

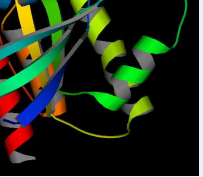
# **Multi-treatment QTL analysis using the complete linkage map**

**Ari Verbyla**

**School of Agriculture, Food and Wine,  
The University of Adelaide, and  
Statistical Bioinformatics — Agribusiness,  
CMIS, CSIRO, Adelaide**

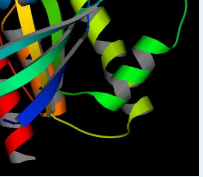
**[Ari.Verbyla@adelaide.edu.au](mailto:Ari.Verbyla@adelaide.edu.au) or [Ari.Verbyla@csiro.au](mailto:Ari.Verbyla@csiro.au)**

**Joint work with Brian Cullis, Christine Hackett, Adrian  
Newton and Bill Thomas**



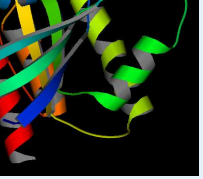
# Acknowledgements

- Grains Research and Development Corporation
- Ali Koshkoih, Wayne Pitchford
- Robin Thompson
- Alison and Emily Smith, Marryann O'Donnell



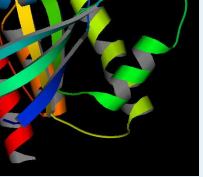
# Outline

- Scientific problem
- Linear mixed model
- Preliminary analysis
- Multi-treatment model
- Conclusions



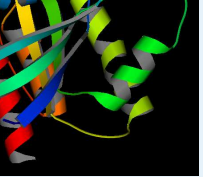
# Scientific Problem

- Yield performance of barley
- Plots infected with powdery mildew
- Treatments applied to plots
- How is the expression of quantitative trait loci influenced by the treatments?
- **Genotype by treatment interaction**



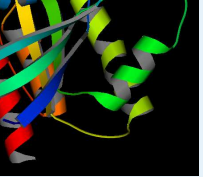
# Derkado × B83.12/21/5 Population

- Scottish Crop Research Institute, Dundee
- Barley Mapping Population
- Doubled Haploid Lines: homozygous
- 156 DH lines used in the trial
- Barley has 7 chromosomes: diploid

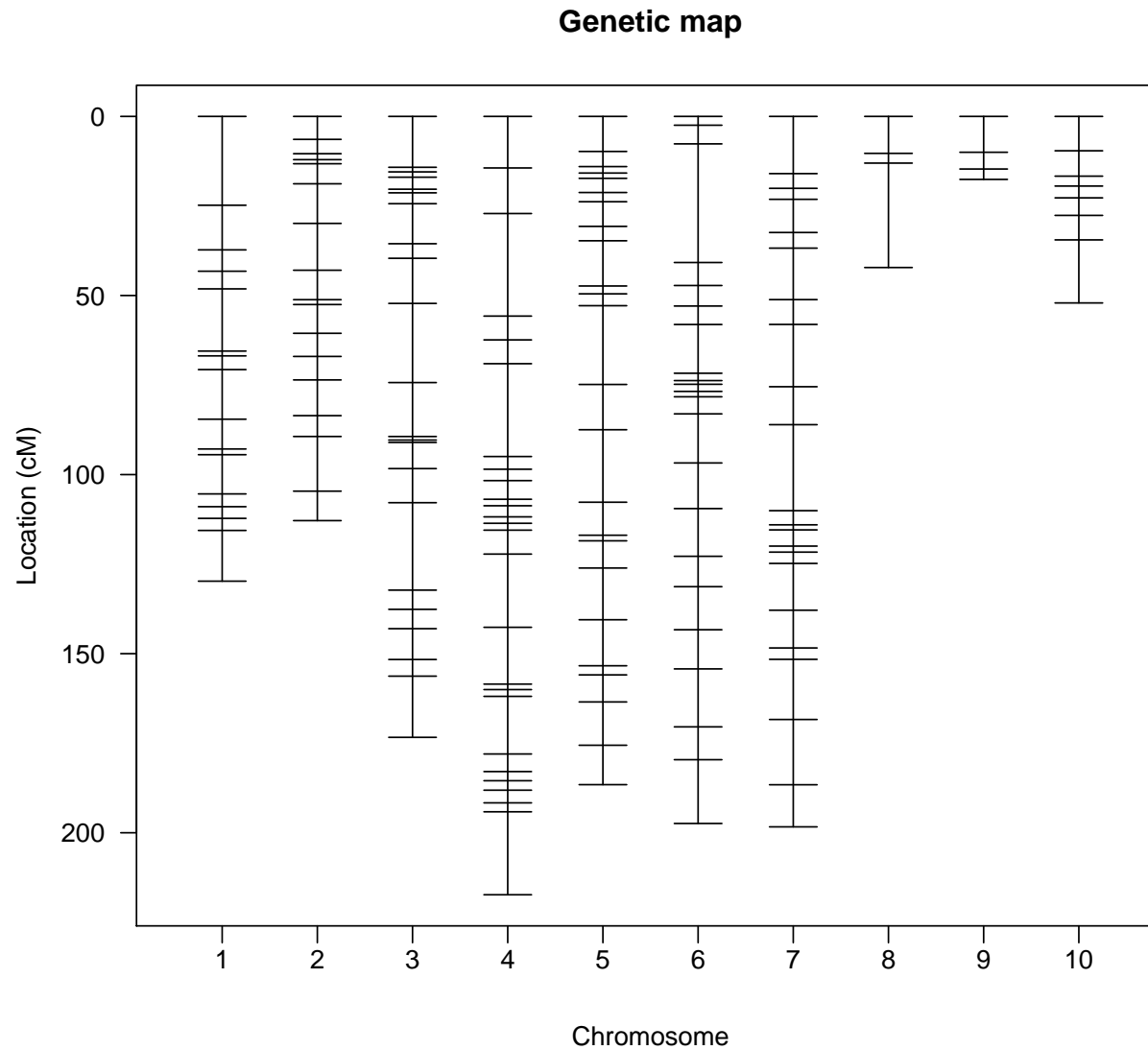


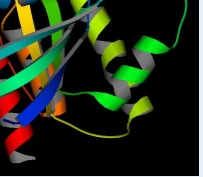
# Derkado × B83.12/21/5 Linkage Map

- 165 markers scored
- 10 linkage groups found (JoinMap)
- 16, 17, 22, 26, 24, 22, 22, 4, 4, 8 markers in the linkage groups
- Average spacing between markers: 9 cM
- 1336 missing marker scores: replaced using the method of Martinez and Curnow (1994)



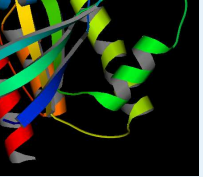
# Derkado × B83.12/21/5 Linkage Map





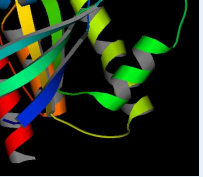
# Field trial Design

- Split-plot design
- 2 blocks of 4 wholeplots
- Wholeplots: 10 beds by 16 rows (160 plots)
- Whole plot treatments: control (nil), Fungicide, Azostrobin, Algal Gold (assigned at random, restricted randomisation)
- Within wholeplot “treatments”: 1 replicate of 156 DH lines, 2 replicates of the parents (assigned at random)
- Full layout: 80 beds by 16 rows



# Oops!

- Azostrobin treatment in block 2
- Mixup — not clear of the actual treatment applied
- Omit the wholeplot
- Limited information on Azostrobin treatment



# Linear mixed model

- Statistical Model

$$y = X\tau + Zu + Z_g g + e$$

- $u \sim N(\mathbf{0}, \sigma^2 \mathbf{G}_u)$
- $e \sim N(\mathbf{0}, \sigma^2 \mathbf{R})$
- Focus on  $g$



# Genetic effects

- Standard quantitative genetics assumption

$$\mathbf{g} \sim N(\mathbf{0}, \sigma_g^2 \mathbf{I}_{n_g})$$

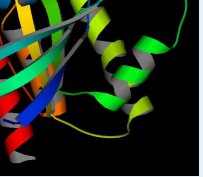
- Multi-treatment model:  $G$  is an  $n_g \times t$  matrix of genetic effects

- $\mathbf{g} = \text{vec}(G)$

- Model:

$$\mathbf{g} \sim N\{\mathbf{0}, \sigma^2(\Gamma \otimes \mathbf{I}_{n_g})\}$$

- What form should  $\Gamma$  take?



# Preliminary analysis

- Symbolic model:  
yield = Type.Treat + **Block** + **Block.Wholeplot** + **diag(Treat).Variety** + e
- Terms in bold: random effects
- Split-plot design BUT
- Treat.Variety is random
- Type distinguishes DH lines from parents (3 level factor)
- $\Gamma$  is a diagonal structure
- 5 outliers detected, removed
- **Next step: correlate the treatments across Varieties: polygenic effect**

# Polygenic Covariance model

- Symbolic model:

$$\text{yield} = \text{Type.Treat} + \text{linrow} + \mathbf{Block} + \mathbf{Block.Wholeplot} + \mathbf{xfa(Treat).Variety} + \mathbf{e}$$

- xfa: Factor analytic model:

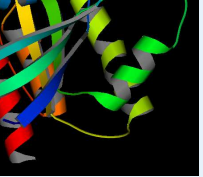
$$\mathbf{g} = (\mathbf{\Lambda} \otimes \mathbf{I}_{n_g}) \mathbf{f} + \boldsymbol{\xi}$$

which leads to

$$\mathbf{\Gamma} = \mathbf{\Lambda}\mathbf{\Lambda}^T + \mathbf{\Psi}$$

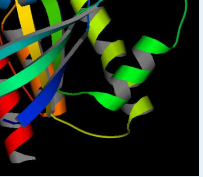
- Why not an unstructured covariance matrix?
- Implementation: **ASReml** and **R**
- Estimated polygenic covariance/correlation matrix:

$$\begin{bmatrix} 0.0528 & 0.7285 & 0.5660 & 0.8371 \\ 0.0329 & 0.0386 & 0.4675 & 0.6914 \\ 0.0337 & 0.0238 & 0.0672 & 0.5372 \\ 0.0460 & 0.0325 & 0.0333 & 0.0572 \end{bmatrix}$$



# Other parameters

Term	Estimate
linrow	.026 (.0029)
<b>Block</b>	.0013
<b>Block.Wholeplot</b>	0
Residual Variance	.049
Residual Row Correlation	.24
Residual Bed Correlation	.44



# Multi-treatment model

- Genetic model (following on from Simon Diffey's talk)

$$g_{ij} = \sum_{k=1}^c \sum_{l=1}^{m_k-1} q_{ijkl} a_{jkl} + p_{ij}$$

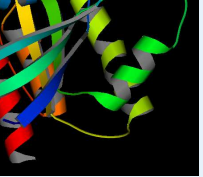
- Regression approach plus uniform distribution for the location of putative QTL for each interval
- Matrix form for genetic effects:

$$G = M \Lambda_E A + P = M_E A + P$$

- Vector form:

$$g = (I_t \otimes M_E) a + p$$

- QTL may be coincident (same interval) or non-coincident across treatments
- Coincident: pleiotropic effect, varying size of effect



# Working model

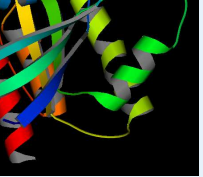
- Use a factor analytic model for intervals

$$\mathbf{a} = (\boldsymbol{\lambda}_a \otimes \mathbf{I}_{n_g}) \mathbf{f}_a + \boldsymbol{\xi}_a$$

- Full genetic model:

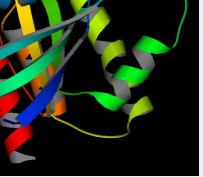
$$\mathbf{g} = (\boldsymbol{\lambda}_a \otimes \mathbf{M}_E) \mathbf{f}_a + (\mathbf{I}_t \otimes \mathbf{M}_E) \boldsymbol{\xi}_a + (\boldsymbol{\Lambda} \otimes \mathbf{I}_{n_g}) \mathbf{f} + \boldsymbol{\xi}$$

- Fitted as two terms: Reduced Rank (RR) and DIAG
- Fitted as: pleiotropic and non-pleiotropic



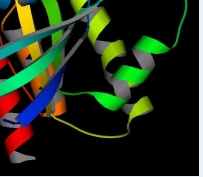
# Algorithm

1. Include a factor analytic or reduced rank (RR) model for pleiotropic effects (random effects)
2. Include a separate term for each treatment contrast (with the control)
3. Examine the significance of the RR term.
4. If significant
  - (a) Use the outlier approach to determine the most likely chromosome for the pleiotropic QTL
  - (b) Use the outlier approach to determine the most likely interval on the selected chromosome
  - (c) Move the interval to the fixed effects, allowing for differential expression across treatments.
  - (d) Fit the new model and repeat until the RR term is no longer significant.



# Algorithm continued

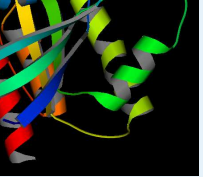
5. If not significant, remove the RR term and refit.
6. Check the significance of the non-pleiotropic terms (one for each treatment level contrast)
7. If significant non-pleiotropic variation exists, for each such term go through step 4.
8. When both the RR term and the individual non-pleiotropic terms are non-significant, all possible QTLs have been determined.



# Fitted models

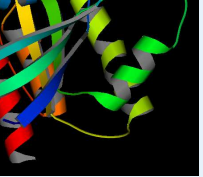
Model	Treat.Variety	Treat.Interval	QTL	log-likelihood	$-2 \log \Lambda$	df
1	DIAG	—	—	829.75		
2	US	—	—	914.06	170.62	6
3	FA	—	—	913.15	1.82	2
4	FA	RR + DIAG <sup>†</sup>	—	948.23	70.16	3
5	FA	RR	—	948.23	0	3
6	FA	RR	Treat.X1	957.20		
7	FA	—	Treat.X1	953.01	8.38	4
8	FA	RR.X1	Treat.X1	958.83	11.64	4
9	FA	—	Treat.X1	950.51	16.64	4
10	FA	RR.X1	Treat.(X1 + X1.X2)	949.24		
11	FA	—	Treat.(X1 + X1.X2)	943.93	10.58	4
12	FA	RR.X1	Treat.)X1 + X1.X2 + X1.X3)	941.91		
13	FA	—	Treat.(X1 + X1.X2 + X1.X3)	940.82	2.18	4

<sup>†</sup> For treatments Fungicide, Azostrobin and Algal Gold



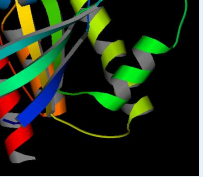
# QTL locations

Indicator	Linkage Group	left.marker	right.marker	left.distance	right.distance
X1	3	sdw1-r	abc08208	90.367	91.021
X2	5	ari-eGP	Bmag357	23.818	30.700
X3	7	P25M42c	P16M47f	51.148	58.082



# QTL

QTL	Treatment	Size	LOD
X1	none	-0.132	10.94
X1	Fungicide	-0.048	1.70
X1	Azostrobin	-0.005	0.01
X1	Alga Gold	-0.176	25.19
X1.X2	none	0.057	1.83
X1.X2	Fungicide	0.089	5.03
X1.X2	Azostrobin	0.109	3.66
X1.X2	Alga Gold	0.061	2.62
X1.X3	none	0.047	1.22
X1.X3	Fungicide	0.095	5.86
X1.X3	Azostrobin	0.052	0.80
X1.X3	Alga Gold	0.055	2.10



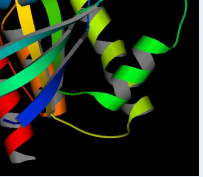
# Polygenic covariance matrices

Pre-QTL

$$\begin{bmatrix} 0.0528 & 0.7285 & 0.5660 & 0.8371 \\ 0.0329 & 0.0386 & 0.4675 & 0.6914 \\ 0.0337 & 0.0238 & 0.0672 & 0.5372 \\ 0.0460 & 0.0325 & 0.0333 & 0.0572 \end{bmatrix}$$

Post-QTL

$$\begin{bmatrix} 0.0342 & 0.6930 & 0.5966 & 0.7958 \\ 0.0212 & 0.0275 & 0.5389 & 0.7189 \\ 0.0267 & 0.0216 & 0.0589 & 0.6189 \\ 0.0230 & 0.0186 & 0.0235 & 0.0245 \end{bmatrix}$$



# Conclusions

- Method extends to multi-trait and multi-environment analysis.
- Pleiotropic effects: the RR term has been the only one required in several applications (but the DIAG term should be fitted routinely).
- Epistatic interactions found in the DH population.