Leveraging bioinformatic breakthroughs into quantitative genetic approaches for crop improvement

ILLINOIS
Crop Sciences
COLLEGE OF AGRICULTURAL, CONSUMER & ENVIRONMENTAL SCIENCES

Alexander E. Lipka

Associate Professor,

Department of Crop Sciences,

University of Illinois, USA

Relationship between my research and bioinformatic breakthroughs

Dr. Matthew D. Murphy

G-to-P models

Genotypic data

https://tincture.io/tagged/dna-sequencing

Advanced statistical models that connect genotype to the phenotype

Phenotypic data

http://boort.com.au/gallery/drone-checking-corn-crop/
https://www.pioneer.com/us/products/soybeans.html

$$\begin{bmatrix} T & G & A & The \\ T & G & & Lipk\alpha \\ T & C & C & La\beta \end{bmatrix}$$

Slide courtesy of Dr. Matthew D. Murphy

Bioinformatic breakthroughs facilitate analysis of more sophisticated phenotypes

Fly drones

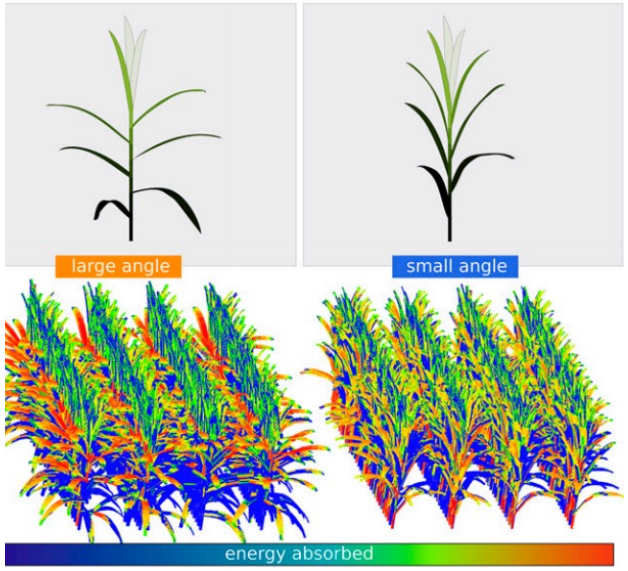
http://boort.com.au/gallery/drone-checking-corn-crop/

https://www.pioneer.com/us/products/soybeans.html

Take images

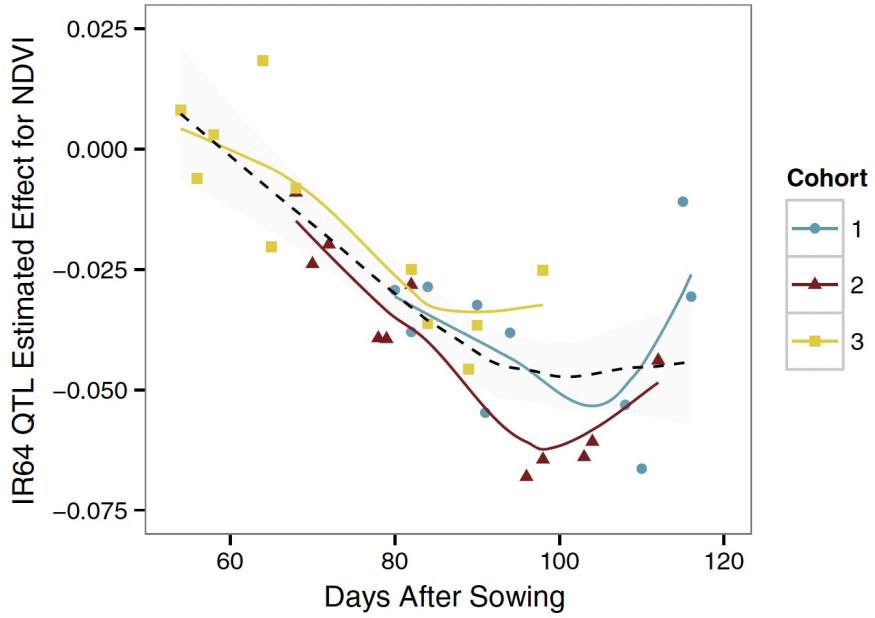
Lipka Lab growout from 2021 UIUC field season

Measure phenotypes



Truong et al., Genetics (2015) Genetics

Bioinformatic breakthroughs make it possible to see how G-to-P relationships change across lifespan



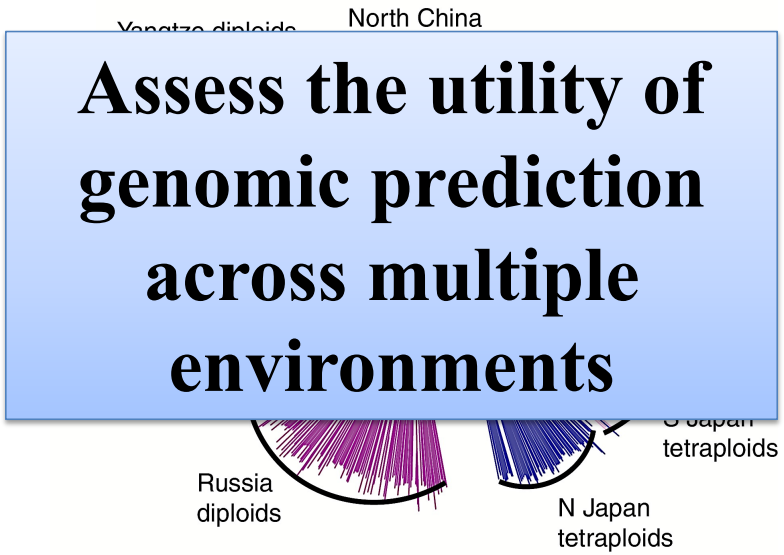
Family of *n* = 1,741 rice RILs

- Trait = Reflectance ratio (NVDI)
- Obtained using multispectral sensors on a tractor

Tanger et al., *Scientific Reports* (2017)

High quality bioinformatic data helped make exciting research possible in plants

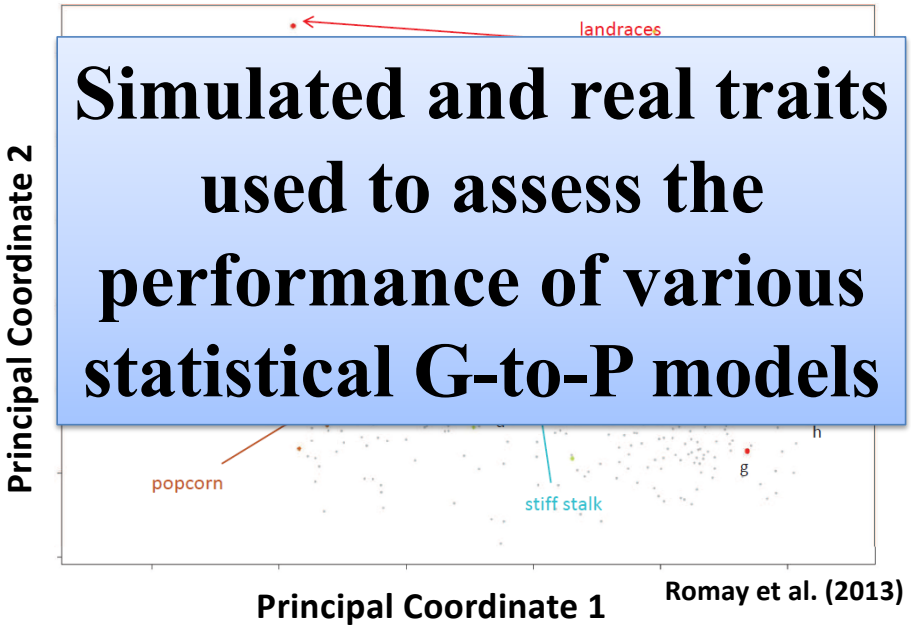
***Miscanthus sinensis* diversity panel: *n* = 538**



Clark *et al*. (2018)

Assess the utility of genomic prediction across multiple environments

NCRPIS ("Ames") maize diversity panel: *n* = 2,815

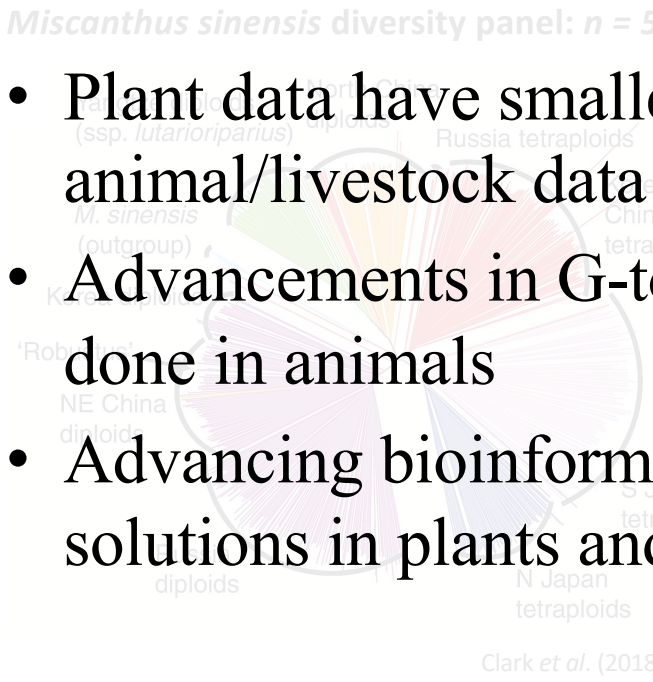


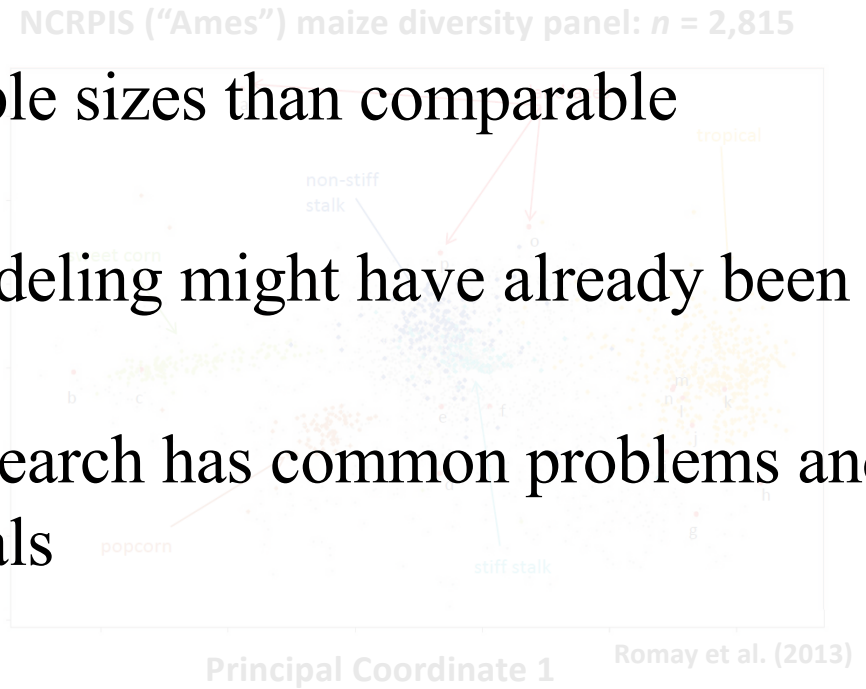
Romay et al. (2013)

Simulated and real traits used to assess the performance of various statistical G-to-P models

AG2PI has put my lab's research experience into a broader context

- Plant data have smaller sample sizes than comparable animal/livestock data
- Advancements in G-to-P modeling might have already been done in animals
- Advancing bioinformatic research has common problems and solutions in plants and animals





AGP2I Thinking Big: Advancing Genomics Research

Current Challenges and Future of Agricultural Genomes to Phenomes in the U.S.

Authors:

Christopher K. Tuggle[1]*#, Jennifer L. Clarke[2]#, Brenda M. Murdoch[3]#, Eric Lyons[4]#, Nicole M. Scott[1]#, Bedrich Beneš[5], Jacqueline D. Campbell[6], Sruti Das Choudhury[2], Henri Chung[1], Courtney L. Daigle[7], Jack C. M. Dekkers[1], Joao R. R. Dórea[8], David S. Ertl[9], Max Feldman[10], Breno O. Fragomeni[11], Janet E. Fulton[12], Carmela R. Guadagno[13], Darren E. Hagen[14], Andrew S. Hess[15], Luke M. Kramer[1], Carolyn J. Lawrence-Dill[1], Alexander E. Lipka[16], Thomas Lübberstedt[1], Fiona M. McCarthy[4], Stephanie D. McKay[17], Seth C. Murray[7], Penny K. Riggs[7], Troy N. Rowan[18], Moira J. Sheehan[19], Juan P. Steibel[1], Addie M. Thompson[20], Kara J. Thornton[21], Curtis P. Van Tassell[22], Patrick S. Schnable[1]*

More genomes need to be available

- **More individuals need to be sequenced**
  - Overreliance on a single reference genome

Sequencing at both genomic and epigenomic levels are needed

- **What this will facilitate**
  - Studying genetic diversity of relevant breeding material
  - Studying structural variation
  - Study history of genetic architecture

Basic science needs to be translated to applications

- **Basic genomic research has advanced**
  - Better understanding of non-additive effects
  - Better understanding of GxE

- **Improving analytical tools can expediate applications**
  - Practical
  - Usable
  - Understandable to users in multiple disciplines

A cohesive agricultural genomics community is needed

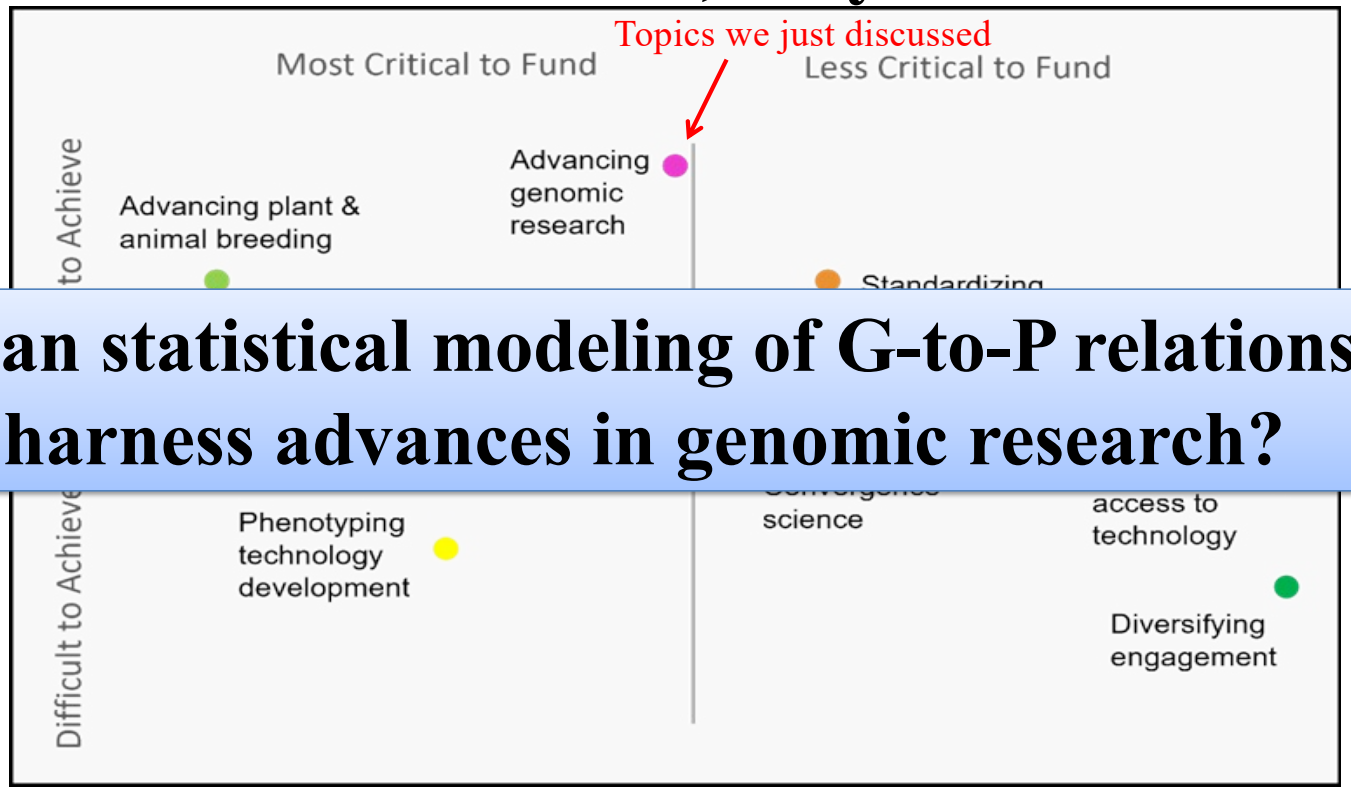
- **Support needed in the following areas**
  - Scientific
  - Funding
  - Human resources

- **Benefits**
  - Scientists can focus on science
  - Innovative scientific thinking is encouraged
  - Latest approaches can be used

Generate high-resolution and more diverse -omics data are needed

- **Species agnostic**
- **Serve as reference data sets**
  - Decrease data collection costs
- **Engaging scientists in industry ensures:**
  - Relevance
  - Accessibility to diverse stakeholders

Advancing genomic research: Critical to fund, easy to achieve



How can statistical modeling of G-to-P relationships harness advances in genomic research?

Tuggle et al., *In Review*

More multi-trait analyses are needed

Vector of multiple traits

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ \cdot \\ Y_k \end{pmatrix} = \mathbf{1}\mu + Q\beta + X\alpha + Zu + \varepsilon$$

Grand mean

Marker effect

Fixed effects: account for population structure

Random effects: account for familial relatedness

Random error vector

Q, ***X***. and ***Z*** are design matrices

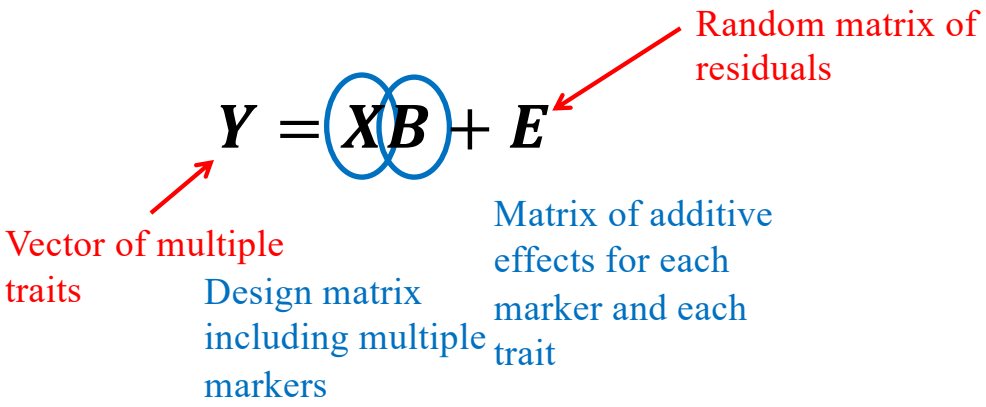
$\boldsymbol{u} \sim \text{MVN}(\mathbf{0}, 2K\sigma_G^2)$

K = kinship matrix

$\boldsymbol{\varepsilon} \sim \text{MVN}(\mathbf{0}, I\sigma_E^2)$

Yu et al. ***Nat. Genet*** **(2006)**

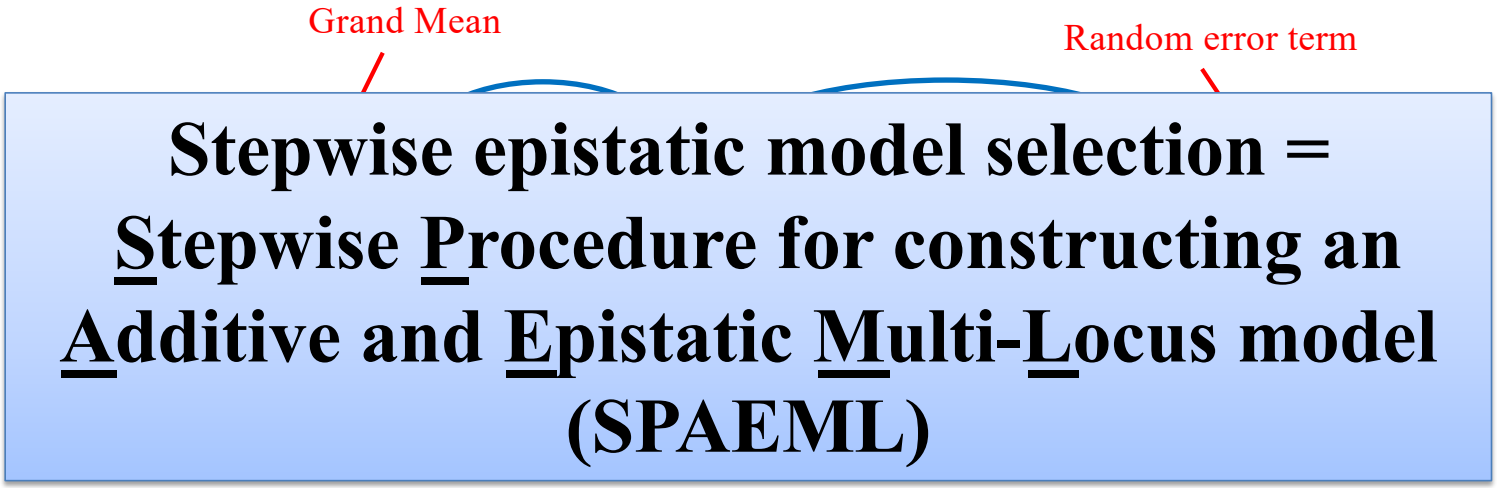
Multiple loci need to be considered in one model (MSTEP shown here)



- Determining the optimal model:
  - AIC, BIC, mBIC
  - Permutation procedure

Fernandes et al., *The Plant Genome* (2022)

Non-additive effects need to be modeled

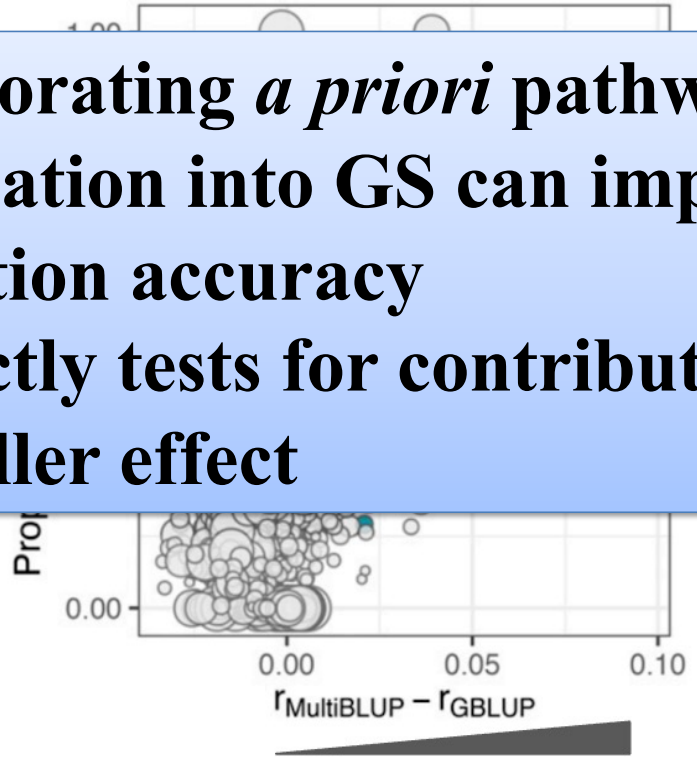


- *I* is a subset of markers with additive effects in model
- *U* is a subset of markers with two-way epistatic effects in model
- Determining the optimal model:
  - AIC, BIC, mBIC
  - Permutation procedure

Bogdan, Ghosh, and Doerge, *Genetics* (2004)

Contribution of non-statistically significant loci needs to be quantified

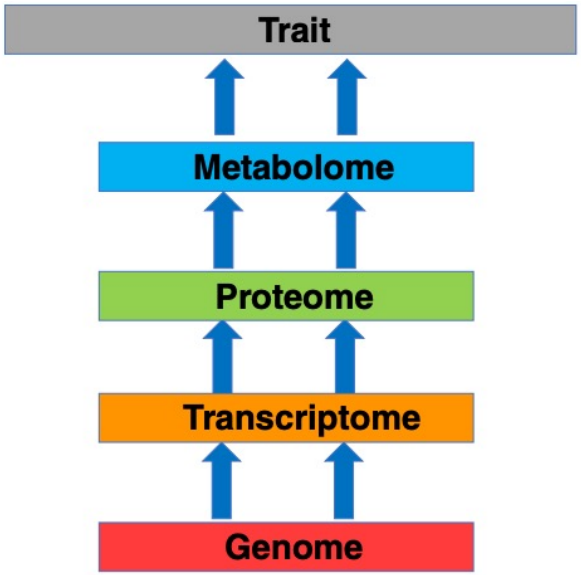
- **Incorporating *a priori* pathway information into GS can improve prediction accuracy**
- **Indirectly tests for contributions of genes of smaller effect**



Multi-kernel performs better than standard

Turner-Hissong et al., *G3* (2020)

-omic levels connecting intermediate steps between "G" and "P" need to be included in the model



$$Y = \mu\mathbf{1} + Zu_G + Zu_T + Zu_P + Zu_M + \varepsilon$$

$$u_G \sim MVN(0, G\sigma_g) \quad u_T \sim MVN(0, T\sigma_T)$$

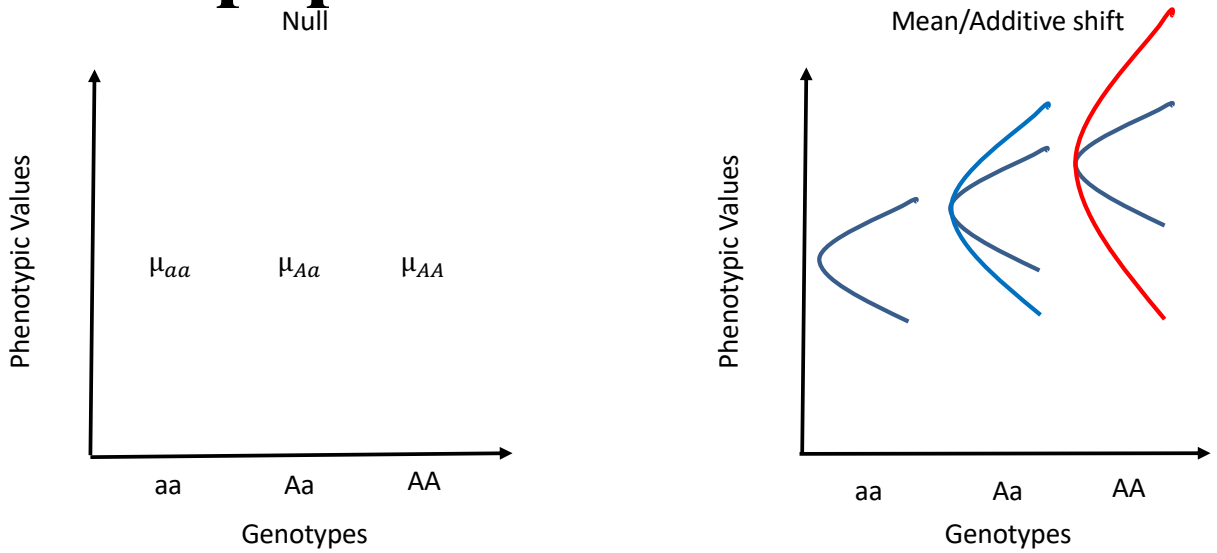
$$u_P \sim MVN(0, P\sigma_P) \quad u_M \sim MVN(0, M\sigma_m)$$

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

Rice and Lipka, *Mol. Breeding* (2021)

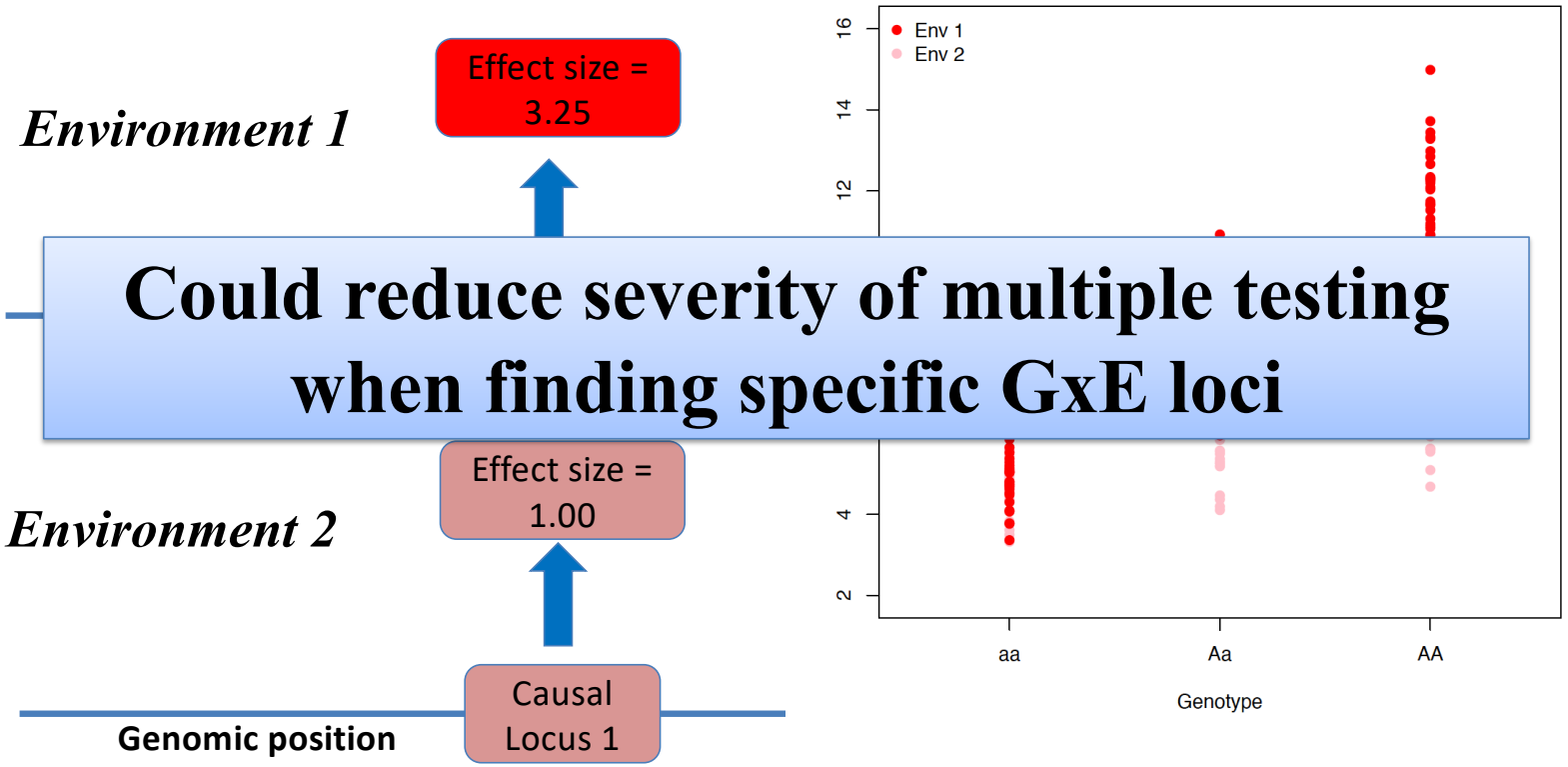
Dr. Matthew D. Murphy

Move beyond testing for differences in population mean trait values



Variance Quantitative Trait Loci (vQTL)
Variance GWAS (vGWAS)

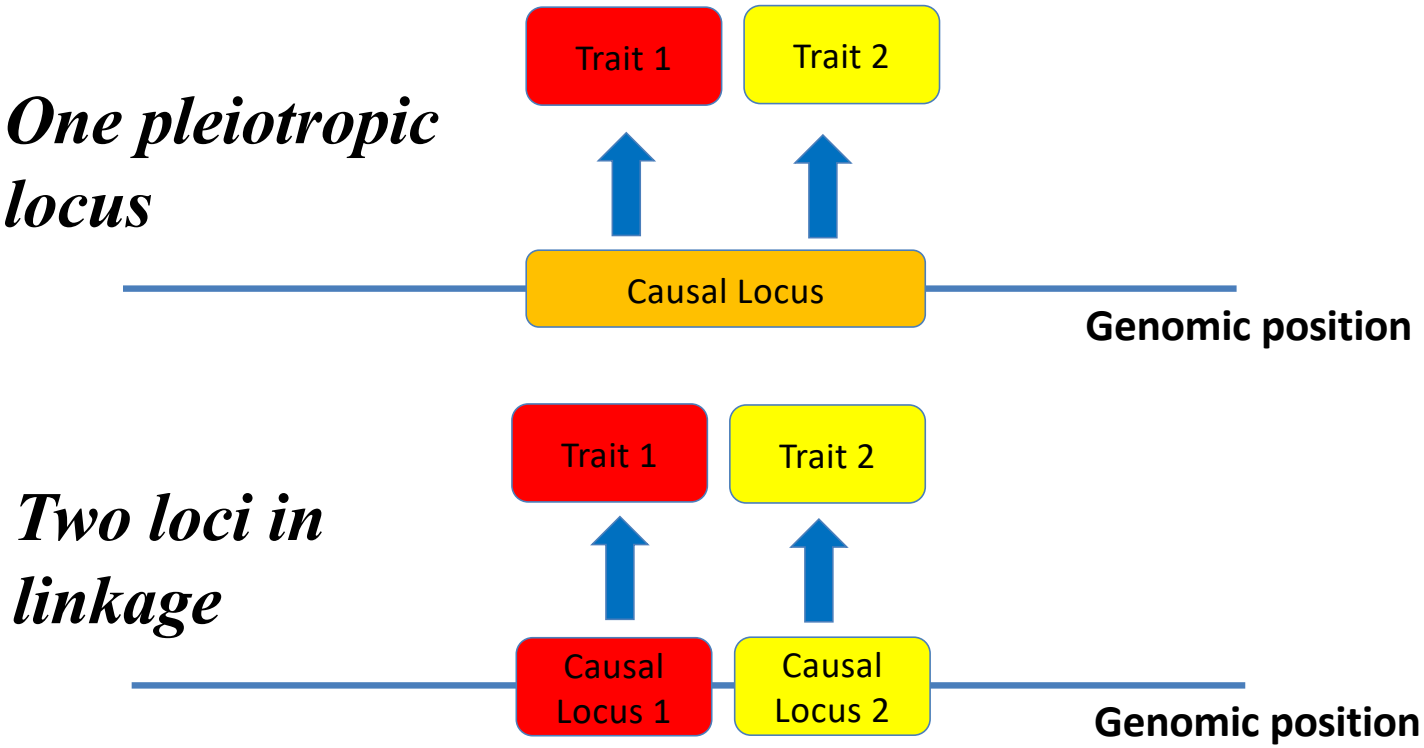
GxE interactions could appear as a vQTL



Murphy et al., *Heredity* (2022)

Bioinformatic breakthroughs can assist with testing scientific hypothesis on genetic architecture

Simple example: Pleiotropy versus linkage?



Fernandes et al., *Frontiers in Genetics* (2021)

Bioinformatic breakthroughs can assist with testing scientific hypothesis on genetic architecture

Complex example: Evidence for omnigenic and/or other genetic architectures?

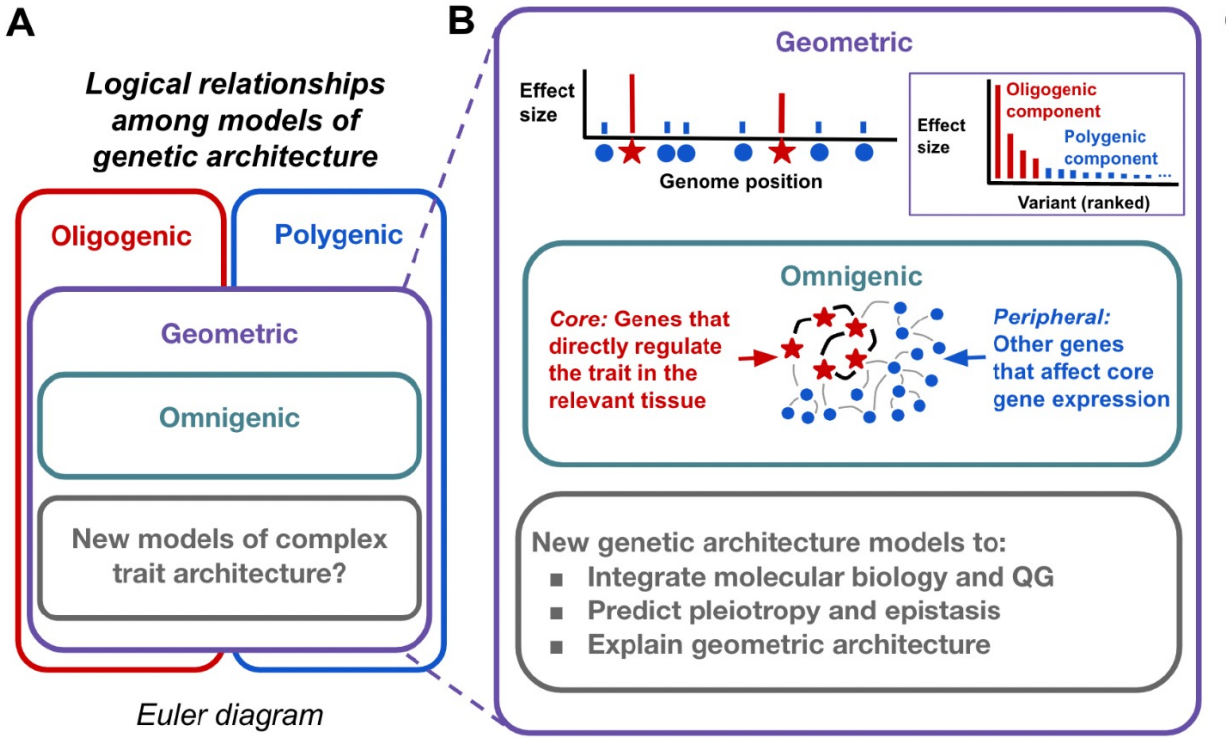


Figure created by Geoffrey P. Morris