



eQTL analysis using a two-colour microarray with a distant pair design

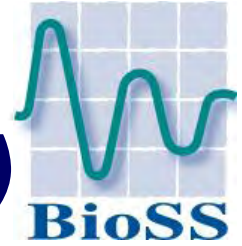
Christine Hackett, Xinwei Chen and Robbie Waugh

Outline

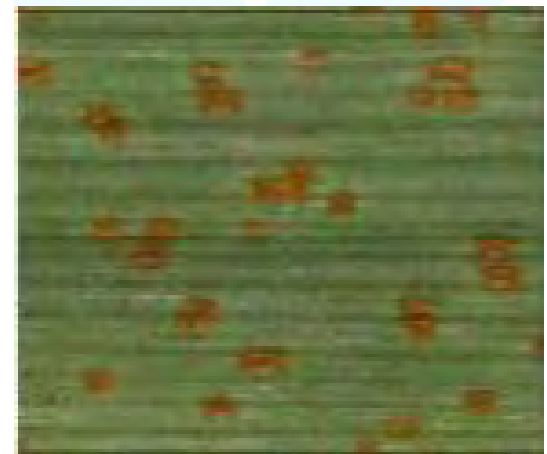


- Barley experiment
- QTL mapping and microarrays
- Designs for 2-colour arrays
- Analysis
- Next steps

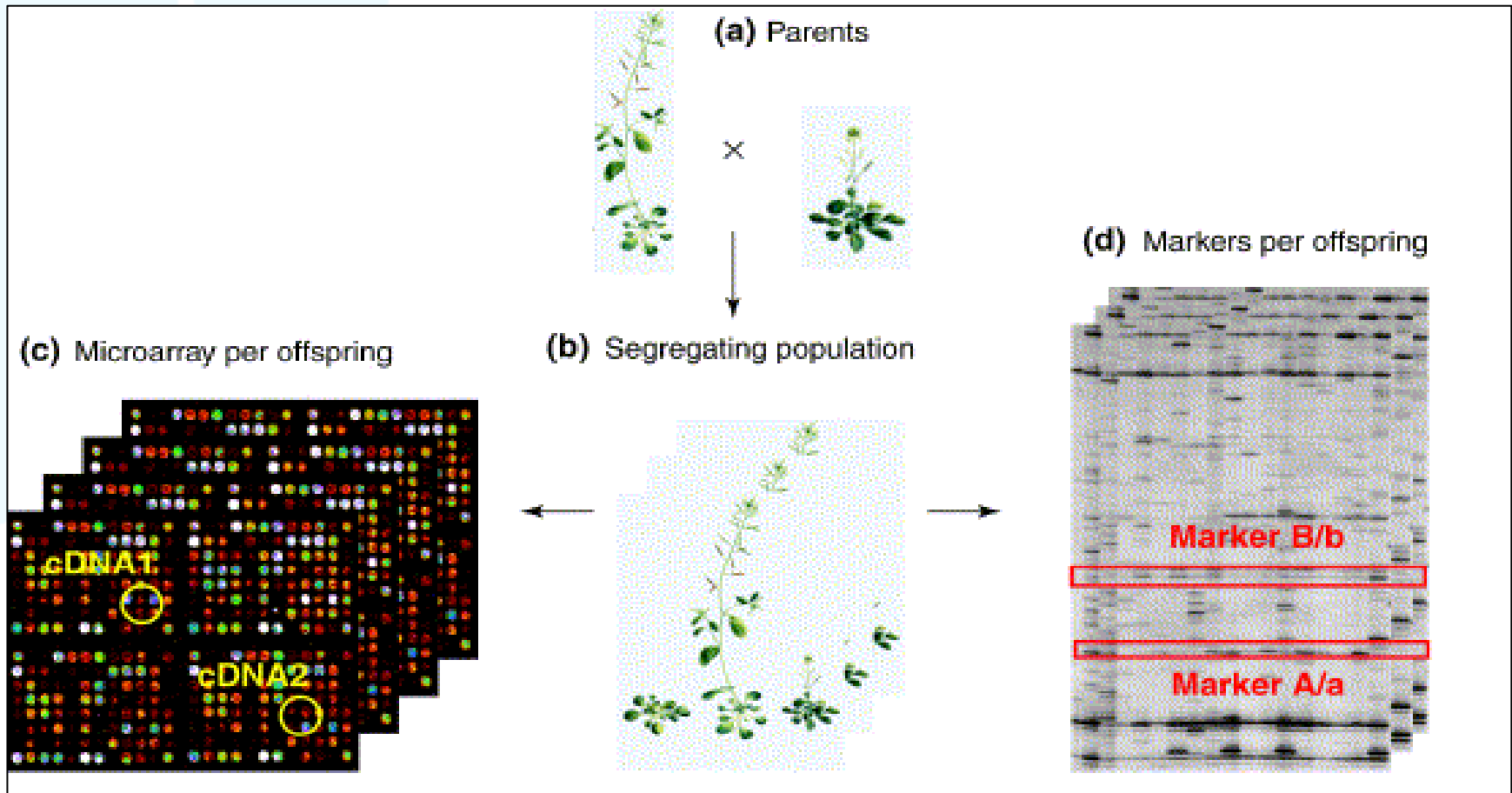
eQTL experiment: Resistance of barley to leaf rust (*Puccinia hordei*)



- Population: barley cross Steptoe x Morex and 144 DH progeny
- Detailed linkage map of SNPs available
- 9 day old seedlings inoculated with leaf rust
- Leaf segments taken after 10, 18, 24, 34, 42, 48 hours to assess rust development
- RNA extracted from segments taken after 18 hours for microarray analysis



Idea of eQTL mapping



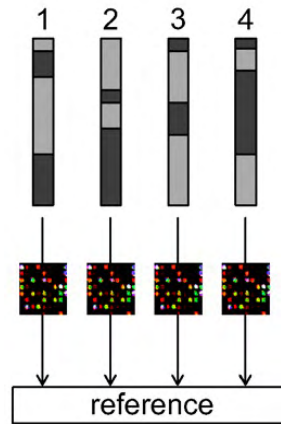
From Jansen & Nap 2001 Trends in Genetics

One- or two-colour array?



- One-colour array eg Affymetrix: use one array for each genotype
- Relate expression data to marker data using usual QTL model
- Simulation study by Fu & Jansen (2006) showed increased efficiency using two-colour arrays.
- Relate ratio of expression data to marker genotypes of the pair

Four experimental designs for two-colour arrays



Common reference design

Fu, J. et al. Genetics 2006;172:1993-1999

Copyright © 2007 by the Genetics Society of America

If a marker is linked to a QTL affecting gene expression we have the following possibilities:

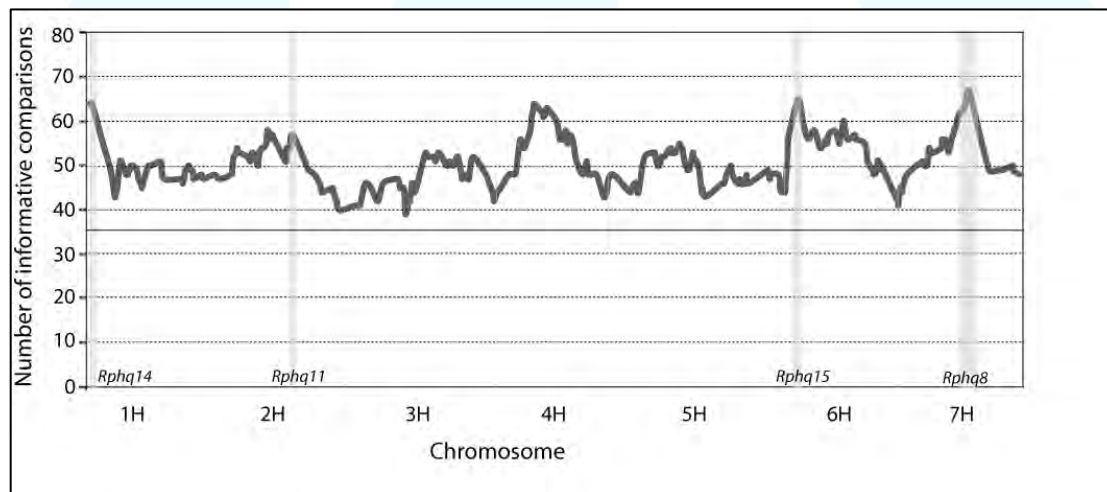
DH1(Cy5)	DH2(Cy3)	Expression
MM	MM	Similar expression
SS	SS	Similar expression
MM	SS	Different expression
SS	MM	Different expression

Recommendation: search for pairing of offspring to maximise number of marker differences

Distant pair design



- Marker data on this population: 466 SNPs
- Select evenly spaced framework (5-10 cM): 119 SNPs
- Search for pairing of 144 offspring to maximise number of differences, using simulated annealing
- Either weight all markers equally, or extra weight to regions of interest
 - Here: weighted regions around 4 possible QTLs for rust from previous studies
- Informative pairs increases from 50% to average of 69%



Array analysis



- Custom barley array
- 15208 genes represented
- 1564 of these genes have known location
- Fluorescence measurements made for each dye
- Calculate log-ratio of gene expression for each gene

QTL model for two-colour array



Proposed by Fu & Jansen (2006)

$$y_i = \alpha + \beta x_i + e_i$$

where

y_i is the gene's log ratio for microarray i

x_i = 1 if the pair (Cy5/Cy3) is SS/MM,

= -1 if the pair is MM/SS

= 0 if the pair is SS/SS or MM/MM

The regression coefficient β shows the effect of the allele difference, while the intercept α should be close to zero unless there is dye bias.

Significance level? 15208 genes regressed on 466 marker locations gives multiple testing issues.

Too large for permutation tests

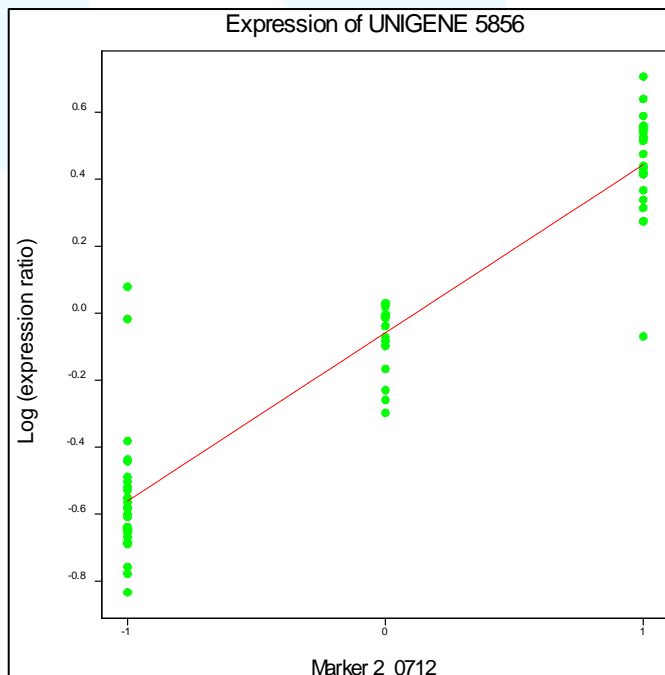
Correlation between markers complicates FDR approach

Scan for $p < 0.001$

Results

- Significant ($p < .001$) QTLs were detected for 9557 out of 15208 genes
- For many genes a single QTL explained a high proportion of variance:
- Other genes had up to six significant QTLs

% variance explained	# genes
90-100%	145
80-89%	595
70-79%	591
60-69%	534
50-59%	548
40-49%	597
Total	3010



Cis- and trans- acting genes

Does eQTL location coincide with gene location?

- eQTL at gene location indicates a *cis*-acting polymorphism in the regulatory region of that gene
- eQTL at distant location indicates a *trans*-acting factor, which may control expression of many genes

• Array has 1564 genes with known location.

• 1065 of these have a significant QTL ($p < 0.001$)

QTL site	Count	%	mean %R ²
< 3cM	613	57	64%
4-20cM	164	15	48%
> 20cM	87	8	27%
unlinked	201	19	22%

Most QTLs are plausibly cis-acting

Trans-acting QTL generally less significant

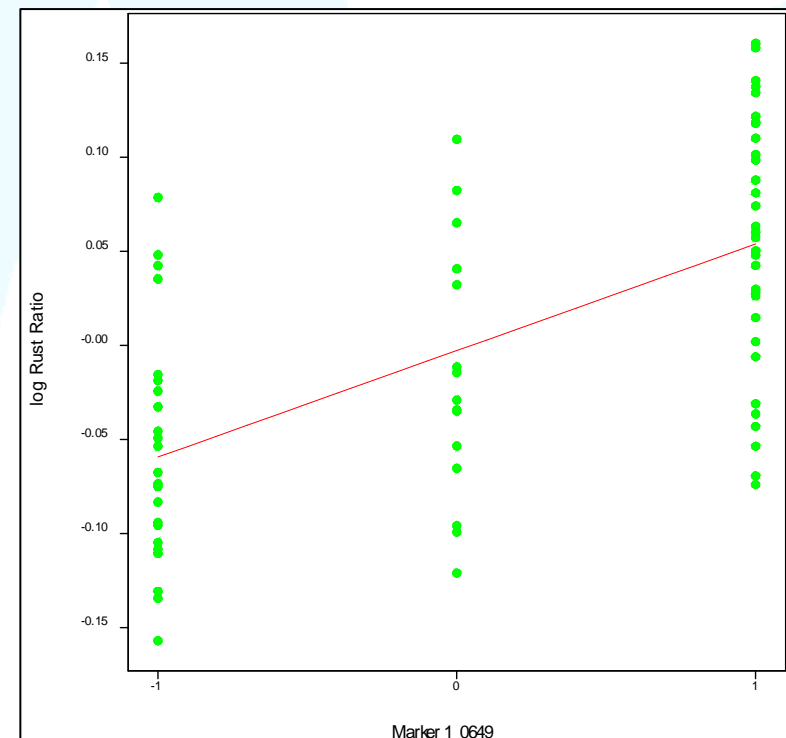
Analysis of rust scores



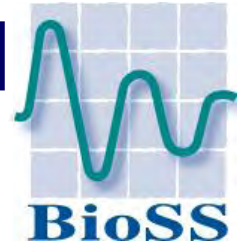
- Rust scored as the relative latency period
- Usual QTL mapping identified 4 QTLs on 1H, 2H, 6H, 7H
- Map $\log(\text{rust-ratio})$ using same model as for expression data
- Most significant QTL on 2H, same location, $R^2 = 35\%$

Further QTLs detected on

- 6H
- 4H
- 1H
- 7H



Correlation between rust score and gene expression



- Examined correlation between ratio of rust scores and ratio of gene expression
- 117 genes correlated with $p < 0.001$
- Top 23 of these have QTLs in a small region of chromosome 2H
 - Good candidate genes for rust resistance locus
 - Also promising candidate genes for other resistance loci

Future work



- *QTL hotspots?*
- Are there regions of the genome that are trans-acting QTLs for an unusually high number of gene expressions?
- How do we test this, given complex correlations between gene expression?
- Permutation approach proposed in Breitling et al. (2008)

Collaborators



- Robbie Waugh
- Xinwei Chen
- Jim McNicol
- Peter Hedley
- Arnis Druka

This study was supported by BioExploit in EU5774 and RERAD