Deciphering the meaning of variancecontrolling quantitative trait locus (vQTL)



Type

Trait

Trait + GC

Residual

DGLM

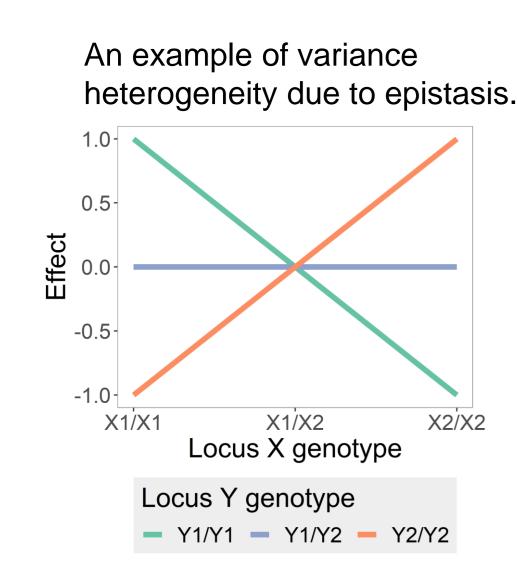
CJ Yang, Wayne Powell, Ian Mackay

Scotland's Rural College (SRUC) West Mains Road Edinburgh, EH9 3JG, UK Email: cyang@sruc.ac.uk Website: cjyang-work.github.io

Feel free to snap a photo, retweet and connect on Twitter-land! @hataraku_cj @SRUCPrincipal @lanJMackay

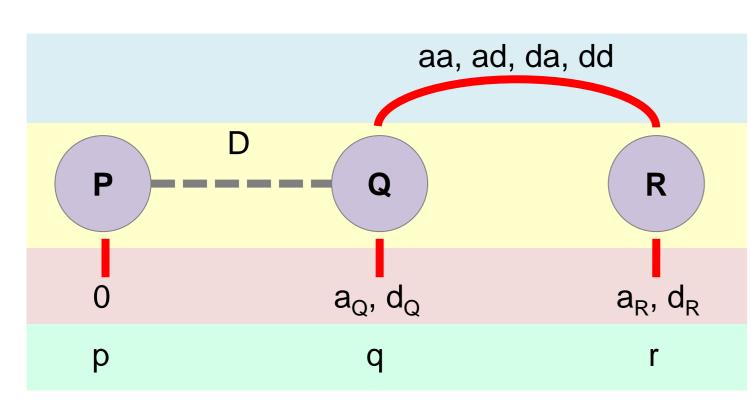
vQTL example

Variance-controlling QTL (vQTL) leads to phenotypic variance heterogeneity.



Locus X	Locus Y	Count	AxA	X mean	X var.
	Y1/Y1	5	1		
X1/X1	Y1/Y2	10	0	0	0.5
	Y2/Y2	5	-1		
	Y1/Y1	10	0		
X1/X2	Y1/Y2	20	0	0	0
	Y2/Y2	10	0		
	Y1/Y1	5	-1		
X2/X2	Y1/Y2	10	0	0	0.5
	Y2/Y2	5	1		

Model for vQTL under epistasis

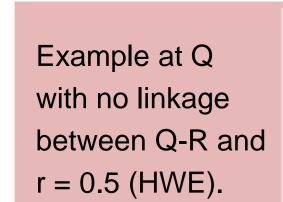


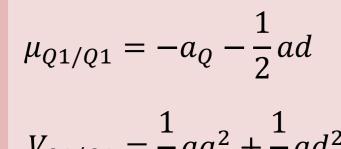
Pairwise interaction effect

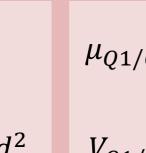
3 loci (P & Q are linked)

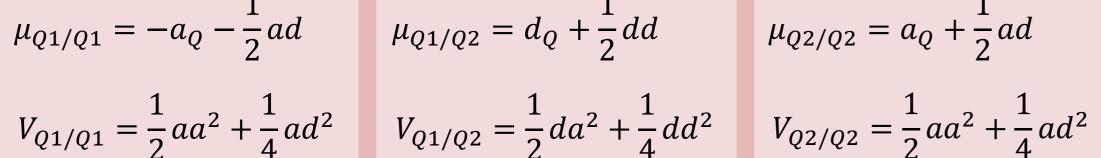
Single locus effect Allele frequency

_										
	Locus Q	Q1/Q1	Q1/Q1	Q1/Q1	Q1/Q2	Q1/Q2	Q1/Q2	Q2/Q2	Q2/Q2	Q2/Q2
	Locus R	R1/R1	R1/R2	R2/R2	R1/R1	R1/R2	R2/R2	R1/R1	R1/R2	R2/R2
	a_Q	-1	-1	-1	0	0	0	1	1	1
	a_R	-1	0	1	-1	0	1	-1	0	1
	d_Q	0	0	0	1	1	1	0	0	0
	d_R	0	1	0	0	1	0	0	1	0
	aa	1	0	-1	0	0	0	-1	0	1
	ad	0	-1	0	0	0	0	0	1	0
	da	0	0	0	-1	0	1	0	0	0
	dd	0	0	0	0	1	0	0	0	0
	Effect Q	-a _o +aa	-a _o -ad	-a _o -aa	d _o -da	d _O +dd	d _o +da	a _o -aa	a _o +ad	a _o +aa





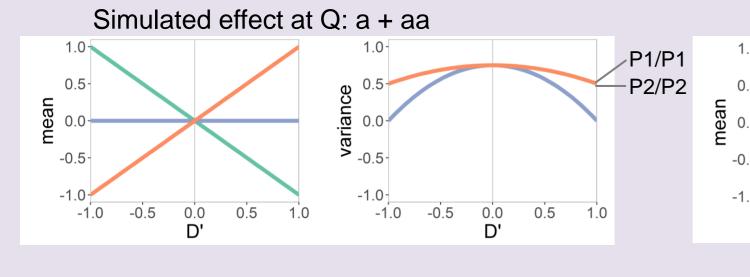


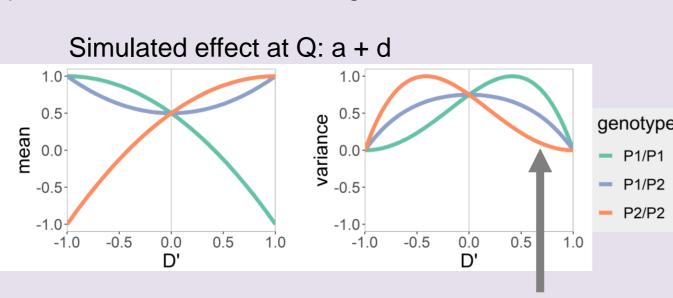


$$\mu_{Q2/Q2} = a_Q + \frac{1}{2}ad$$

$$1 \qquad 1$$

Example at P with no linkage between Q-R and p = q = r = 0.5. Note: D' = linkage between P-Q.





Variance heterogeneity can arise in the absence of interaction!

vQTL tests

Brown-Forsythe (BF) test

$$y_{ij} = \left| x_{ij} - median(x_{ij}) \right|$$

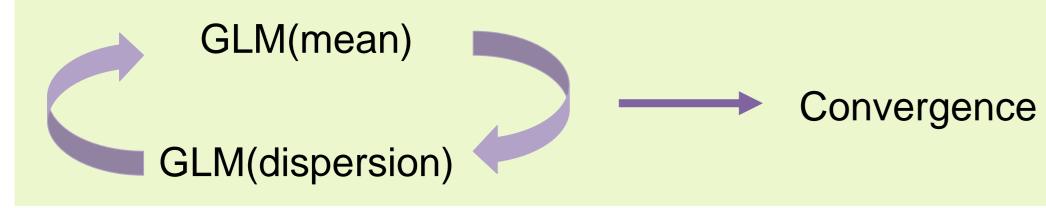
$$Y = Y_{i=1} \cup Y_{i=2} \cup Y_{i=3}$$

$$F_{p-1,N-p} = \frac{N-p}{p-1} \cdot \frac{\sum_{i=1}^{p} n_i \cdot \left(\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} - \frac{1}{N} \sum_{k=1}^{N} y_k\right)^2}{\sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(y_{ij} - \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}\right)^2}$$

Note: p = 2 in inbred markers.

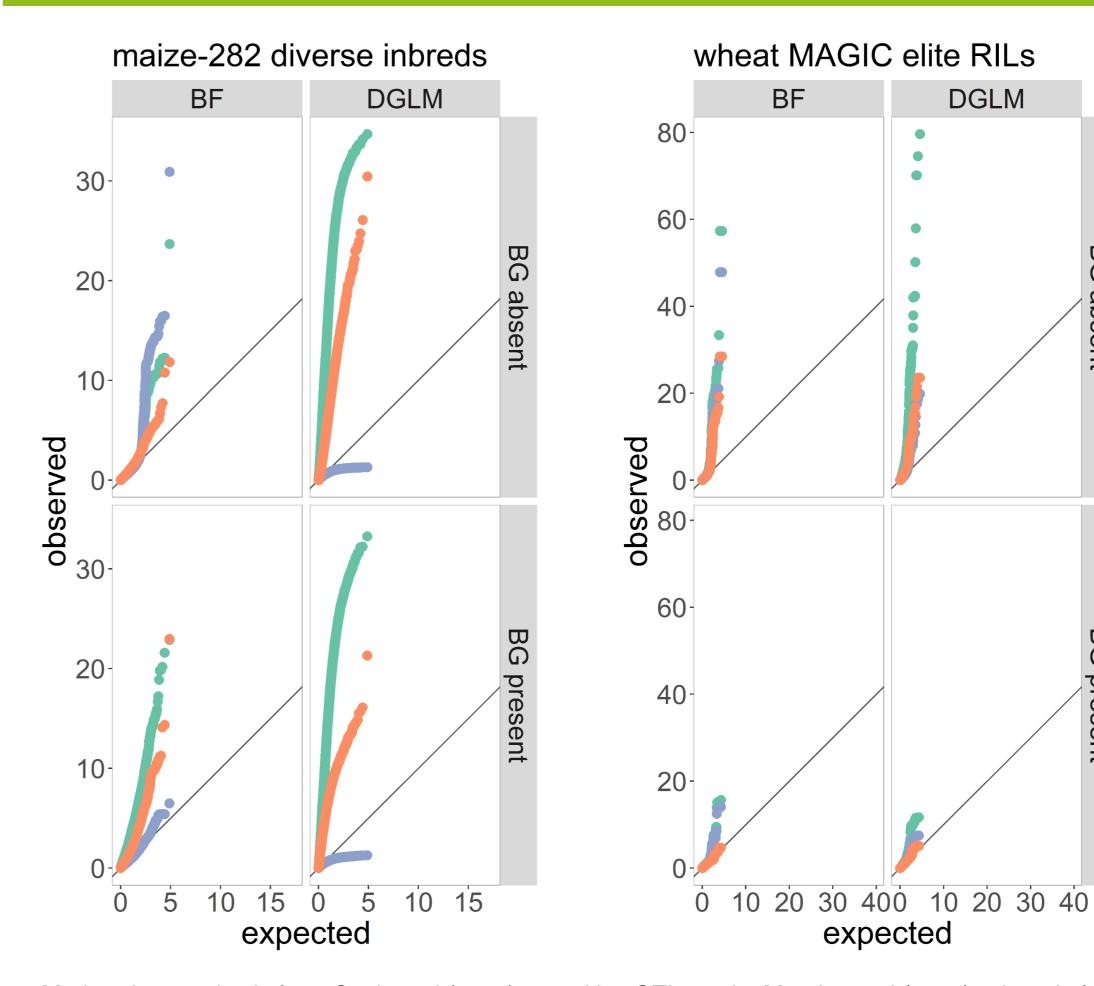
p = 3 in non-inbred markers. Brown and Forsythe (1974)

Double Generalized Linear Model (DGLM)



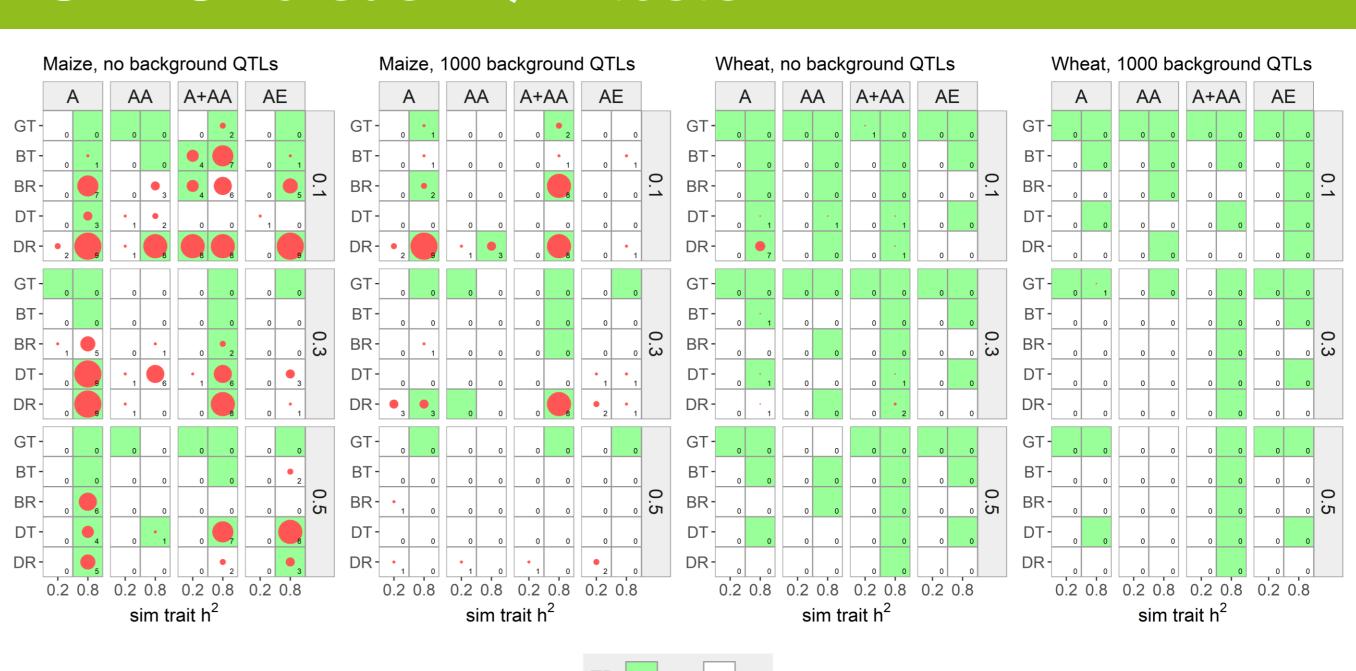
Smyth (1989)

Inflation control



- Marker data: maize is from Cook et al (2012) used in vQTL test by Murphy et al (2022), wheat is from Mackay et al (2014). • Simulation: 2 main A+AA QTLs with 0.1 frequencies (h² = 0.8 if BG is absent, otherwise 0.4), 1,000 BG A QTLs with h² = 0.4.
- BG: background, A: additive effect, AA: additive-by-additive effect, GC: genomic control (Devlin and Roeder 1999).

GWAS versus vQTL tests



True positive (TP) and false positive (FP) for GWAS and vQTL tests in simulated traits.

- TP/FP is loosely measured as chromosomes with significant hits that match (or not) to the simulated main QTLs.
- TP is indicated by green/white squares, and FP is shown as red circles proportional to the number of FP.
- Y-axis: GT (GWAS, trait), BT (BF test, trait), BR (BF test, residual), DT (DGLM, trait), DR (DGLM, residual).
- X-axis: total trait heritabilities which include simulated background and main QTLs.
- Column panel: main QTLs as A (additive), AA (additive-by-additive), AE (additive-by-environment).
- Row panel: main QTL allele frequencies.

Conclusions (tl;dr)

- Variance heterogeneity can arise due to epistasis and genetic-by-environment interaction, and in markers with incomplete linkage to non-interacting QTL.
- Variance heterogeneity may be absent between homozygous genotypes.
- Current methods of controlling inflation in test statistics due to population structure are imperfect – a robust method is needed.
- vQTL tests have high false positive rate especially at low QTL frequency.
- The original motivation for vQTL test is to act as a quick alternative to a full interaction test. A follow-up (reduced) interaction test is recommended.

References

- Brown MB & Forsythe AB (1974). Robust tests for the equality of variances. J Amer Stat Assoc 69: 364-367.
- Cook JP, et al (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. Plant Physiol 158: 824-834.
- Devlin B & Roeder K (1999). Genomic control for association studies. Biometrics 55: 997-1004.
- Mackay IJ, et al (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties and validation. G3 4: 1603-1610.
- Murphy M, et al (2022). Assessment of two statistical approaches for variance genome-wide association studies in plants. Heredity 129: 93-102.
- Smyth G (1989). Generalized linear models with varying dispersion. J Royal Stat Soc 51: 47-60.

Acknowledgement

We would like to thank Rajiv Sharma, Ian Dawson, David Marshall and Nicola Rossi for their help and discussion in the project. All the names listed (including us authors) are members of the Principal's Research Group at SRUC. Please check us out for cool research interests and projects!

