Defining the causes and consequences of phantom epistasis in heterotic hybrids

CJ Yang, Wayne Powell, Ian Mackay

18th Eucarpia Biometrics in Plant Breeding Conference

22nd September 2022

1. Definition and example of phantom epistasis.

2. Quick introduction to heterosis.

3. Distribution of mid-parent heterosis.

4. Estimating variance components.

5. Single locus and interaction GWAS.

Phantom/apparent epistasis:

Significant epistatic effect when the causative QTL effect is purely additive and/or dominance.

Imperfect Linkage Disequilibrium Generates Phantom Epistasis (& Perils of Big Data)

Gustavo de los Campos ✉, Daniel Alberto Sorensen, Miguel Angel Toro

G3 Genes|Genomes|Genetics, Volume 9, Issue 5, 1 May 2019, Pages 1429–1436,
https://doi.org/10.1534/g3.119.400101
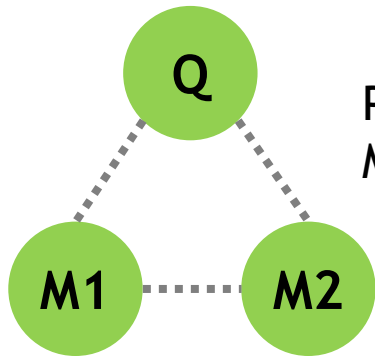
Phantom epistasis can arise between M1 and M2 if there are imperfect LDs.

Example of phantom epistasis

RETRACTED ARTICLE: Detection and replication of epistasis influencing transcription in humans

Gibran Hemani ✉, Konstantin Shakhbazov, Harm-Jan Westra, Tonu Esko, Anjali K. Henders, Allan F. McRae, Jian Yang, Greg Gibson, Nicholas G. Martin, Andres Metspalu, Lude Franke, Grant W. Montgomery, Peter M. Visscher & Joseph E. Powell

Nature **508**, 249–253 (2014) | Cite this article

Another explanation for apparent epistasis

Andrew R. Wood, Marcus A. Tuke, Mike A. Nalls, Dena G. Hernandez, Stefania Bandinelli, Andrew B. Singleton, David Melzer, Luigi Ferrucci, Timothy M. Frayling ✉ & Michael N. Weedon ✉

*Nature* **514**, E3–E5 (2014) | Cite this article

Phantom epistasis between unlinked loci

Gibran Hemani ✉, Joseph E. Powell, Huanwei Wang, Konstantin Shakhbazov, Harm-Jan Westra, Tonu Esko, Anjali K. Henders, Allan F. McRae, Nicholas G. Martin, Andres Metspalu, Lude Franke, Grant W. Montgomery, Michael E. Goddard, Greg Gibson, Jian Yang & Peter M. Visscher
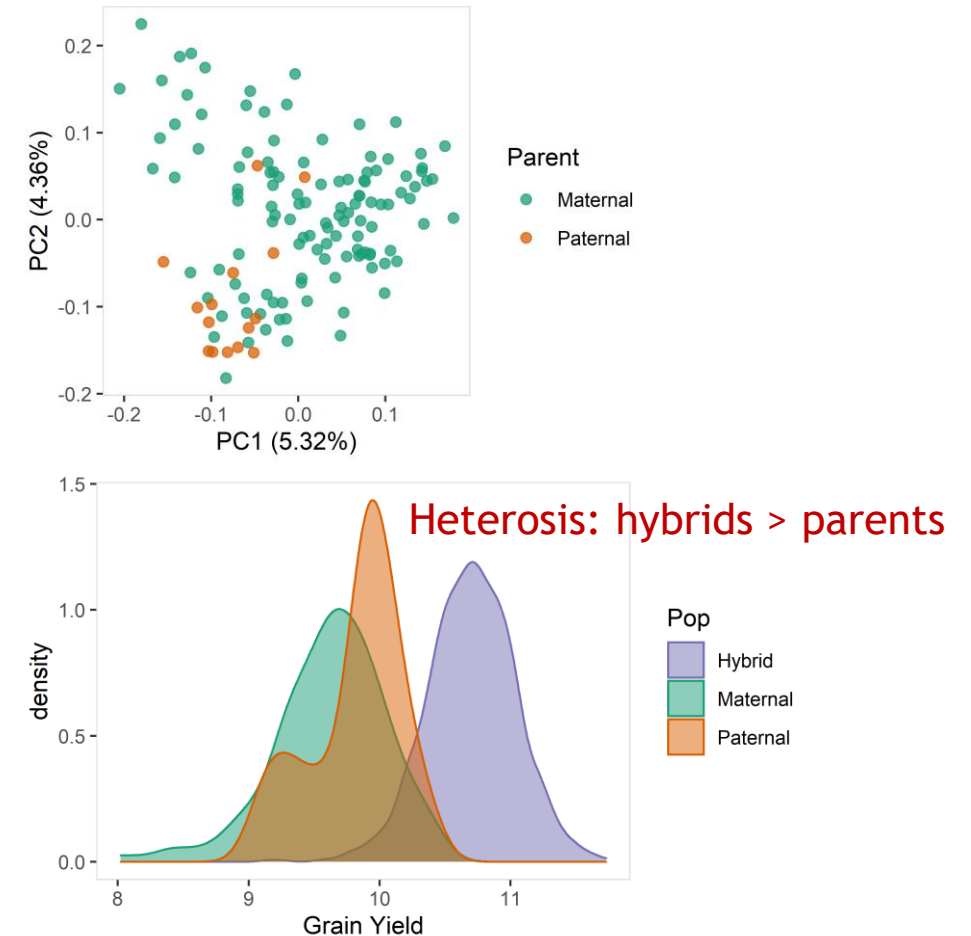
Nature **596**, E1–E3 (2021) | Cite this article

Heterotic pools have been long established in maize.

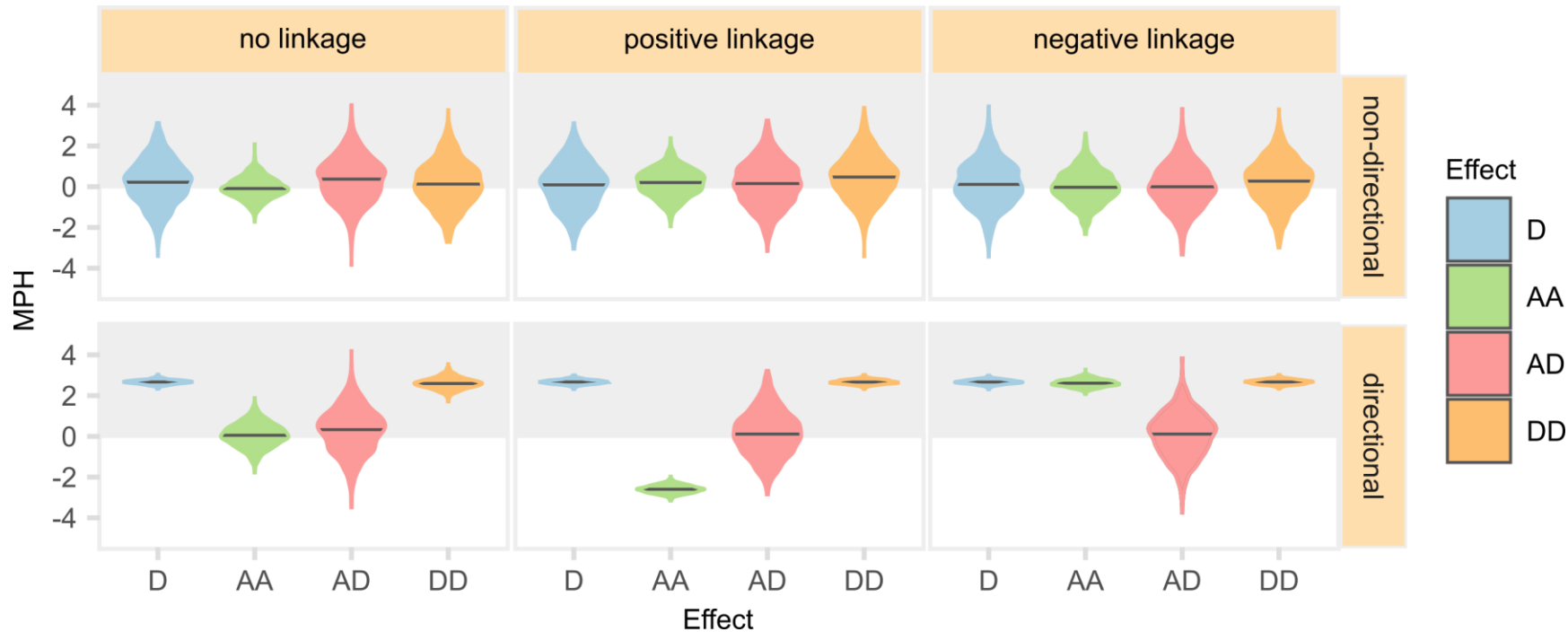Heterotic pools don't quite exist in wheat, yet.
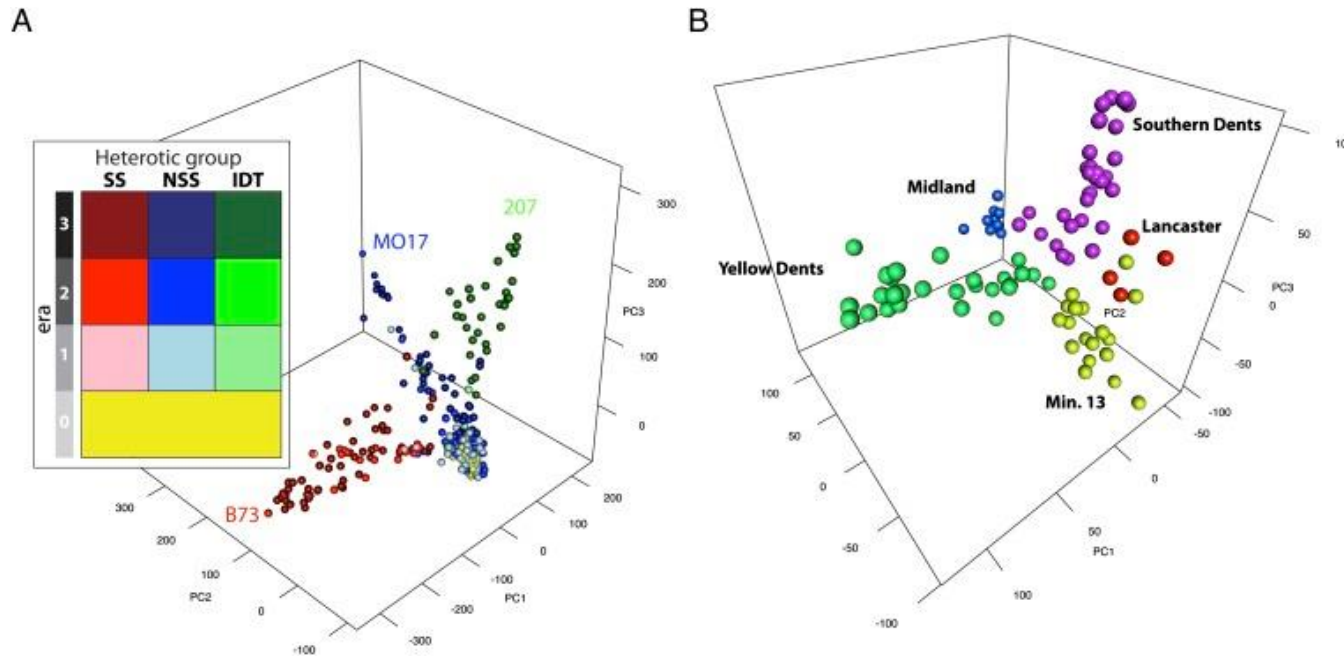
van Heerwaarden et al (2012) Historical genomics of North American maize.
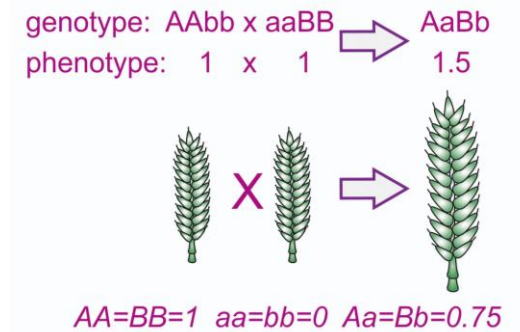
Heterosis: hybrids > parents

Zhao et al (2015) Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding.

Mid-parent heterosis (MPH)

(e) Heterosis is also explained by dispersed dominant alleles

Example of heterosis due to dispersed favorable alleles. Mackay et al (2021)

- Simulate 4 effects, 1,000 loci (×2 if linked), 100 × 2 parents, 1,000 hybrids.

- MPH = Hybrid – Parental Average.

- Directional effect is needed for non-zero MPH.

- Linkage is needed for non-zero MPH due to AA.

Variance components in hybrids

1. Standard mixed model

$$y_h = X\beta + \sum g_i + \varepsilon \qquad i = A, D, AA, AD, DD, \ldots$$

Fixed effects   Random effects

$$g_i \sim N(0, K_i \sigma_i^2)$$

$$\varepsilon \sim N(0, I\sigma_\varepsilon^2)$$

y_h is a vector of phenotypic trait in hybrids.

K_i is the genomic relationship matrix.

Both models are fitted using:

- R/sommer (Covarrubias-Pazaran 2016)
- R/BGLR + RKHS (Pérez and de los Campos 2014)

2. Transformed mixed model

$$y = X\beta + \sum g_i + \varepsilon \qquad y = y_h \cup y_p$$

Transformation #1

$$Ty = TX\beta + \sum T g_i + T\varepsilon$$

$$\tilde{g}_i \sim N(0, \tilde{K}_i \sigma_i^2)$$

$$\tilde{y} = \sum \tilde{g}_i + \tilde{\varepsilon} \qquad \tilde{\varepsilon} \sim N(0, TT'\sigma_\varepsilon^2)$$

TT' is non-trivial; get a 2nd transformation matrix V^{-1} by eigen-decomposition.

Transformation #2

$$TT' = Q\Lambda Q' = VV'$$

$$V^{-1}\tilde{y} = \sum V^{-1}\tilde{g}_i + \tilde{\varepsilon}$$

$$\hat{g}_i \sim N(0, \hat{K}_i \sigma_i^2)$$

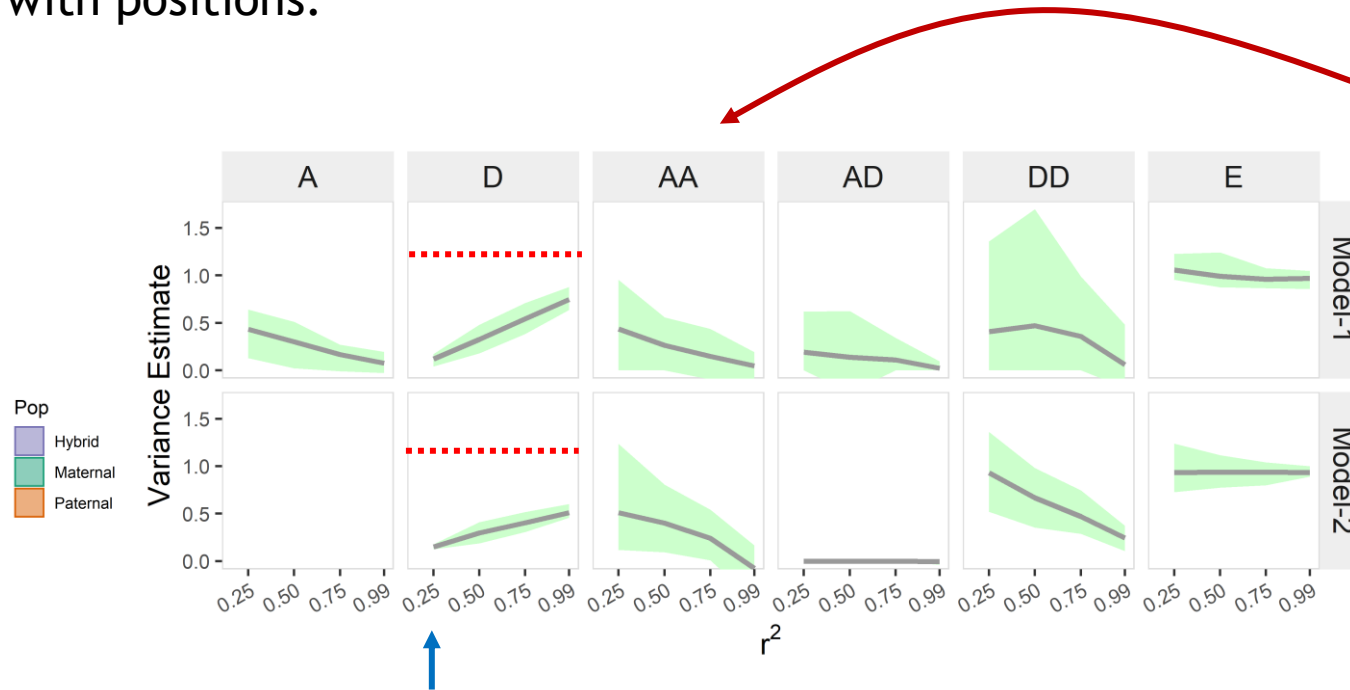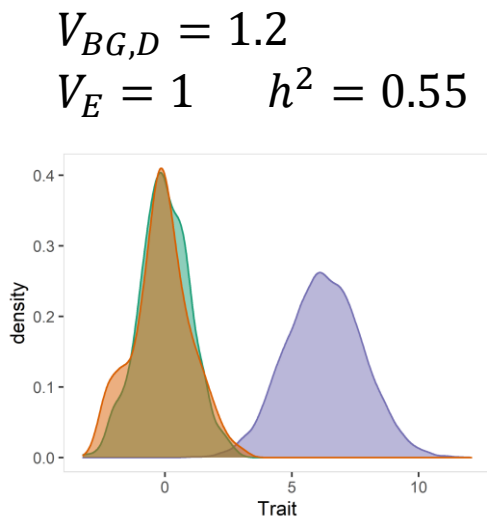$$\hat{y} = \sum \hat{g}_i + \hat{\varepsilon} \qquad \hat{\varepsilon} \sim N\left(0, V^{-1}VV'V^{-1'}\sigma_\varepsilon^2\right) = N(0, I\sigma_\varepsilon^2)$$

Jiang et al (2017)

- Hybrid wheat data from Zhao et al (2015).
  - 120 maternal parents + 15 paternal parents.
  - 1,604 hybrids
  - 2,701 markers with positions.

- Simulate 100 background QTL with various effects.
- Thin markers based on linkage to QTLs.
- 10 simulations for each genetic architecture.

$$V_{BG,D} = 1.2$$
$$V_E = 1 \qquad h^2 = 0.55$$

Presence of epistatic variance in the absence of epistatic QTLs

Huang and Mackay (2016) The genetic architecture of quantitative traits cannot be inferred from variance component analysis.

Loss in dominance variance at lower marker density.

Variance components in simulated traits
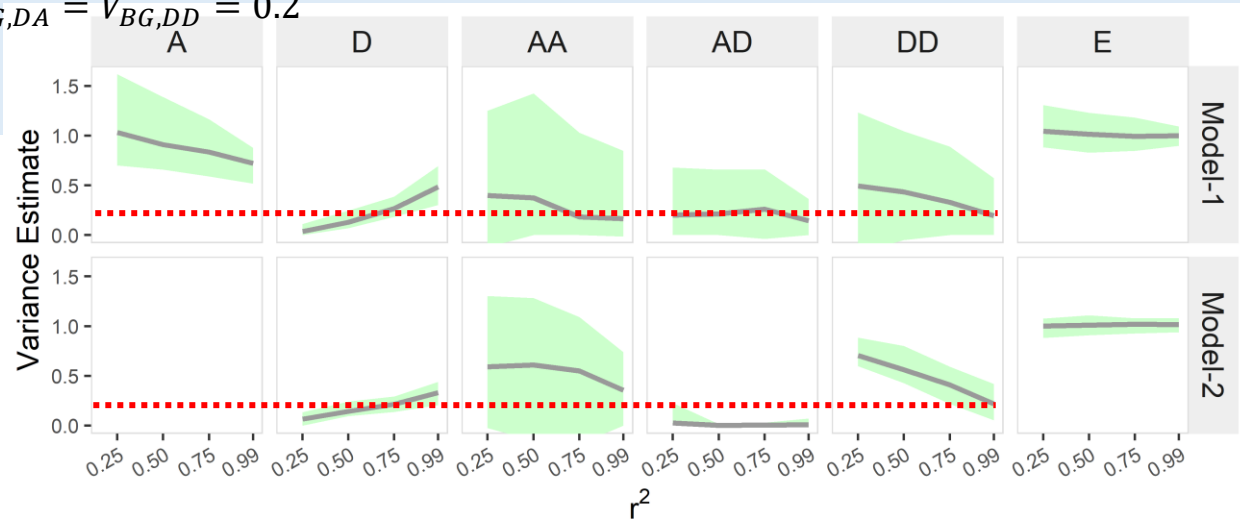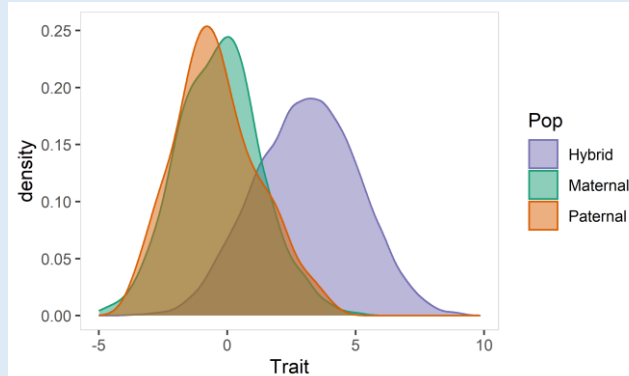
$V_{BG,AA} = 1.2$

$V_E = 1 \quad h^2 = 0.55$
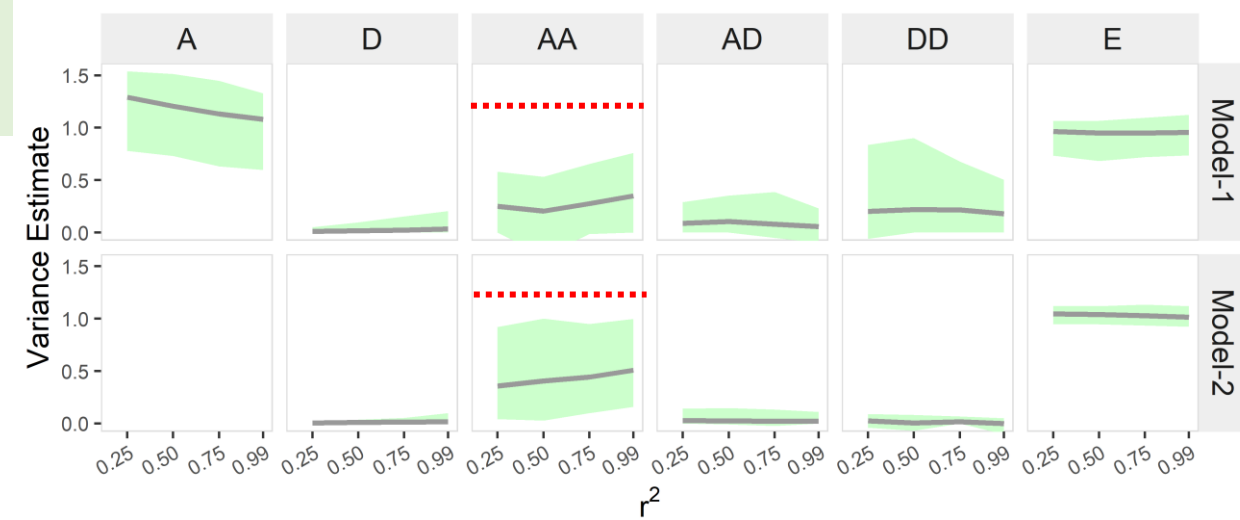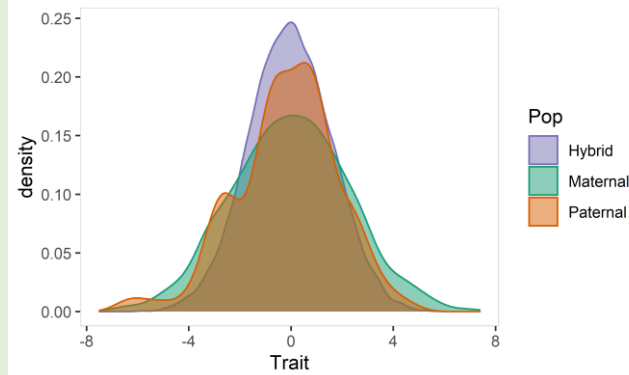
Additive x additive (AA) variance manifests as Additive (A) variance.

Hill et al (2008) Data and theory point to mainly additive genetic variance for complex traits

$V_{BG,A} = V_{BG,D} = V_{BG,AA} = V_{BG,AD} = V_{BG,DA} = V_{BG,DD} = 0.2$
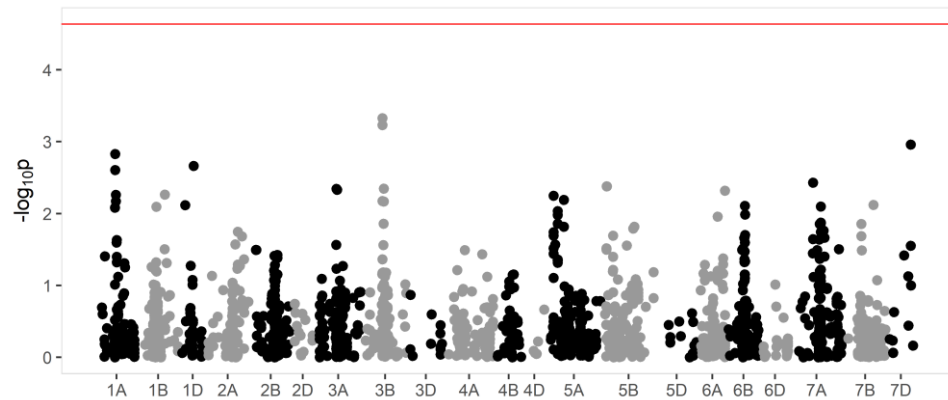
$V_E = 1 \quad h^2 = 0.55$

It gets harder to interpret when the underlying genetic architecture increases in complexity.

Not easy to determine where each variance is getting absorbed into, and the estimates are less consistent across simulations.
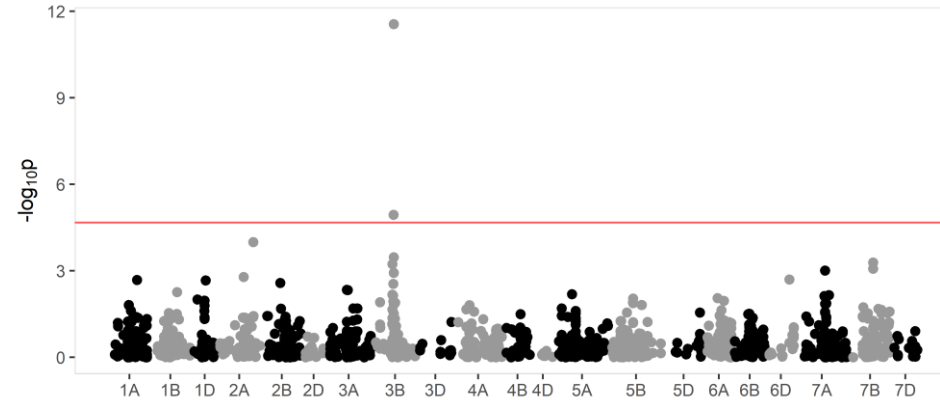
Simulated 200 background QTLs ($V_{BG,A} = 0.8, V_{BG,D} = 0.4$), 1 main QTL on 3B ($V_{M,A} = 0.4, V_{M,D} = 0.2$) and $V_E = 1$.
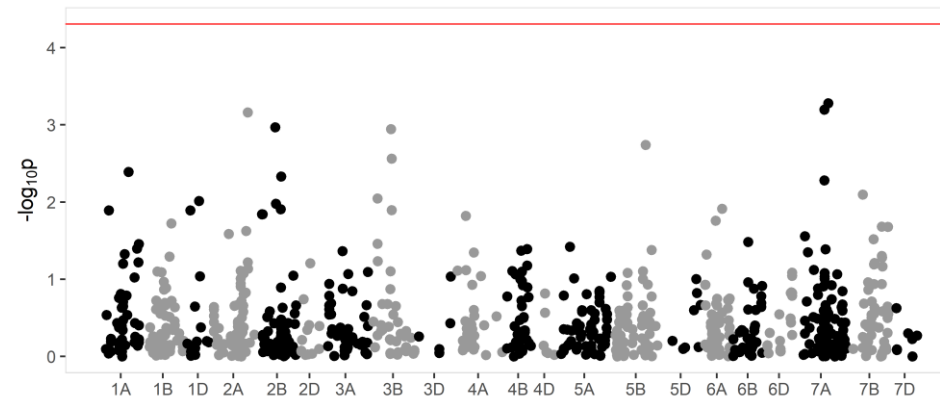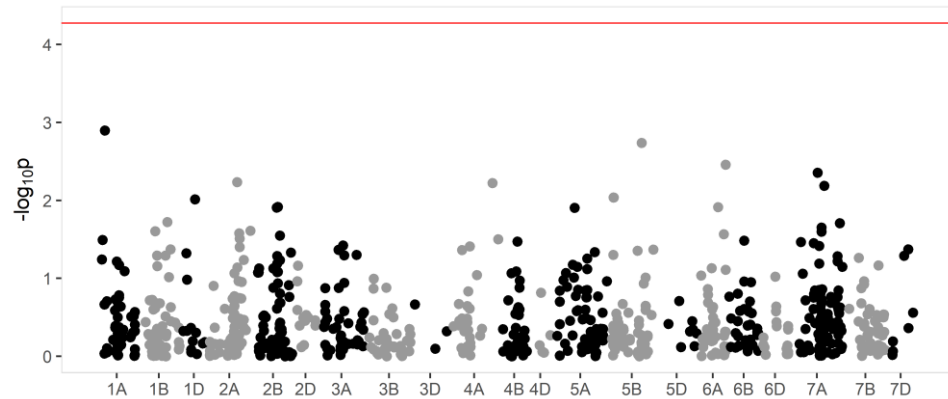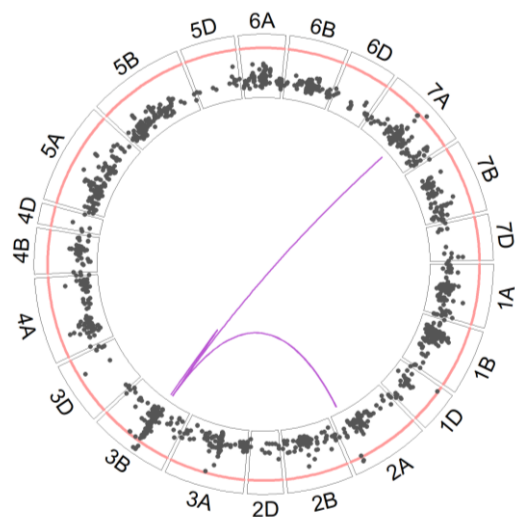
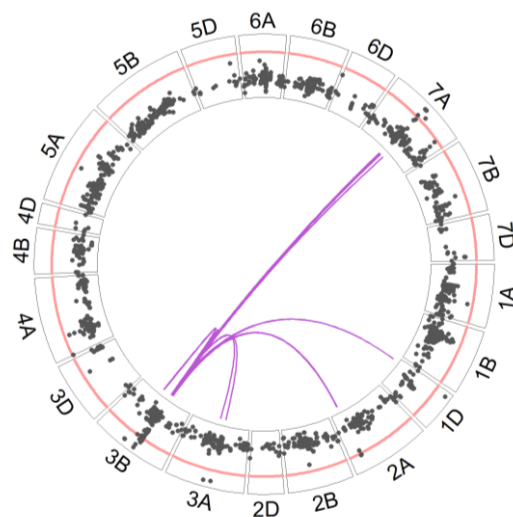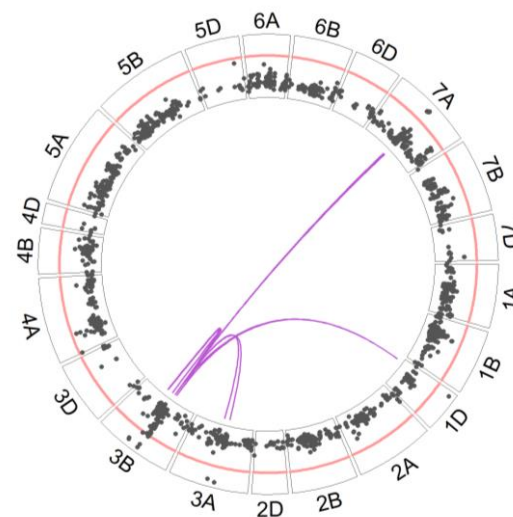**GWAS – Additive**

GWAS – Dominance

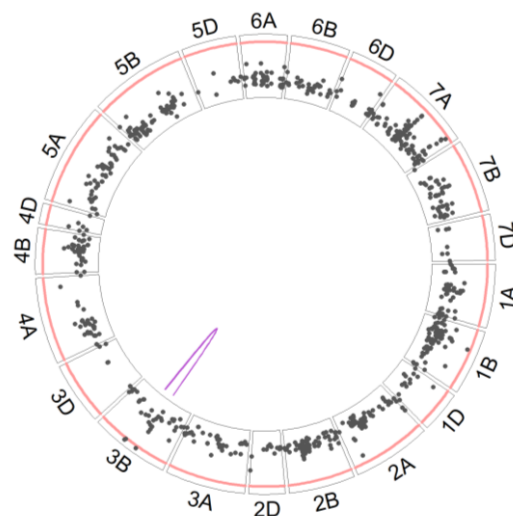No markers with $r^2 > 0.99$.

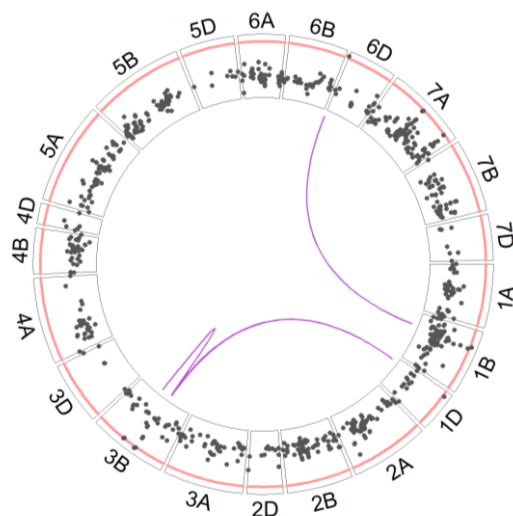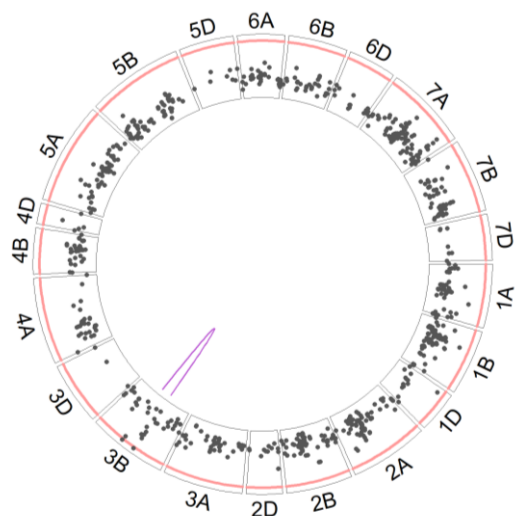No markers with $r^2 > 0.25$.

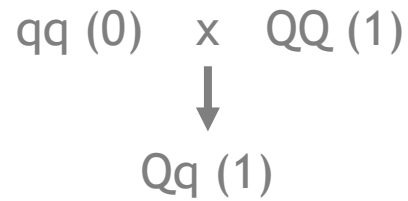GWAS – A x A

GWAS – A x D

GWAS – D x D

No markers with $r^2 > 0.99$.

No markers with $r^2 > 0.25$.

Take home messages

1. Heterosis arises due to directional dominance (and dominance x dominance).

$$qq\ (0) \quad x \quad QQ\ (1)$$

$$Qq\ (1)$$

2. Heterosis can also arise due to linked (dispersed) and directional additive x additive interaction.
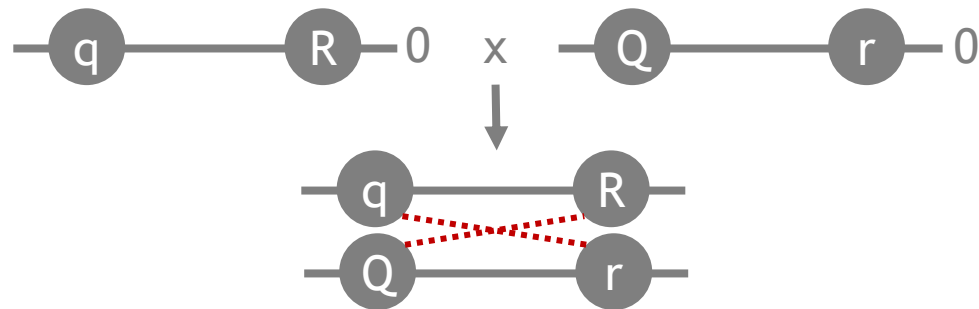
3. Inference of genetic architecture from variance components is hard – further complicated by marker density.

$$V_D = 1.2$$

$$V_D = ?\,, V_{AA} = ?$$
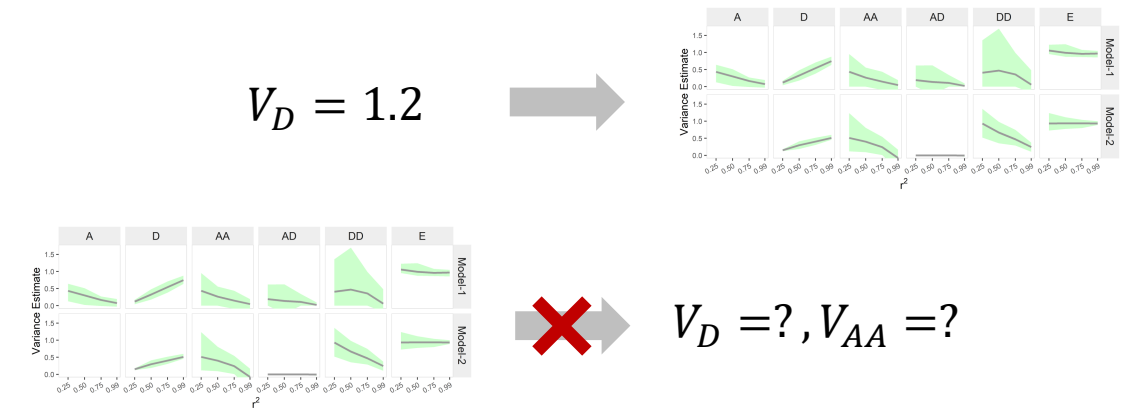
4. Inference of genetic architecture from GWAS is not any better – hard to control for false positive in interaction GWAS.

Acknowledgement

Principal's Research Group

The Principal's Research Group brings together plant breeding expertise from across SRUC to boost the productivity, sustainability and resilience of food systems across the world.
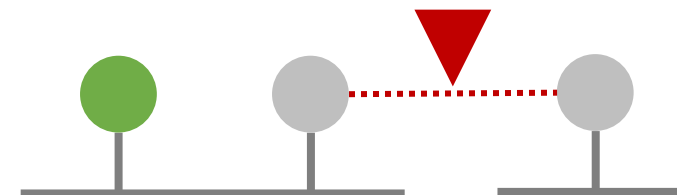
Wayne Powell

Ian Mackay

Rajiv Sharma

Ian Dawson

David Marshall

Nicola Rossi

Please feel free to reach out to us if you have any question or interest in collaboration!

@hataraku_cj
@SRUCPrincipal
@IanJMackay

cyang@sruc.ac.uk

This work is not possible without the following supports:

Conference organizers

BIOMETRICS IN PLANT BREEDING
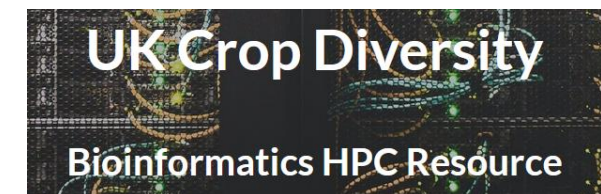EUCARPIA
2022

Data source

Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding

Yusheng Zhao[a], Zuo Li[a], Guozheng Liu[a], Yong Jiang[a], Hans Peter Maurer[b], Tobias Würschum[b], Hans-Peter Mock[c], Andrea Matros[c], Erhard Ebmeyer[d], Ralf Schachschneider[e], Ebrahim Kazman[f], Johannes Schacht[g], Manje Gowda[b,1], C. Friedrich H. Longin[b], and Jochen C. Reif[a,2]

[a]Department of Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, 06466 Stadt Seeland, Germany; [b]State Plant Breeding Institute, University of Hohenheim, 70593 Stuttgart, Germany; [c]Department of Physiology and Cell Biology, Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, 06466 Stadt Seeland, Germany; [d]Wheat breeding, KWS LOCHOW GmbH, 29296 Bergen, Germany; [e]Wheat breeding, Nordsaat Saatzuchtgesellschaft mbH, 38895 Langenstein, Germany; [f]Wheat breeding, Lantmännen SW Seed Hadmersleben GmbH, 39398 Hadmersleben, Germany; and [g]Wheat breeding, Limagrain GmbH, 31226 Peine-Rosenthal, Germany

External computational resource

UK Crop Diversity
Bioinformatics HPC Resource

Slides are available at: https://github.com/cjyang-work/presentation/