes / approximate phenotypes
      - image analysis, fluorescence measurements

SPICY participating organizations & PI's

- Wageningen University (WU) – WP 0/2/3/4
  - F.A. van Eeuwijk
- Plant Research International (PRI) – WP 0/2/3/4
  - A. Dieleman
- Institut National de la Récherche Agronomique (INRA) – WP 1/3
  - A. Palloix
- Flemish Institute for Biotechnology (VIB) – WP 1/5
  - M. Vuylsteke
- Estación Experimental de la Fundación Cajamar (EEFC) WP 2/4
  - J. Magán
- Budapest University of Technology and Economics (BME) – WP 2
  - A. Baróczi
- Scottish Crop Research Institute (SCRI) – WP 2
  - C. Glasbey

- 0 = coordination; 1 = genomics; 2 = phenotyping; 3 = modeling; 4 = validation; 5 = dissemination

Industrial Advisory Board

- Participating companies
  - ENZA
  - LemnaTec
  - Limagrain
  - Monsanto
    - De Ruiter
    - Seminis
  - Nunhems
  - Ramiro Arnedo
  - Rijk Zwaan
  - Syngenta
- Tasks
  - At start project
    - Discussing aims and deliverables
  - During project
    - Discussing results in relation to current developments in the European breeding industry & suggesting improvements

WELCOME TO THE SPICY WEB

Welcome

This website was generated to provide both the scientific community and the general public background information about the SPICY project.

The aim of the **S**mart tools for **P**rediction (and) **I**mprovement of **C**rop **Y**ield project is to develop a suite of tools based on molecular breeding to help breeders in predicting phenotypic response of genotypes for complex traits like yield under a range of environmental conditions.

The project particularly focuses on developing tools for predicting phenotypic performance (growth, yield) of a genotype by means of an integrated gene-to-phenotype model, thereby reducing the effort of phenotyping new genotypes. This will be achieved by extending a crop growth model with a separation between species specific and genotype (variety, line) specific parameters and the development of smart tools for integrating QTL analysis with crop growth models. In addition, gene expression studies will try to identify genes within the QTL while imaging and fluorescence tools will be developed for fast and automated large scale phenotyping.

SPICY is supported by the European Comunity and funded by the KBBE FP7 programme.
(Grant agreement number *KBBE-2008-211347*)

WAGENINGEN UR
For quality of life

www.spicyweb.eu

BioSS

spicy

Philosophy of SPICY

- Dissect complex traits (yield) with context dependent variation into component traits that
  - do *not* show strong context dependencies
    - no QTLxE, no QTLxQTL, no QTLxQTLxE
  - can be integrated with environmental information over the growing season to produce yield
    - which Genotype-to-Phenotype (G2P) model to use?
  - have relatively simple genetic basis
    - few additive genes/QTLs

G2P models, more generally

- G2P models predict the phenotypic performance of organisms on the basis of genetic, genotypic, genomic and environmental information
- G2P models help in defining strategies for improvement of plant performance

Some G2P models (ignoring error variation)

- $P = G + E + GxE$
  - GxE = genotype by environment interaction
- $P = G + \beta Z$
  - Z = environmental quality
- $P = f(\mathbf{G},\mathbf{Z})$
  - f: non-linear function
  - **G**: vector of genotype related parameters
  - **Z**: environmental triggers/ inputs
- $P_t = \int f_t(\mathbf{G}_t,\mathbf{Z}_t)dt$
  - as above, but now with integration over time to account for development

Some QTL models as G2P models

- P = QTL + residual G + error

- P = ΣQTL + Σ (QTLxQTL) + residual G +
  - E +
  - Σ(QTLxE) + residual GxE +
  - error

- P = QTL + residual G +
  - E +
  - QTLxE$_{\text{(environmental covariable)}}$ + residual QTLxE +
  - residual GxE +
  - error

QTL by Environment interaction: genetic basis GEI

- Model for fitting environment-specific QTL effects;
  - $P = E + GGE = E + \Sigma(QTL_E) + \varepsilon$
    - GGE random
    - $VCOV(\varepsilon) = VCOV(GGE)$ = factor analytic
- Eco-physiological QTL model
  - Regression of environment specific QTL on environmental information
  - $QTL_E = QTL(Z_E) + \delta$

G2P models

- Statistical genetic models (REML and Bayesian mixed models)
  - multi-trait multi-environment QTL models/ hierarchical models
  - models for QTL x E, QTL x QTL, QTL x genetic_background
- Crop growth models
  - ordinary differential equation models
  - mechanistic model with modules for morphology, leaf area, photosynthesis, assimilate partitioning, nutrient dynamics and fruit growth dynamics
- Integration of statistical genetic and crop growth models
  - express individual physiological parameters in crop growth models in terms of underlying QTLs
  - impose network relations on multi-trait QTL models using prior physiological knowledge (Bayesian belief networks)

SPICY: phenotyping and molecular tools

- New phenotyping tools:
  - Image analysis tool (BIOSS and WUR)
  - Fast fluorescence tool (Univ Budapest)
- Role of image analysis/ fluorescence
  - Provide proxies to physiological components
  - Provide G2P inputs with high heritability (cf. correlated response theory)
  - Give details of the growing process over time to help in development of G2P models (provide a genotype specific movie)
- New molecular tools:
  - Approaches for finding candidate genes (INRA + VIB)
    - bio-informatic strategies on gene data bases of related species
  - eQTL mapping & genetic pathway reconstruction
    - co-localization of QTLs for gene expression, eQTLs, and QTLs for physiological parameters

Phenotyping experiments

- Genotypes
  - RIL-population, 149 genotypes
  - Parents: Yolo Wonder, CM 334
  - F1

- Two locations
  - Netherlands (Wageningen) & Spain (Almer
- Two seasons
  - Spring & Autumn

Relevant measurements

- Crop measurements
  - Plant weight (stem, leaves, fruit; initial and fin
  - Number of internodes
  - Leaf area (initial and final)
  - Fruit harvest (number, weight)

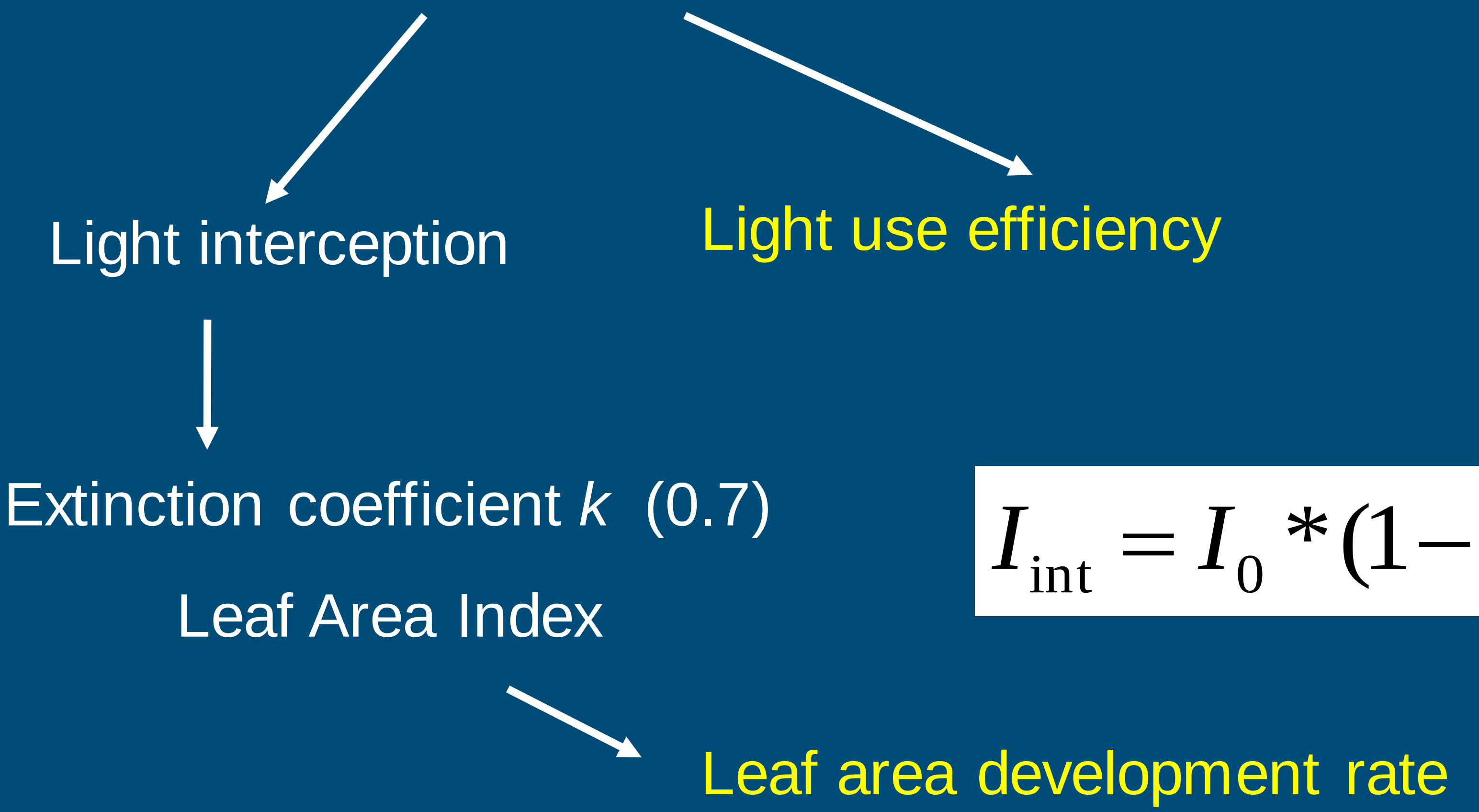
- Environment
  - Temperature
  - Radiation

‘ Simulating’ yield from identified QTLs

Yield of sweet pepper (g m^{-2})

Maaike Wubs
Ep Heuvelink

Yield = total biomass production * Partitioning into fruits



Light interception

Light use efficiency

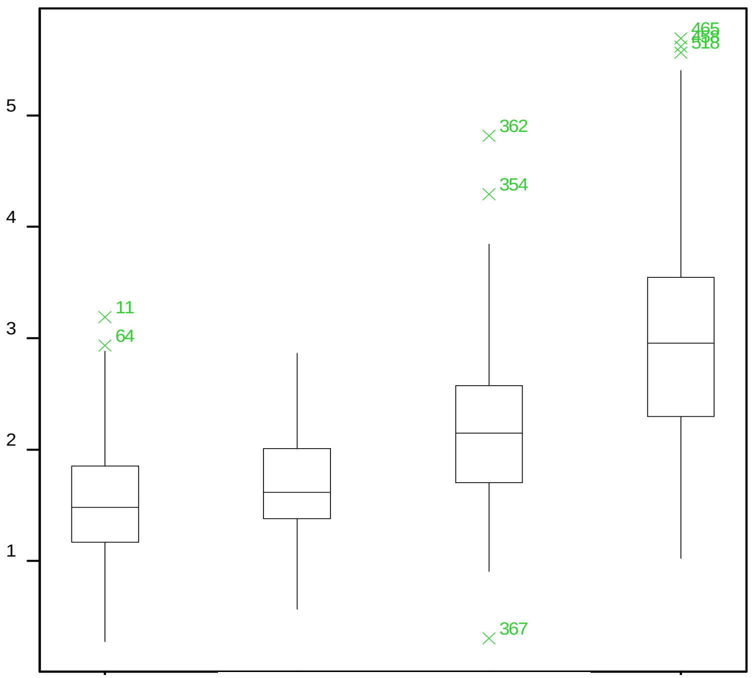
Extinction coefficient *k* (0.7)

Leaf Area Index

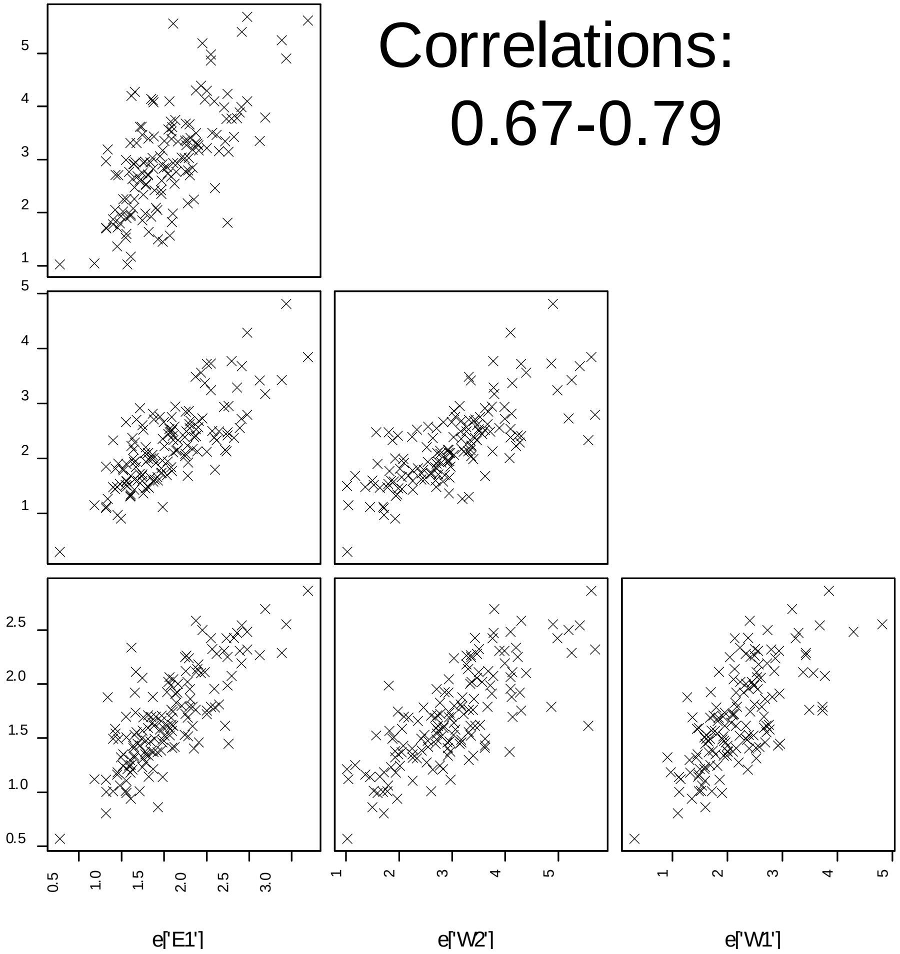
Leaf area development rate

$$I_{\text{int}} = I_0 * (1 - e^{-k*LAI})$$

Leaf development rate



Spain 1	Spain 2	Ned 1	Ned 2
0.9	0.94	0.71	0.88



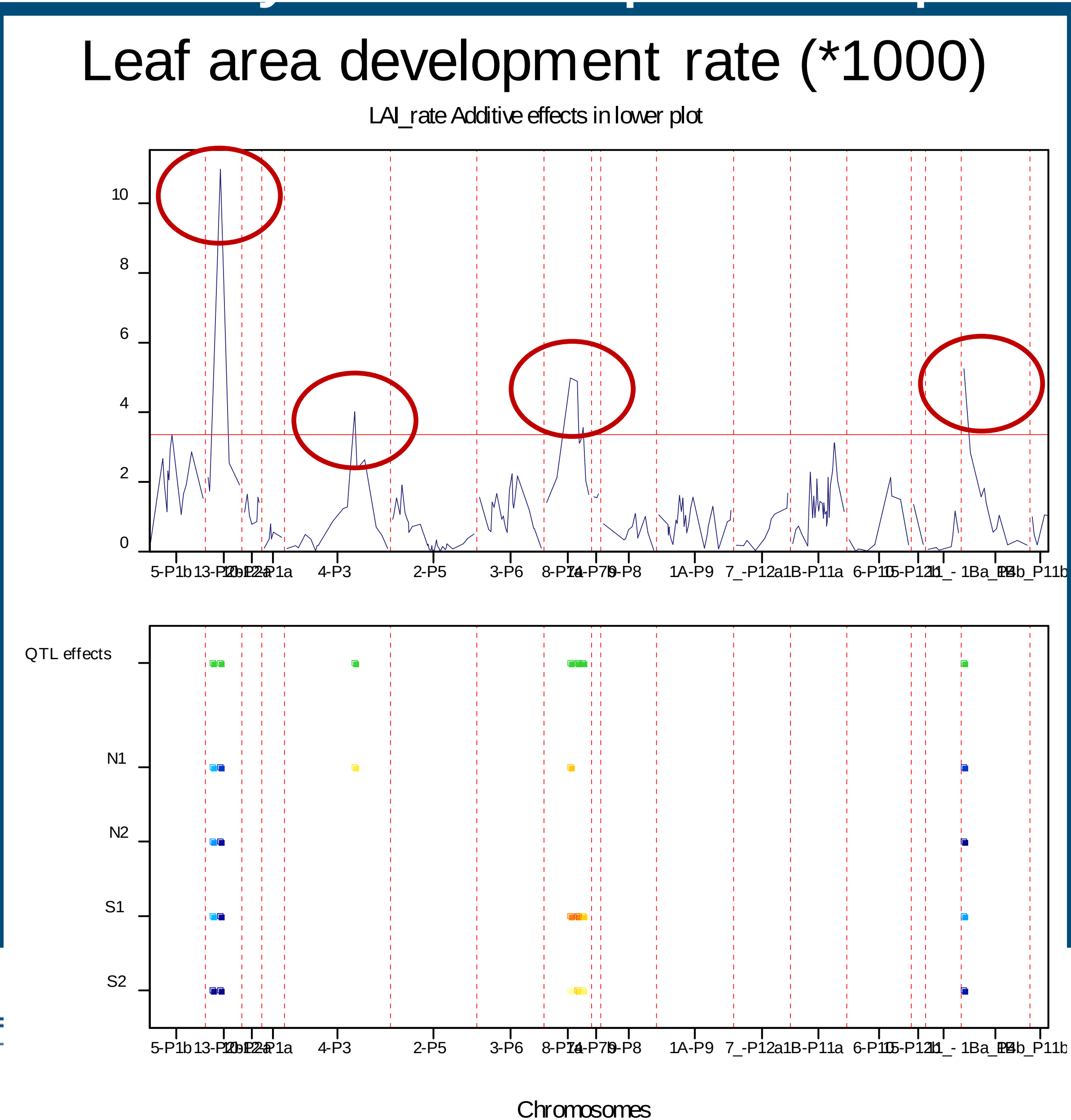
Correlations:
0.67-0.79

G2P models for yield and its components

- yield =
  - environment +
  - env. specific QTLs +
  - residual G and GxE +
  - error

- yield component =
  - environment +
  - QTL +
  - residual G +
  - error

QTL analysis on crop model parameters


Leaf area development rate (*1000)

Multi-environment QTL analysis

Effect of QTL's in different environments

Leaf area development rate

- Average fitted value 2.11
- Effect env. from -0.65 to 0.79

QTL	**S1**	**S2**	**N1**	**N2**
1 (P2b)	-0.16 (0.04)	-0.23 (0.03)	-0.21 (0.05)	-0.33 (0.07)
2 (P7a)	0.14 (0.04)	0.06 (0.03)	0.13 (0.05)	0.02 (0.07)
3 (P4)	-0.10 (0.04)	-0.13 (0.03)	-0.21 (0.05)	-0.35 (0.07)
4 (P3)	-0.05 (0.04)	0.02 (0.03)	0.08 (0.05)	-0.09 (0.07)

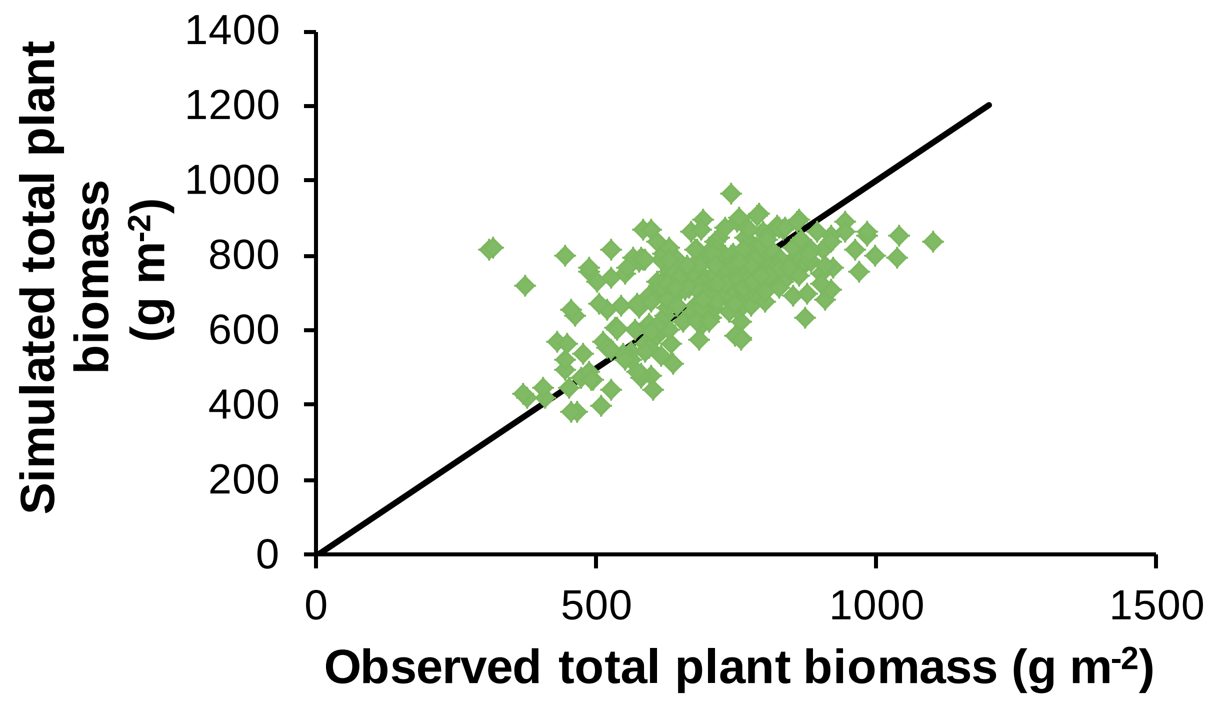
Variance explained by QTL's
N1 24% N2 24%
S1 23% S2 33%

Yield component traits

- Model parameters have high heritabilities and high genetic correlations

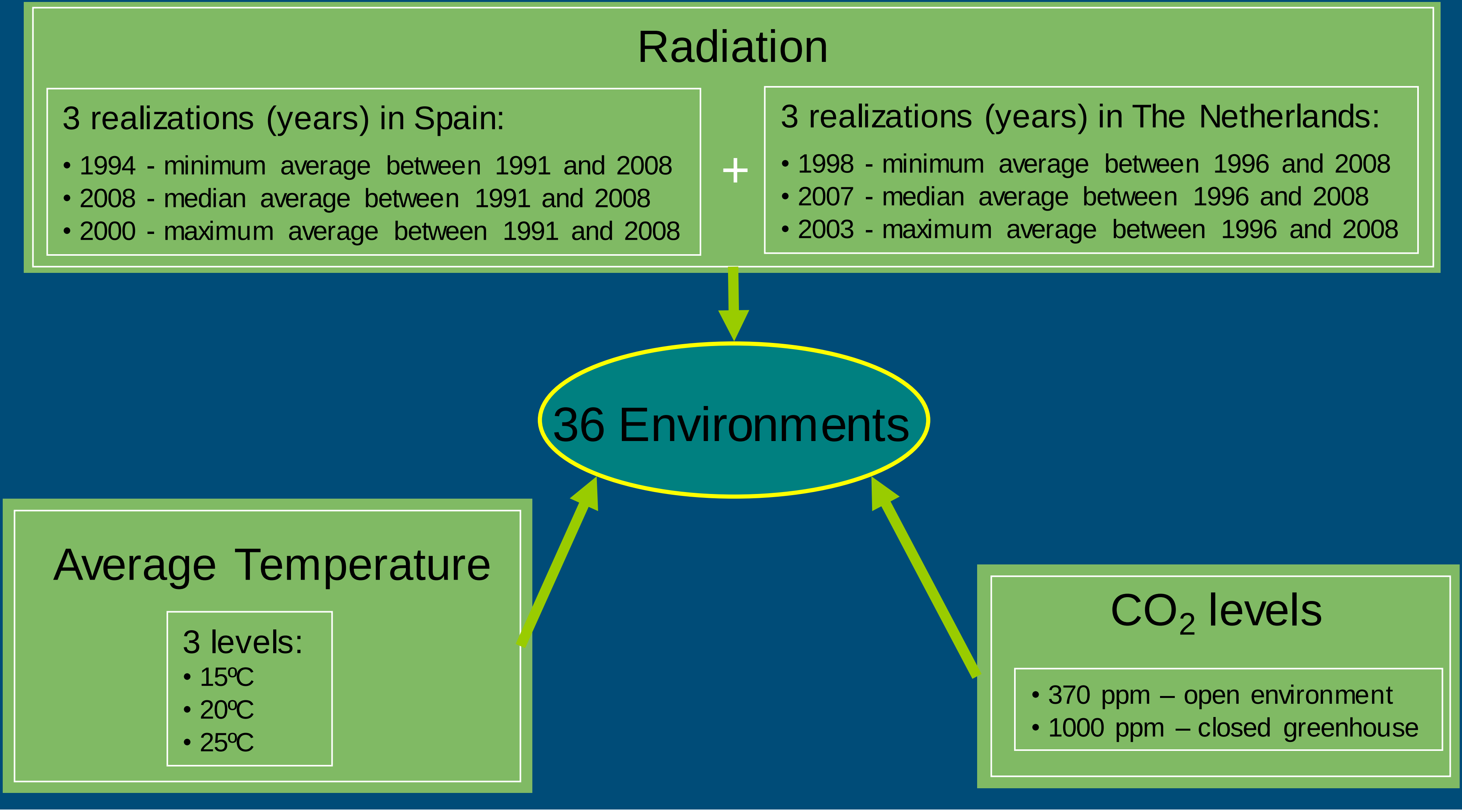
- QTL analysis
  - Analysis of leaf area dev. rate was successful
  - Other two parameters:
    - Not very high scores in QTL profiles
    - Effect of the QTL's is not very large

Predicting yield from crop growth model using ‘real’ QTLs



Predictions Spain 2 from QTLs identified in Spain 2
Rank correlation 0.54

Crop growth simulation: Scenario studies



Radiation

3 realizations (years) in Spain:

- 1994 - minimum average between 1991 and 2008
- 2008 - median average between 1991 and 2008
- 2000 - maximum average between 1991 and 2008

\+

3 realizations (years) in The Netherlands:

- 1998 - minimum average between 1996 and 2008
- 2007 - median average between 1996 and 2008
- 2003 - maximum average between 1996 and 2008

36 Environments

Average Temperature

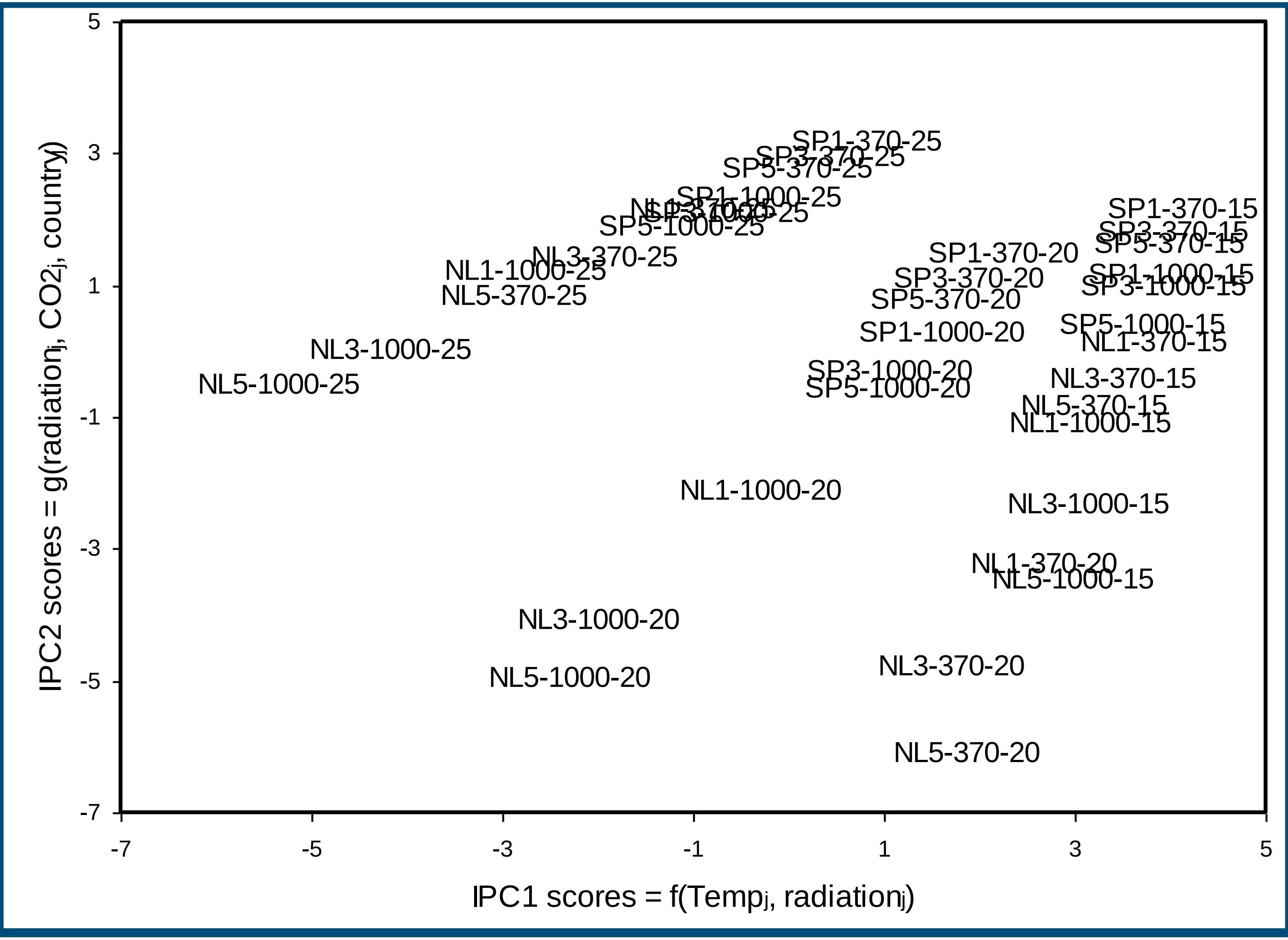
3 levels:

- 15ºC
- 20ºC
- 25ºC

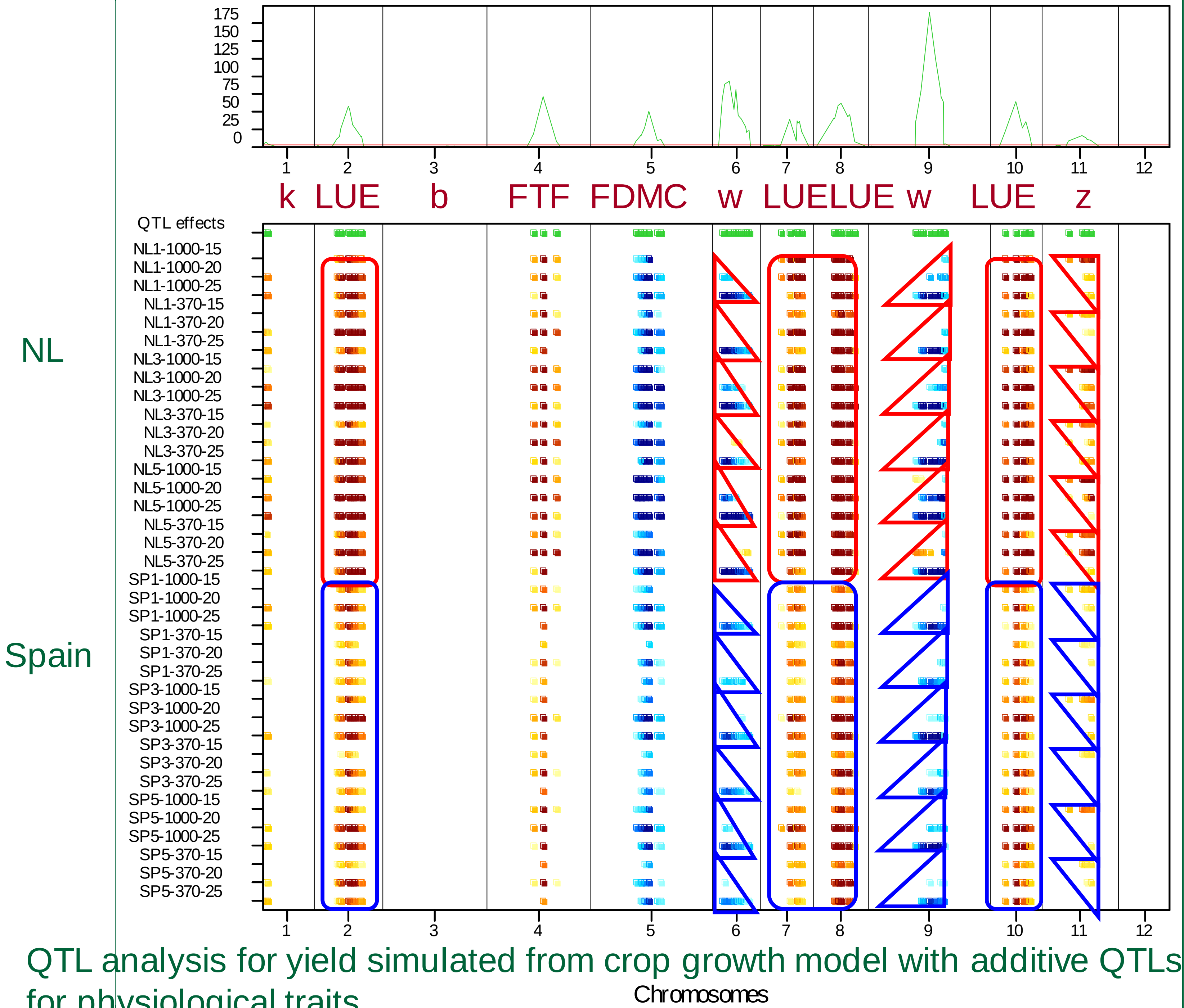
CO_2 levels

- 370 ppm – open environment
- 1000 ppm – closed greenhouse

GxE analysis on simulated yield (AMMI2 bip



G2P model realistic?



QTL analysis for yield simulated from crop growth model with additive QTLs for physiological traits

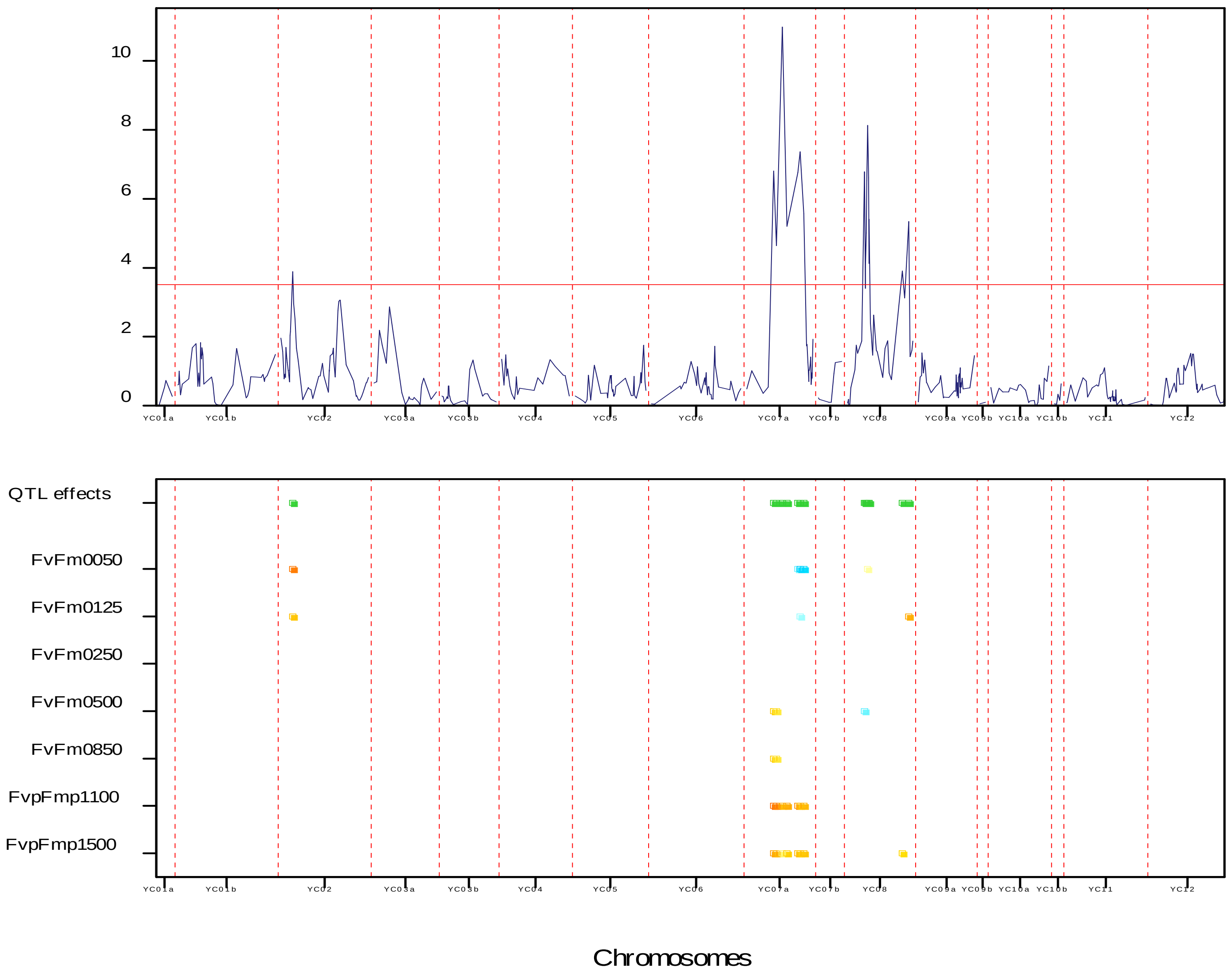
Results scenario study in pepper

- Crop growth model containing 7 genotype specific parameters (components) could produce GxE and QTLxE
- QTLs for yield appeared there were QTLs for components were simulated
- Patterns of QTLxE for yield could be understood given knowledge of location and effect of QTLs for yield components and environmental variables

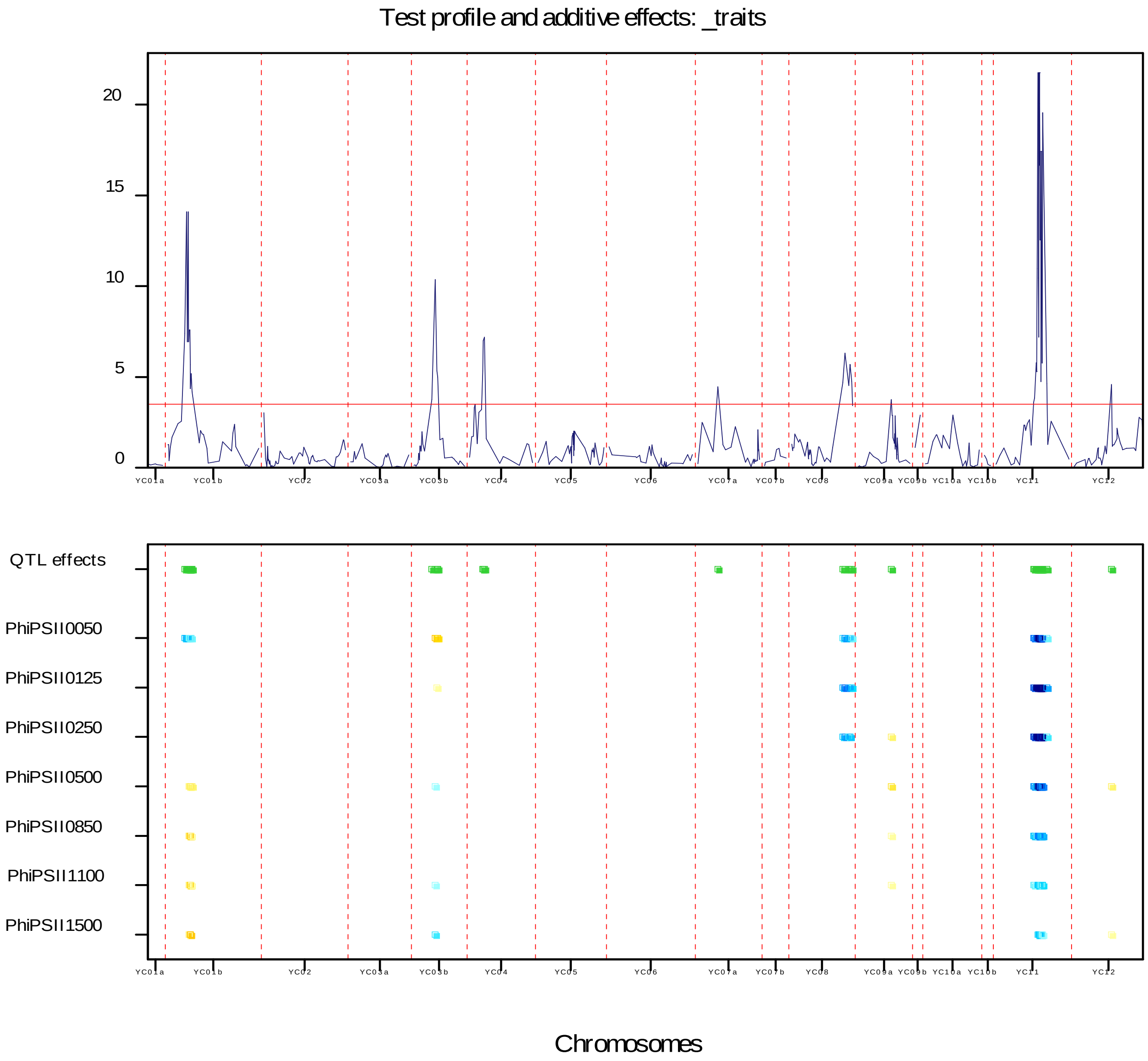
Fluorescence traits

- Can we find QTLs for fluorescence traits?
- Do these QTLs co-locate with physiologically interesting QTLs?

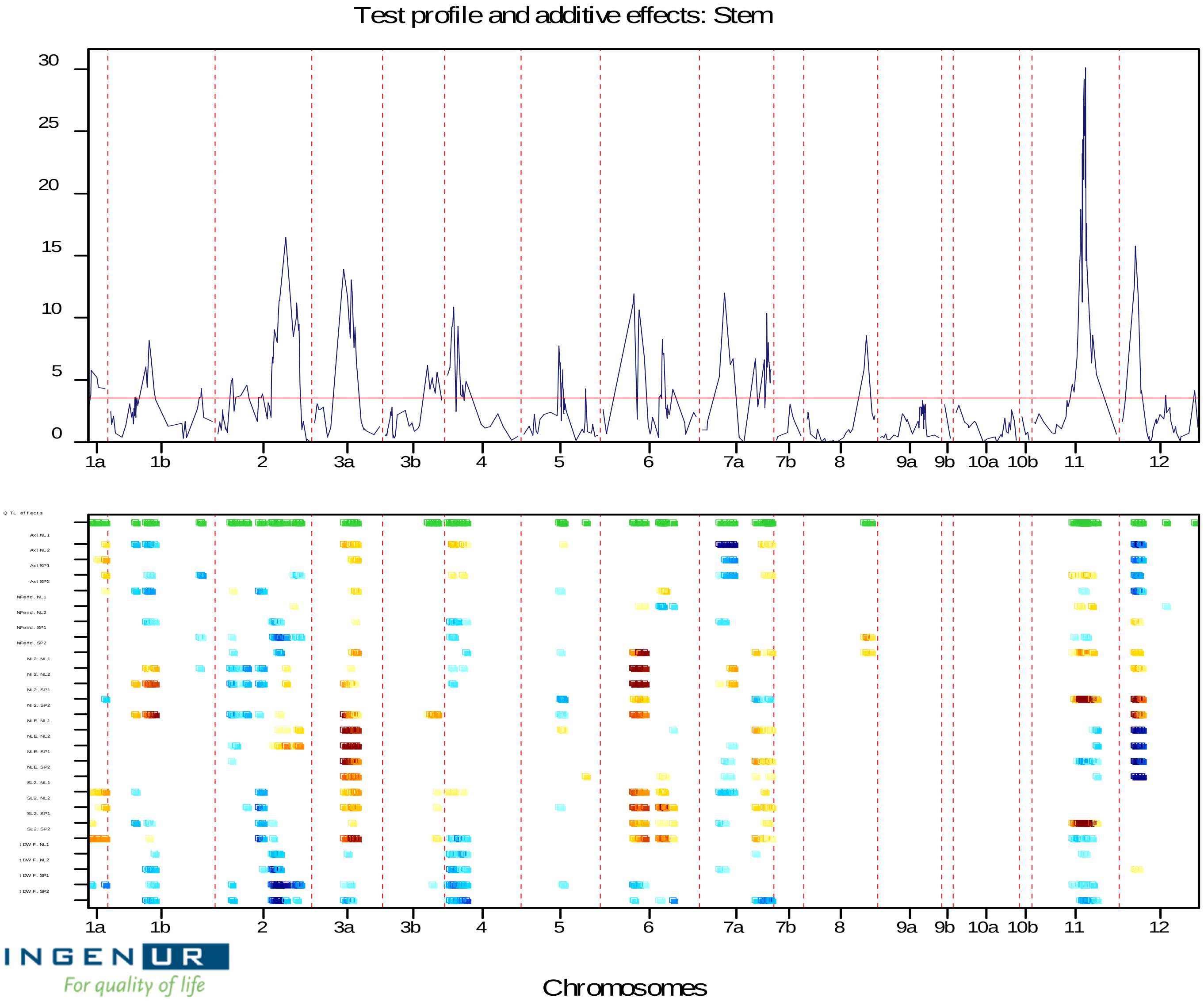
Mapping fluorescence traits I Fv’/Fm’


Test profile and additive effects: _traits

Mapping fluorescence traits PhiPSII



Mapping plant physiological traits:
Search for co-location with fluorescence traits


Test profile and additive effects: Stem

Chromosomes

Image analysis

Find summaries of spatial/spectral characteristics of images which:

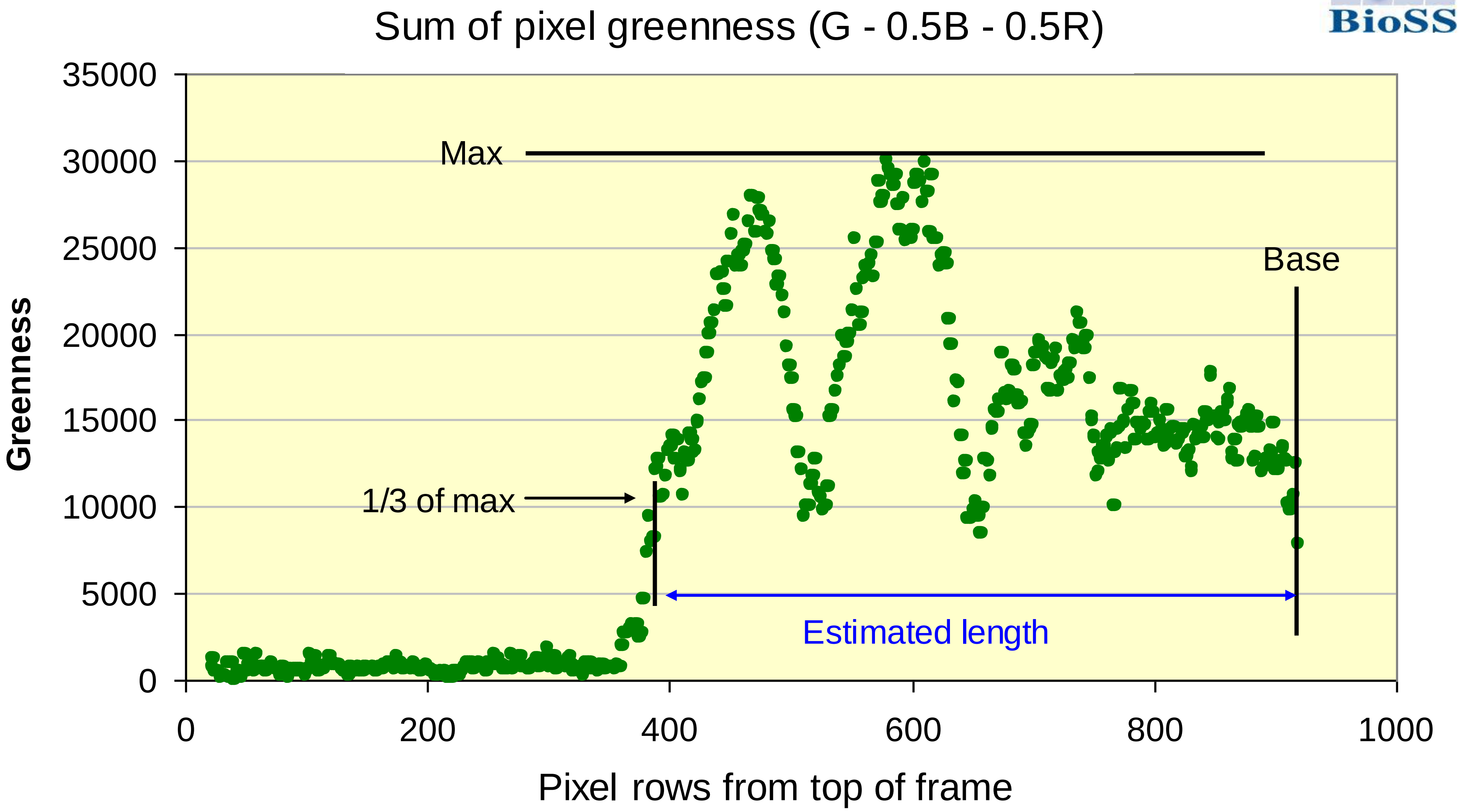
- **Correlate with physiological features (leaf area, fruit weight etc)**
- **Show consistent differences between genotypes: heritability**

Measuring plant characteristics

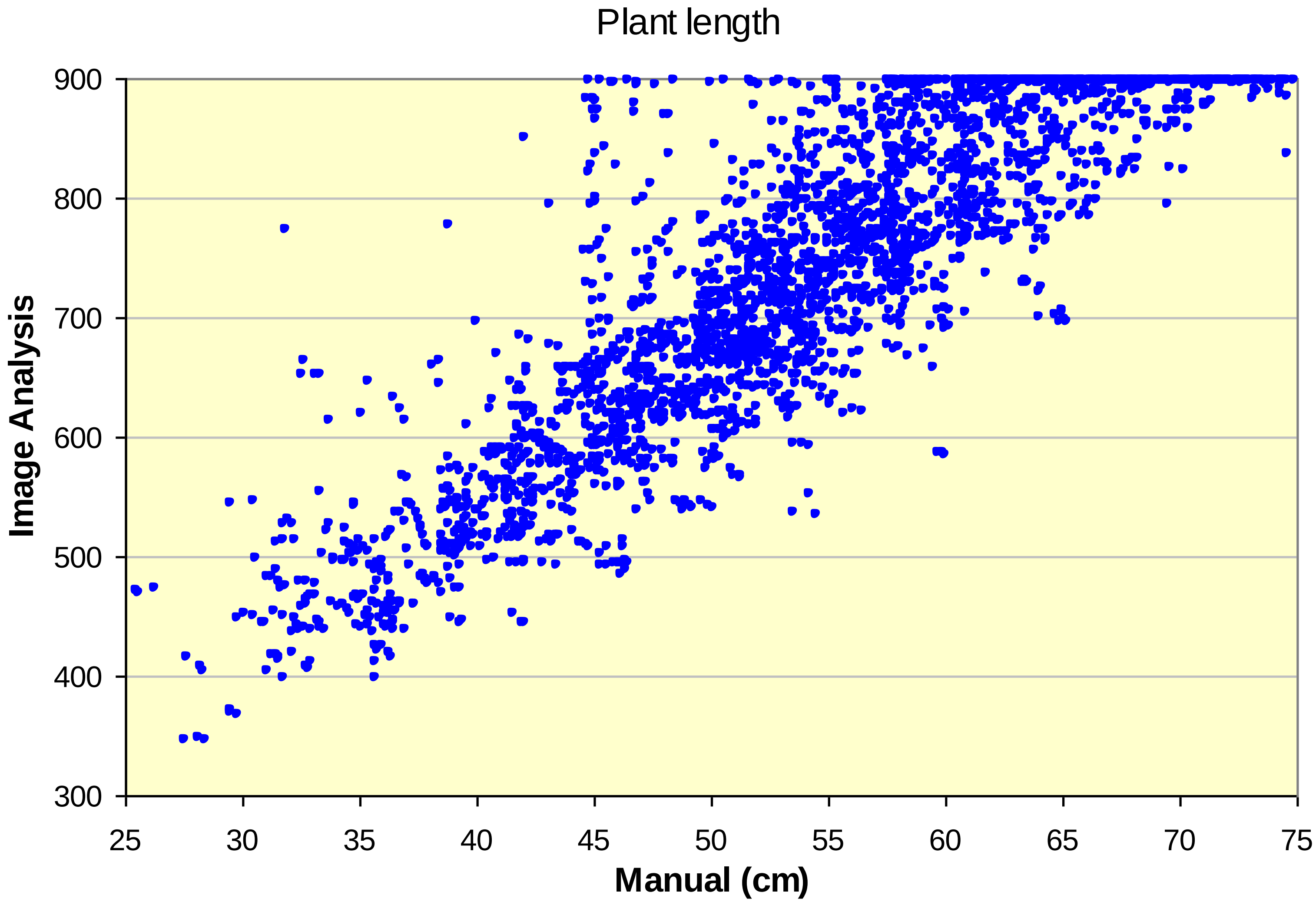
Barcode reading finds the position of each plant and reads its ID number.

This allows us to draw a frame around the plant.

Estimating height



Automatic vs manual



Correlation = 0.86

Manual length (sum of length 1 and 2) versus first estimate by image analysis

Consistent differences between genotypes

Manual length

```
Error: plant
           Df Sum Sq Mean Sq F value    Pr(>F)
genotyp   151 251646 1666.53  34.904 < 2.2e-16 ***
Residuals 738  35236   47.75

Error: Within
            Df     Sum Sq    Mean Sq F value Pr(>F)
Residuals 2077 1.5789e-26 7.6016e-30
```

Automatic length

```
Error: plant
            Df   Sum Sq Mean Sq F value    Pr(>F)
genotyp    153 61549300  402283  31.214 < 2.2e-16 ***
Residuals 1639 21123263   12888

Error: Within
            Df  Sum Sq Mean Sq F value Pr(>F)
Residuals 4273 6287244  1471.4
```

Discriminant analysis

- Data: red, green and blue spectrum (256 traits/intensities)
- Discriminant analysis constructs combinations of variables that best separate genotypes

Spectral histogram discriminant functions

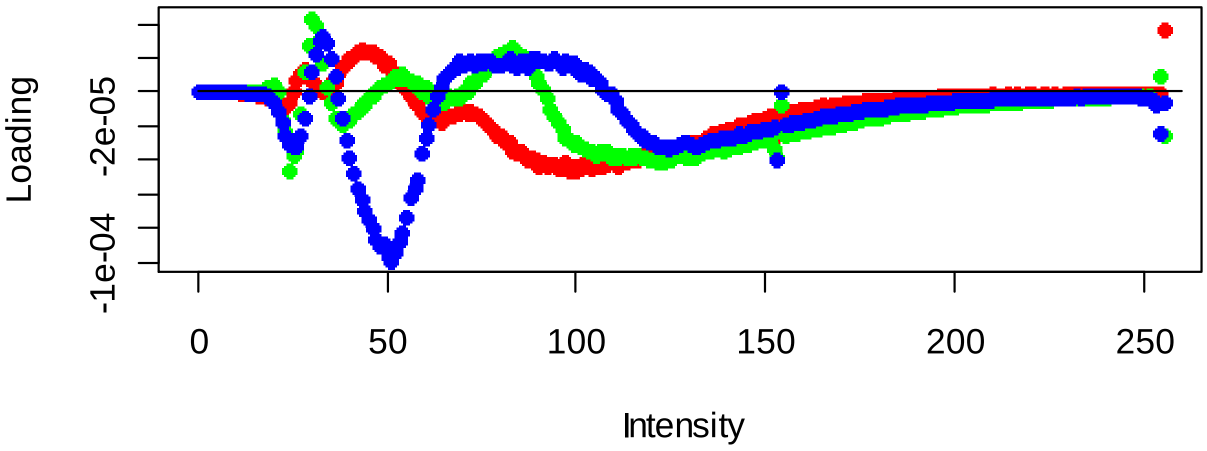
First discriminant function

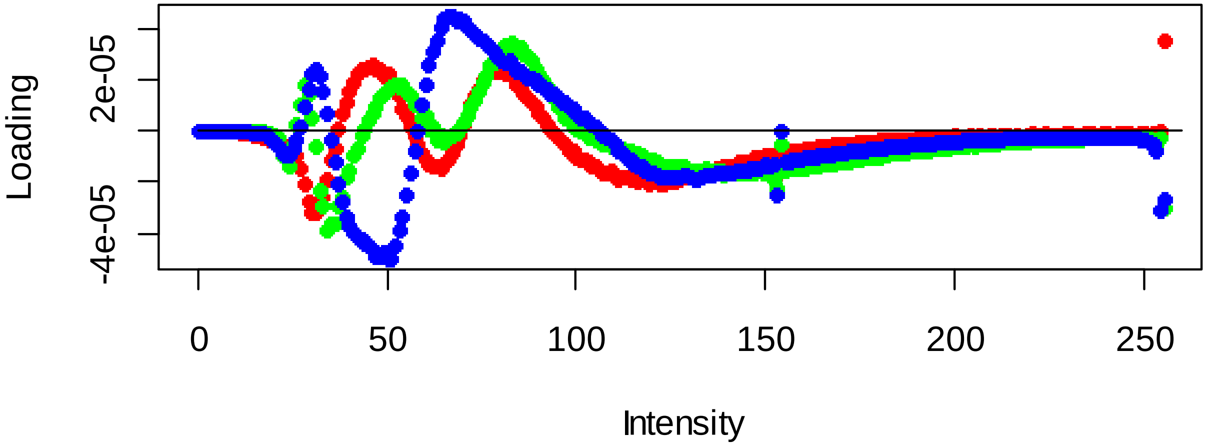
```
Error: plant
            Df  Sum Sq Mean Sq F value    Pr(>F)
genotype   153 16845.9 110.104  38.831 < 2.2e-16 ***
Residuals 1639  4647.3   2.835

Error: Within
            Df Sum Sq Mean Sq F value Pr(>F)
Residuals 4273 1264.7 0.29597
```

Second discriminant function

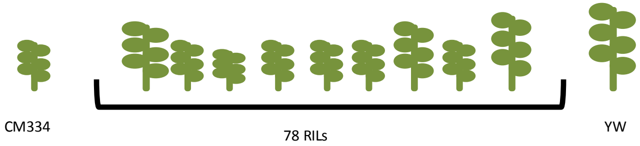
```
Error: plant
            Df Sum Sq Mean Sq F value    Pr(>F)
genotyp    153 7384.5  48.265  16.059 < 2.2e-16 ***
Residuals 1639 4925.9   3.005

Error: Within
            Df Sum Sq Mean Sq F value Pr(>F)
Residuals 4273 986.14 0.23078
```

Spectral histogram discriminant functions
What do they mean?

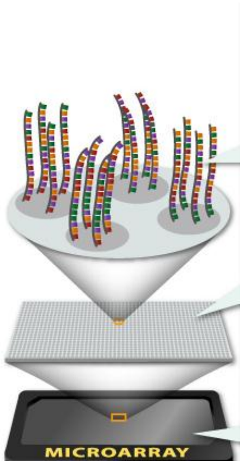

1st discriminant function


2nd discriminant function

Which genes can be identified behind QTLs?
eQTL mapping

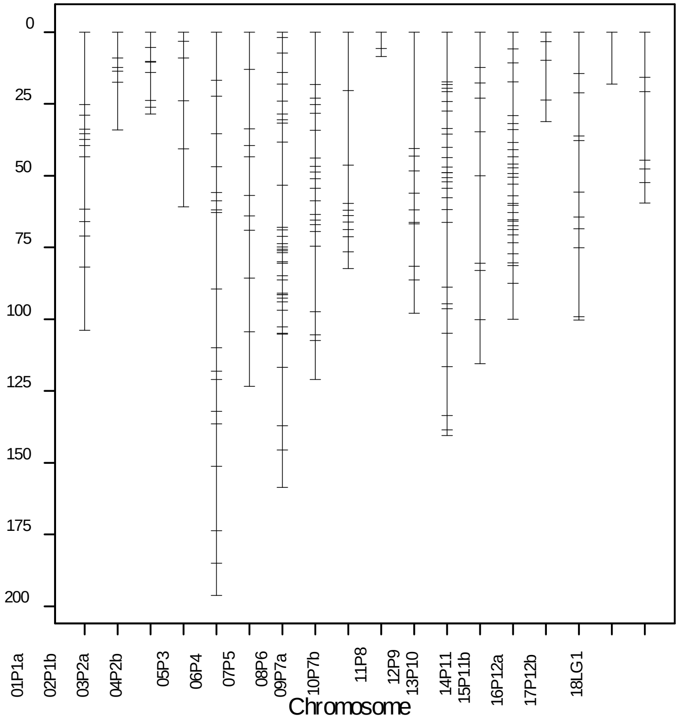


expression data



42779 probesets

+



Physiol. QTL

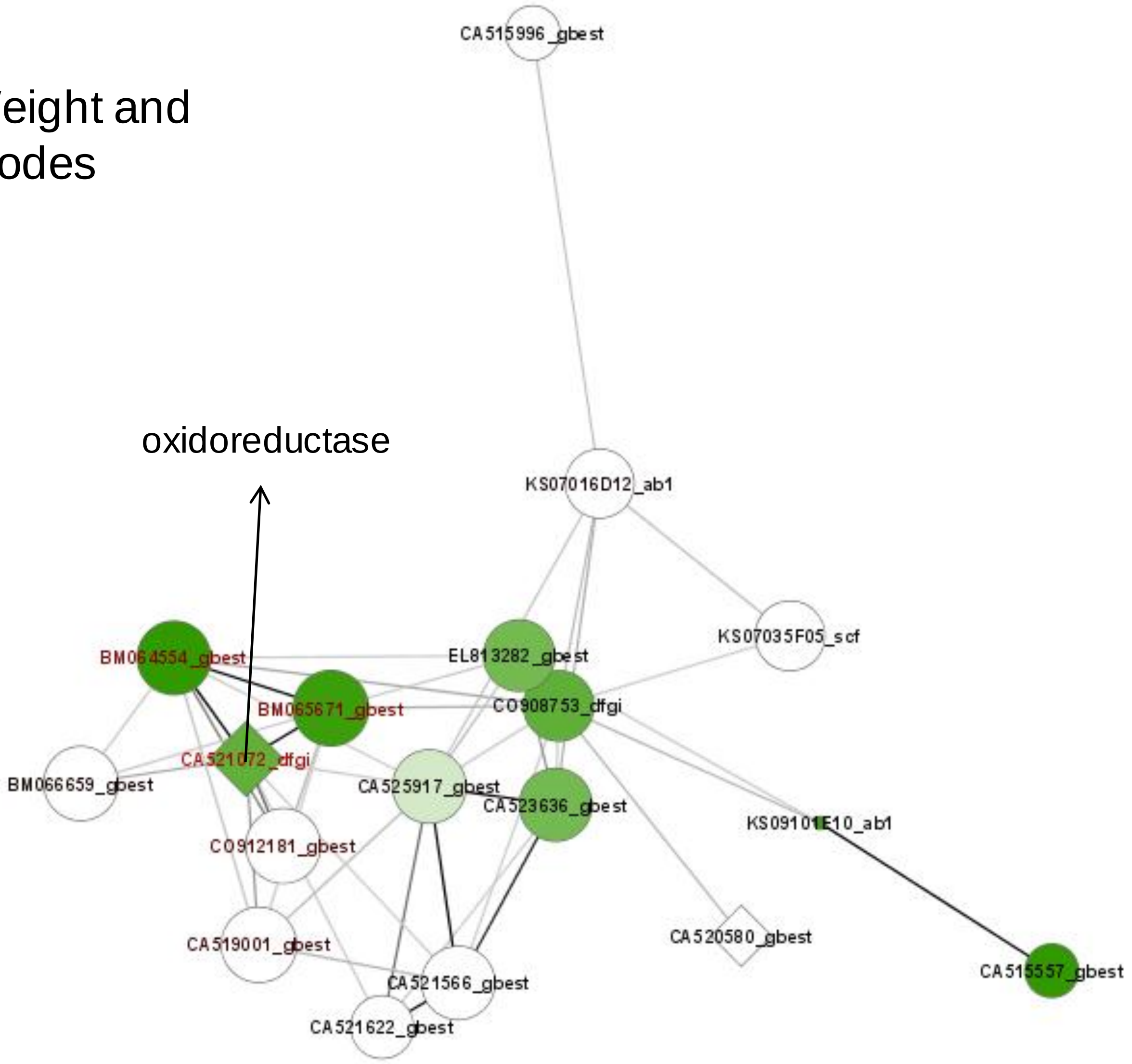
gene1

gene2

gene3

P3 (46 cM)

QTLs for Fruit Weight and
Number of Internodes


oxidoreductase

Final remarks

- Within context SPICY project a 3-5 day course/meeting will be organized on G2P modelling in general in spring 2011
- A special symposium on G2P modelling / crop growth modelling / genetics will be held 2 september 2012 in Hohenheim
- EU-DROPS project led by Francois Tardieu is another project with a strong G2P modelling component in it, comparing statistical and physiological approaches