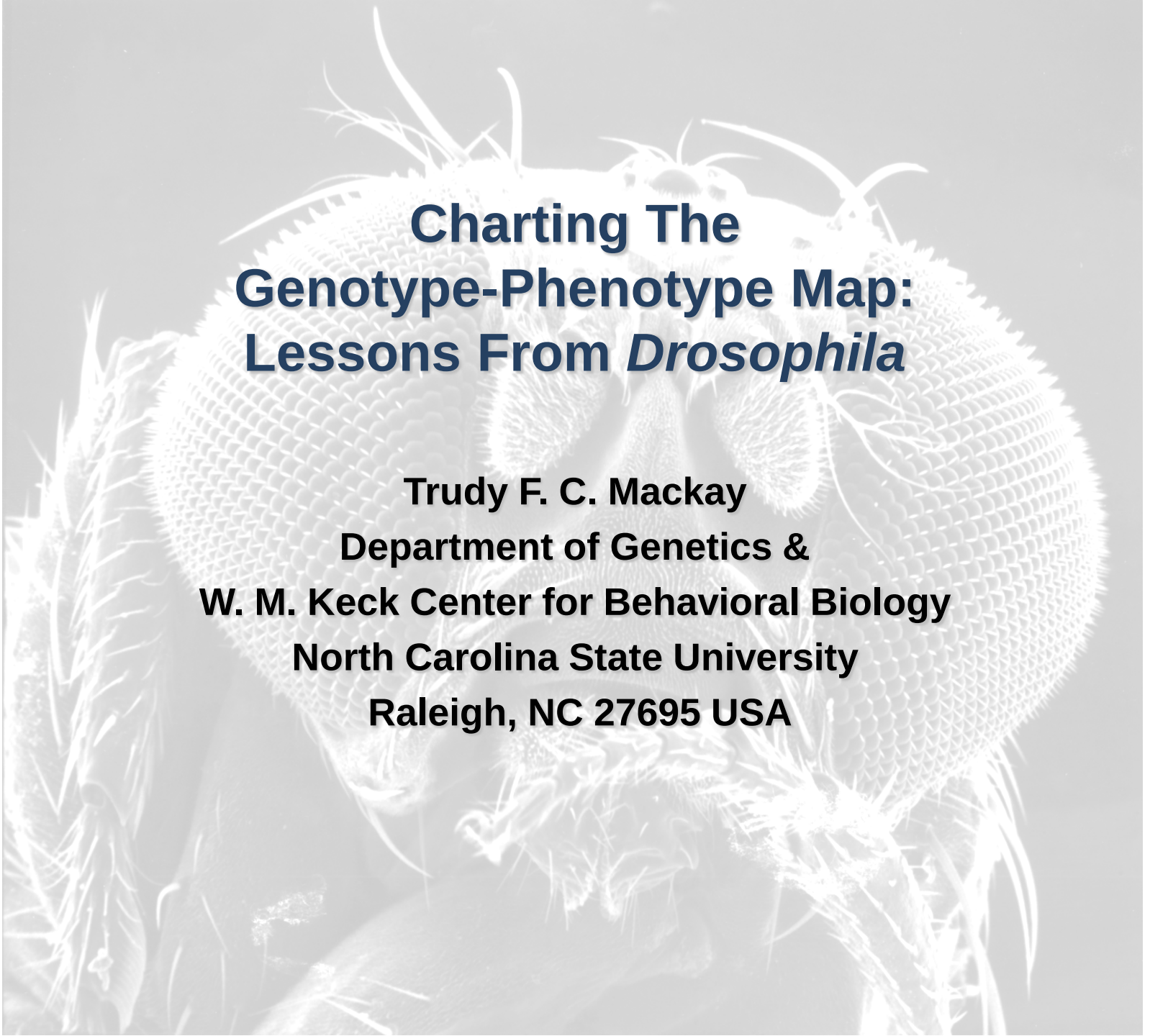
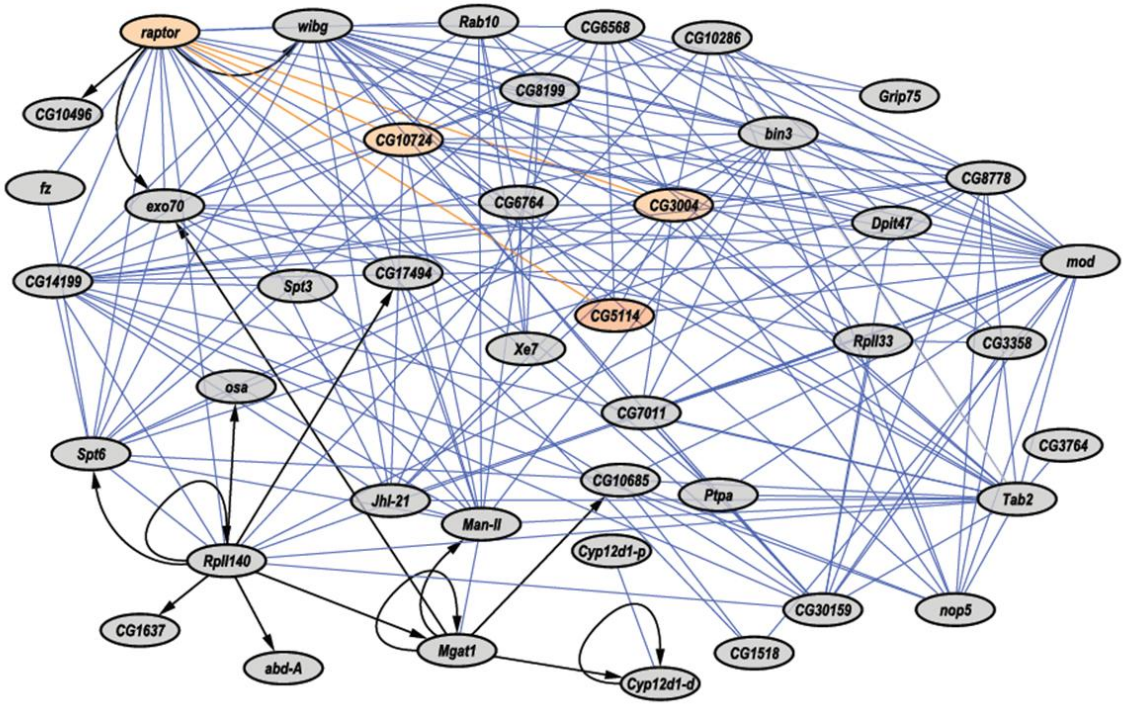


Charting The Genotype-Phenotype Map: Lessons From *Drosophila*

Trudy F. C. Mackay
Department of Genetics &
W. M. Keck Center for Behavioral Biology
North Carolina State University
Raleigh, NC 27695 USA

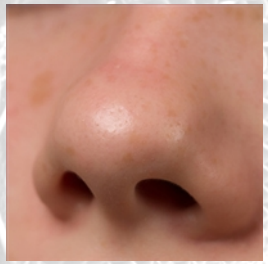
AACTGGCCTCCTCCTCCTCCTATACGTTACCGG……CGATTCCAAA
AACTGGCCTCCT……ATACGTTACCGG……AAACCTTAGC
AACTGGCCTCCT……ATACGTTACCGC……AAACCTTAGC
AACTGGCCTCCT……ATACGTTACCGC……Δ3kb……AAACCTTAGC



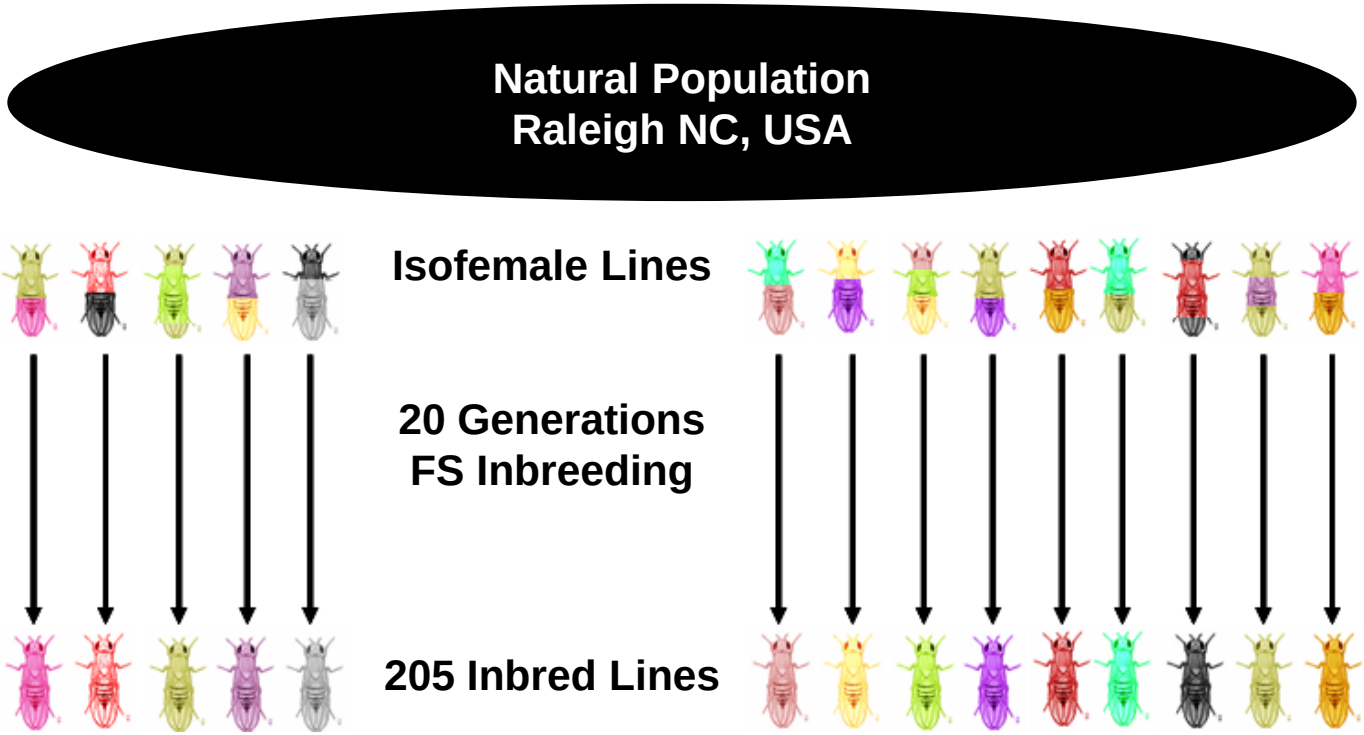
Genetic Resources

Genetic Manipulation

Model Organism

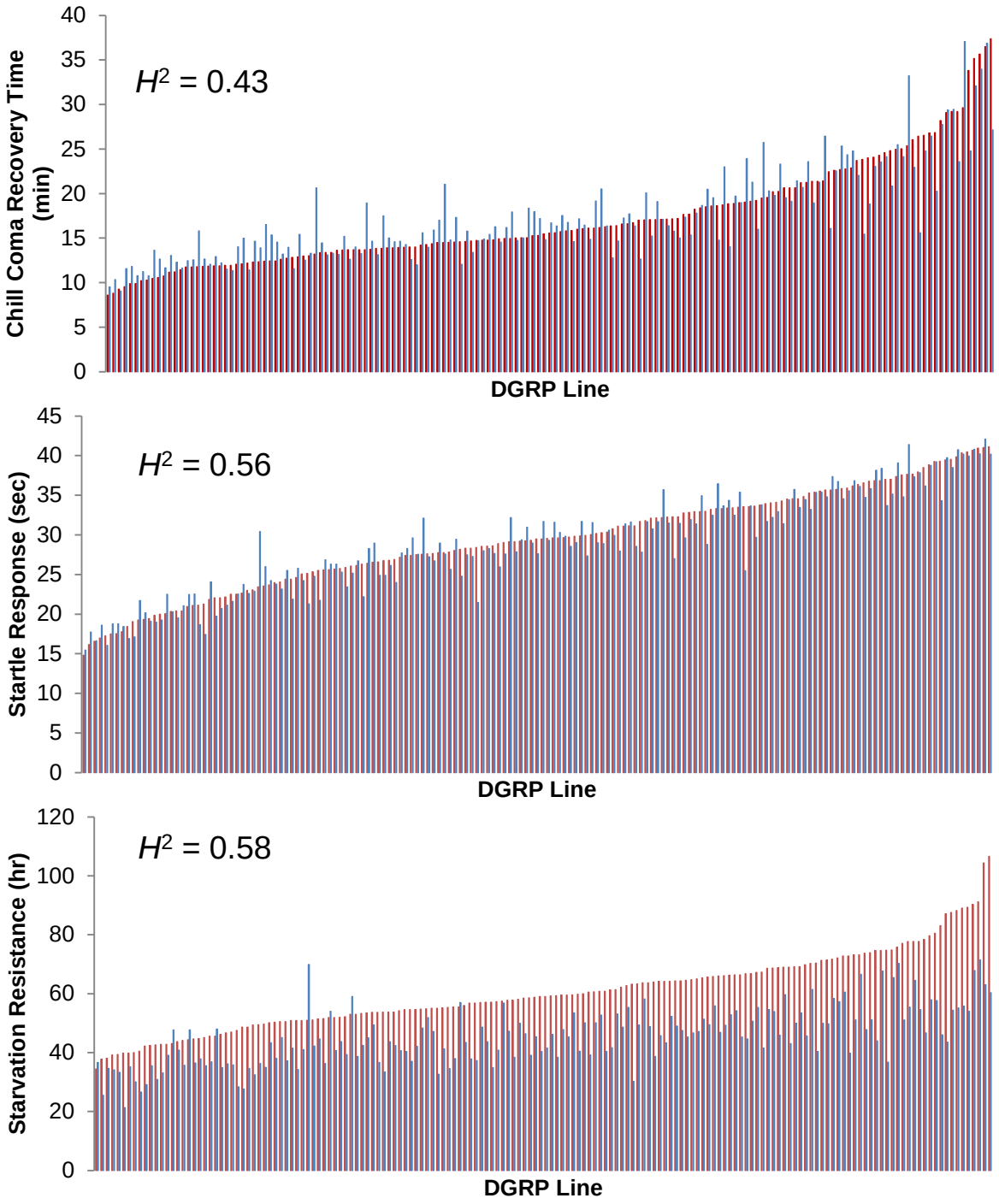


Drosophila melanogaster Genetic Reference Panel (DGRP)



- ❖ Living library of genetic variation
- ❖ Community resource for association mapping complex traits
- ❖ Replicated genotypes
- ❖ Accurate phenotypes, multiple phenotypes (pleiotropy), multiple environments (GEI)
- ❖ Freeze 1 genome sequences of 168 lines

Mackay *et al.* 2012. Nature 482:173


Chill Coma Recovery Time (min)
H^2 = 0.43
DGRP Line
Startle Response (sec)
H^2 = 0.56
DGRP Line
Starvation Resistance (hr)
H^2 = 0.58
DGRP Line

Freeze 2 Molecular Variation in DGRP Lines

Freeze 2 Genotyping

- 205 lines sequenced on Illumina platform
- Initial variant discovery using 7 existing callers
- Global list of variants grouped into haplotype bins
- Reads re-aligned to haplotypes to count supporting and opposing reads
- Final genotypes called by the Joint Genotyper for Inbred Lines (JGIL), which computes a quality score for each genotype and for each variant

The Numbers

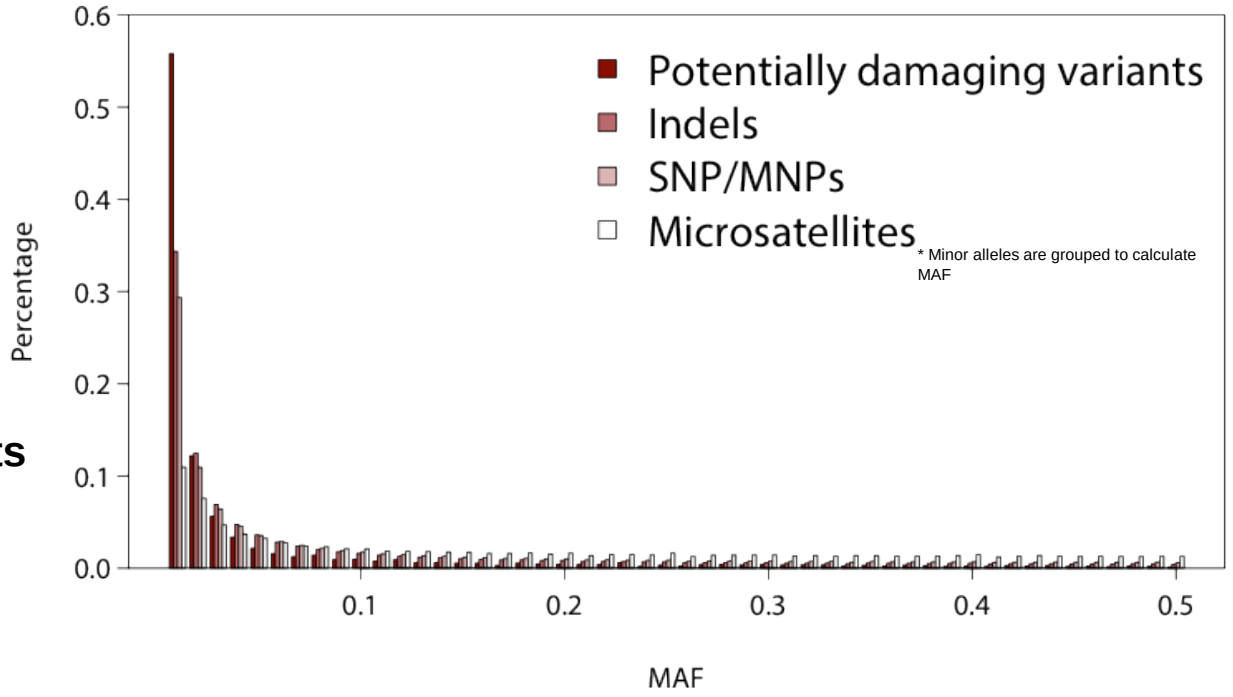
SNP/MNPs
3,976,011

Indels
462,416

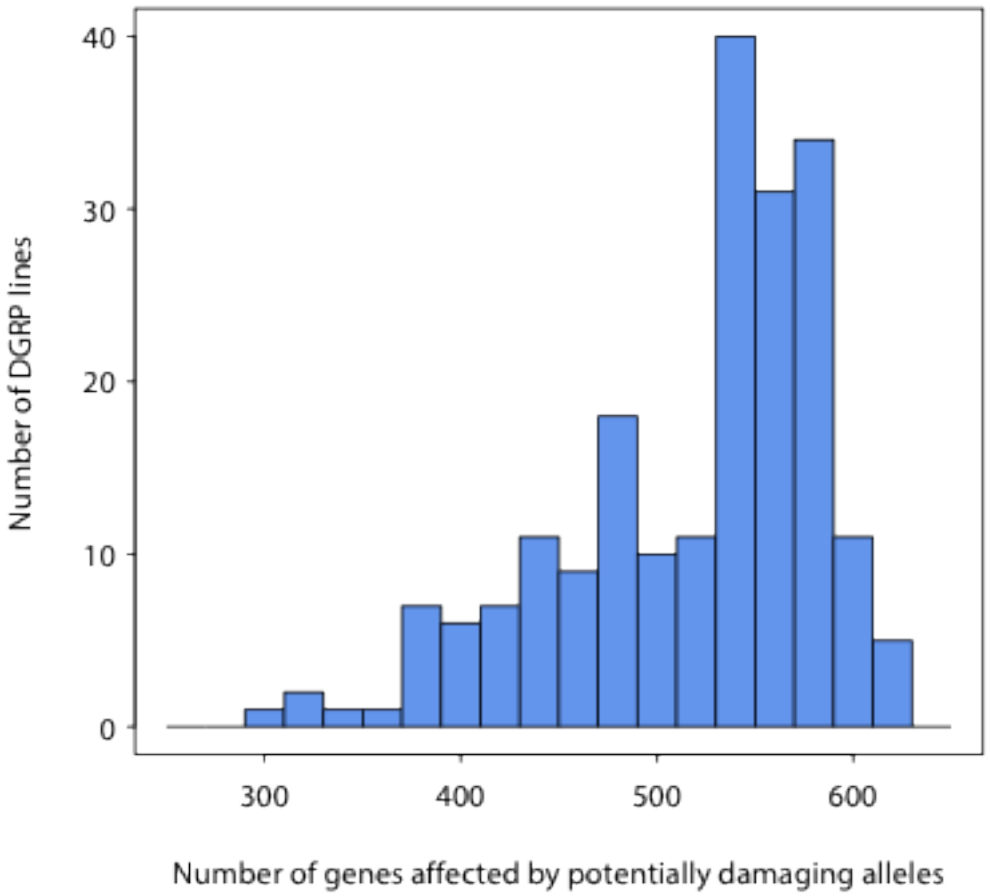
Polymorphic microsatellites
23,754

Potentially damaging variants
6,523

Site Frequency Spectra

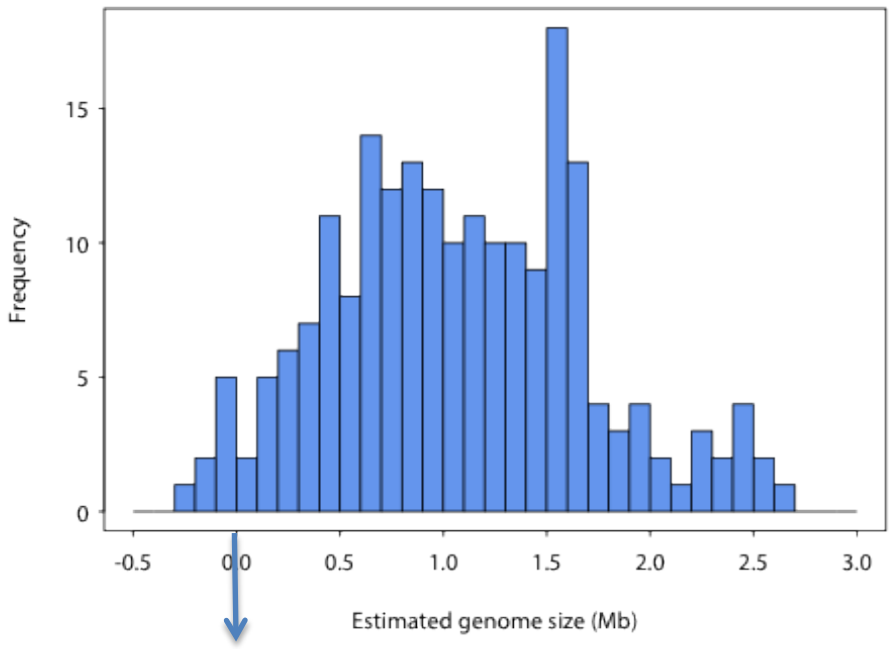


Potentially Damaging Variants in DGRP Lines



- Potentially damaging annotations: SPLICE_SITE, START_LOST, EXON_DELETED, FRAME_SHIFT, STOP_GAINED, STOP_LOST
- Hundreds of genes affected per line
- Large gene families enriched for affected genes (e.g. olfactory receptors)
- Novel alleles for functional analysis

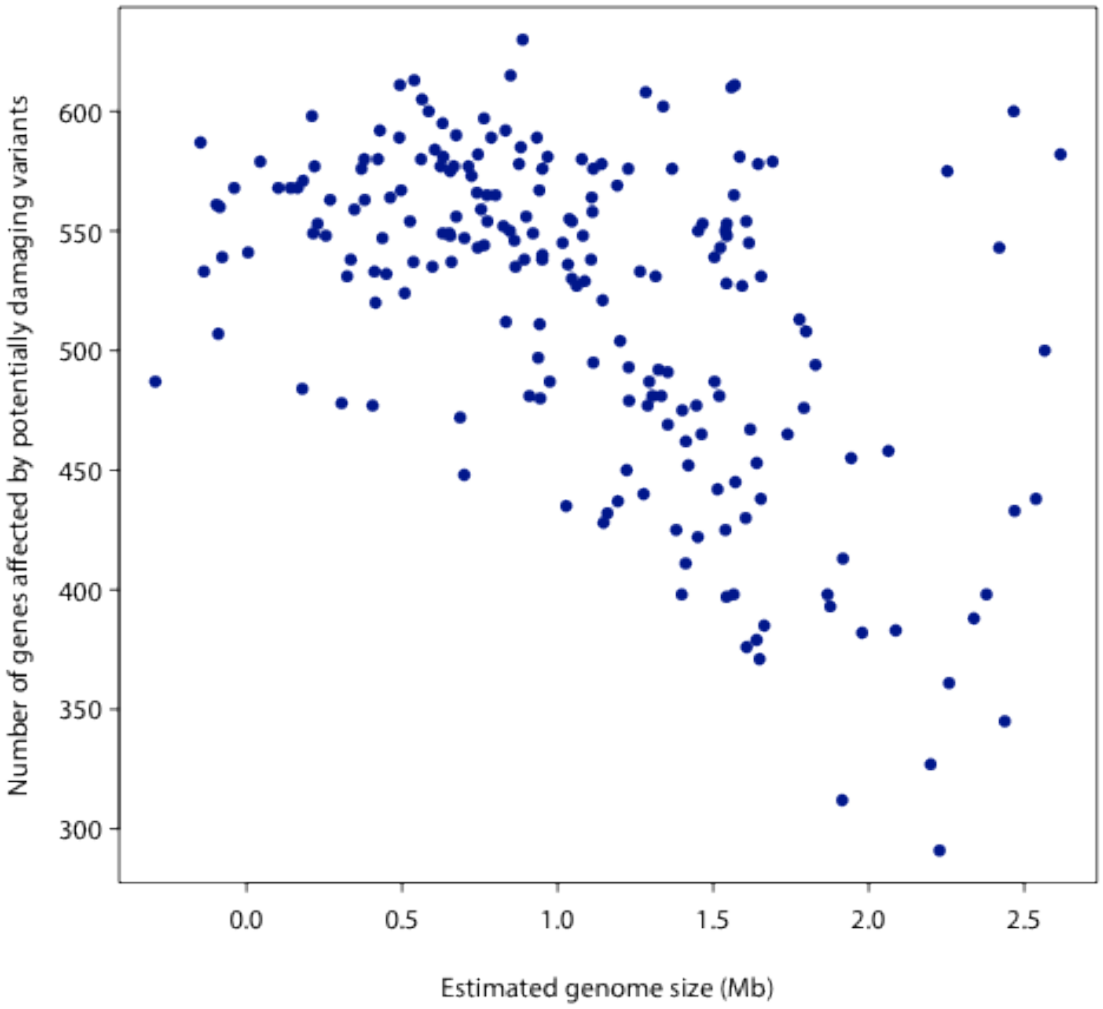
Variation in Genome Size From Indels in DGRP Lines



Reference Strain

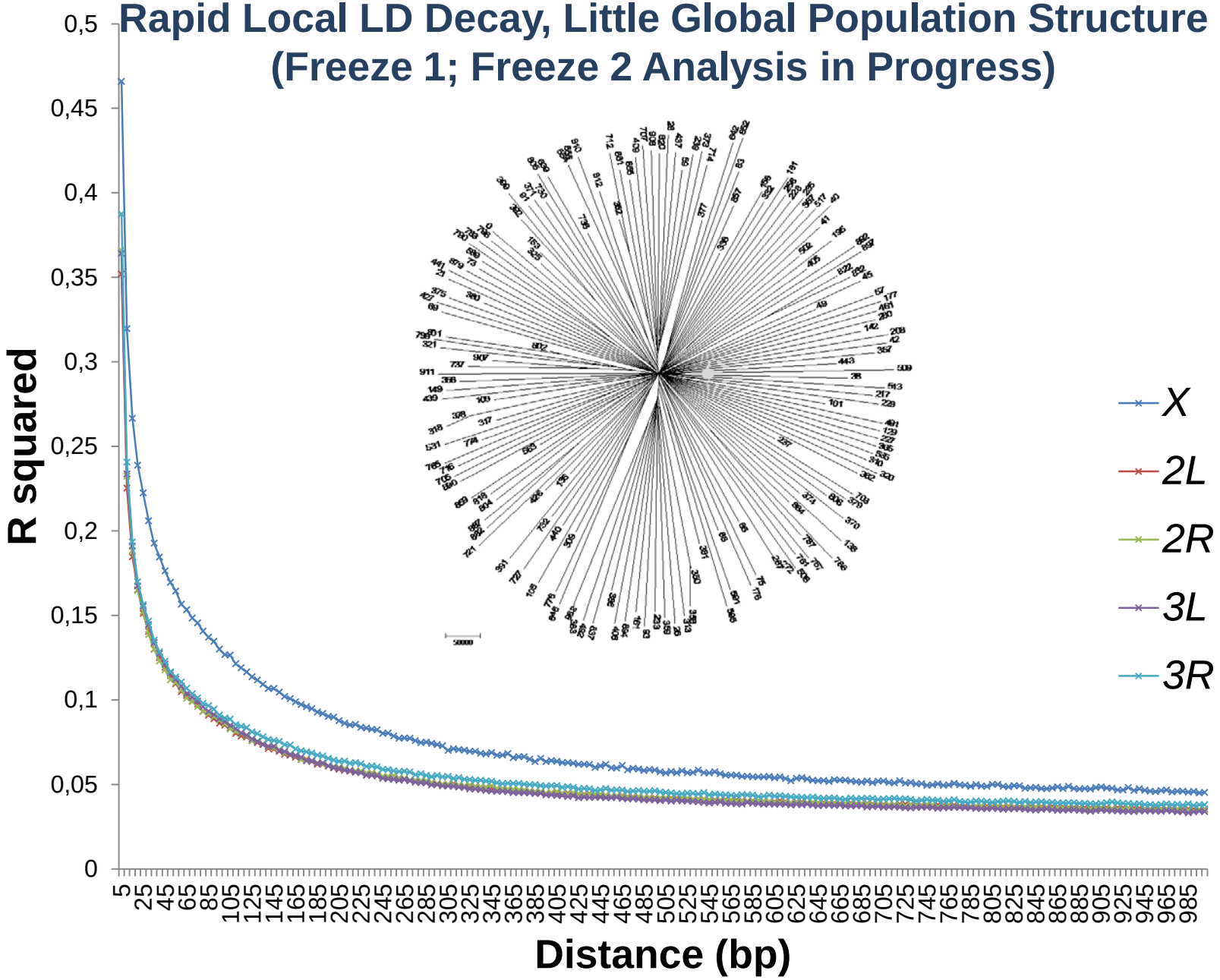
???

Genome Size vs. Damaging Mutations



Deletions (relative to the reference genome) have more damaging effects

Rapid Local LD Decay, Little Global Population Structure (Freeze 1; Freeze 2 Analysis in Progress)



Mackay *et al.* 2012. Nature 482:173

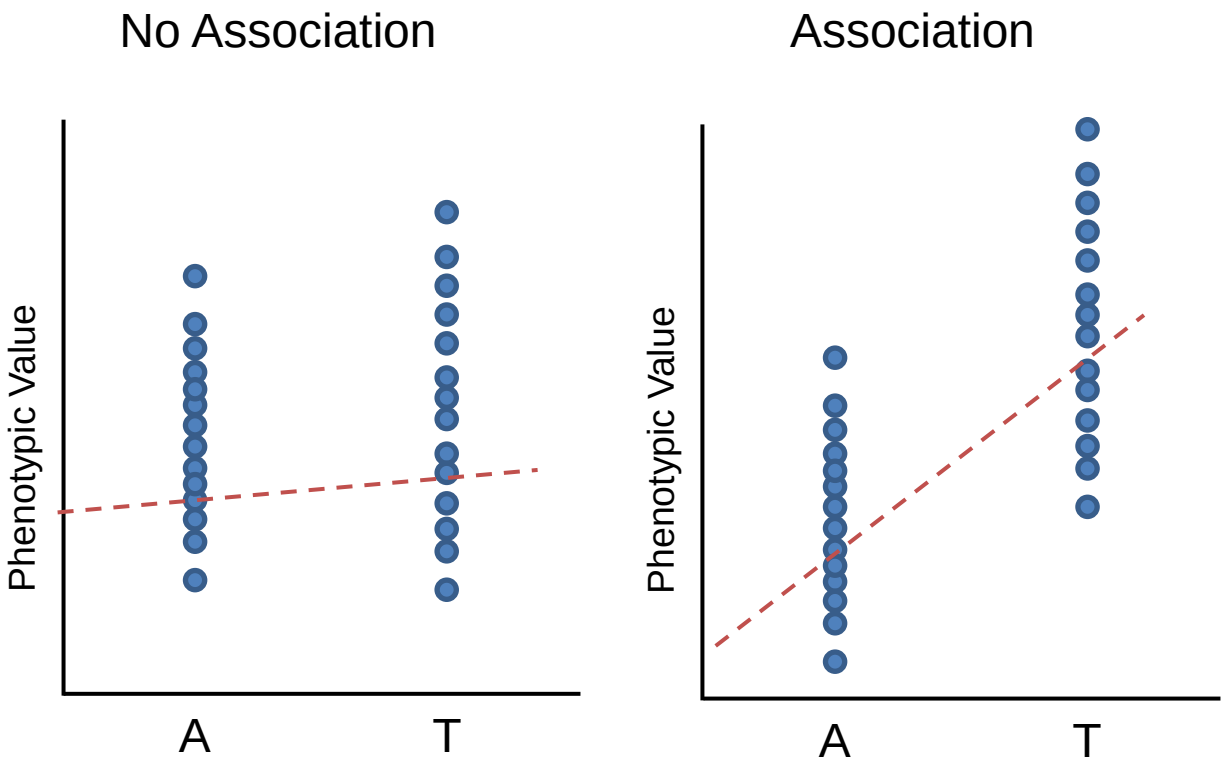
Other Information

- *Wolbachia* infection status of each DGRP line

- Cytogenetic analysis of polymorphic inversions:

In(2L)t
In(2R)Nova Scotia
In(3L)Payne
In(3R) Mourad
In(3R)Payne
In(3R)Kodani
In(3R)C

Association Analyses



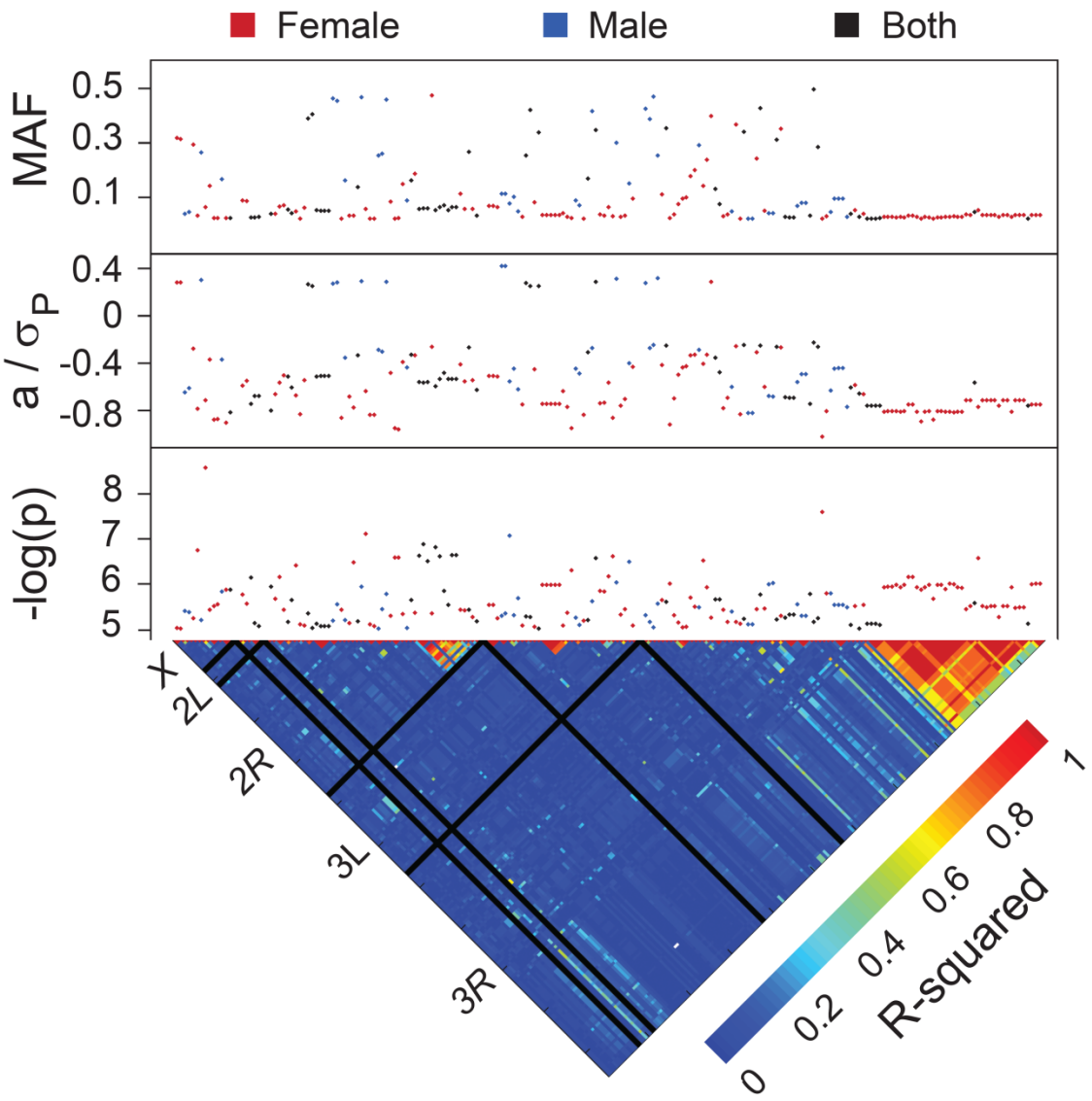
http://dgrp.gnets.ncsu.edu/

Faster GWAS

GWAS includes relationship matrix

Latest FlyBase annotations of variants

GWAS in the DGRP: Starvation Resistance



Most significant SNPs are low frequency

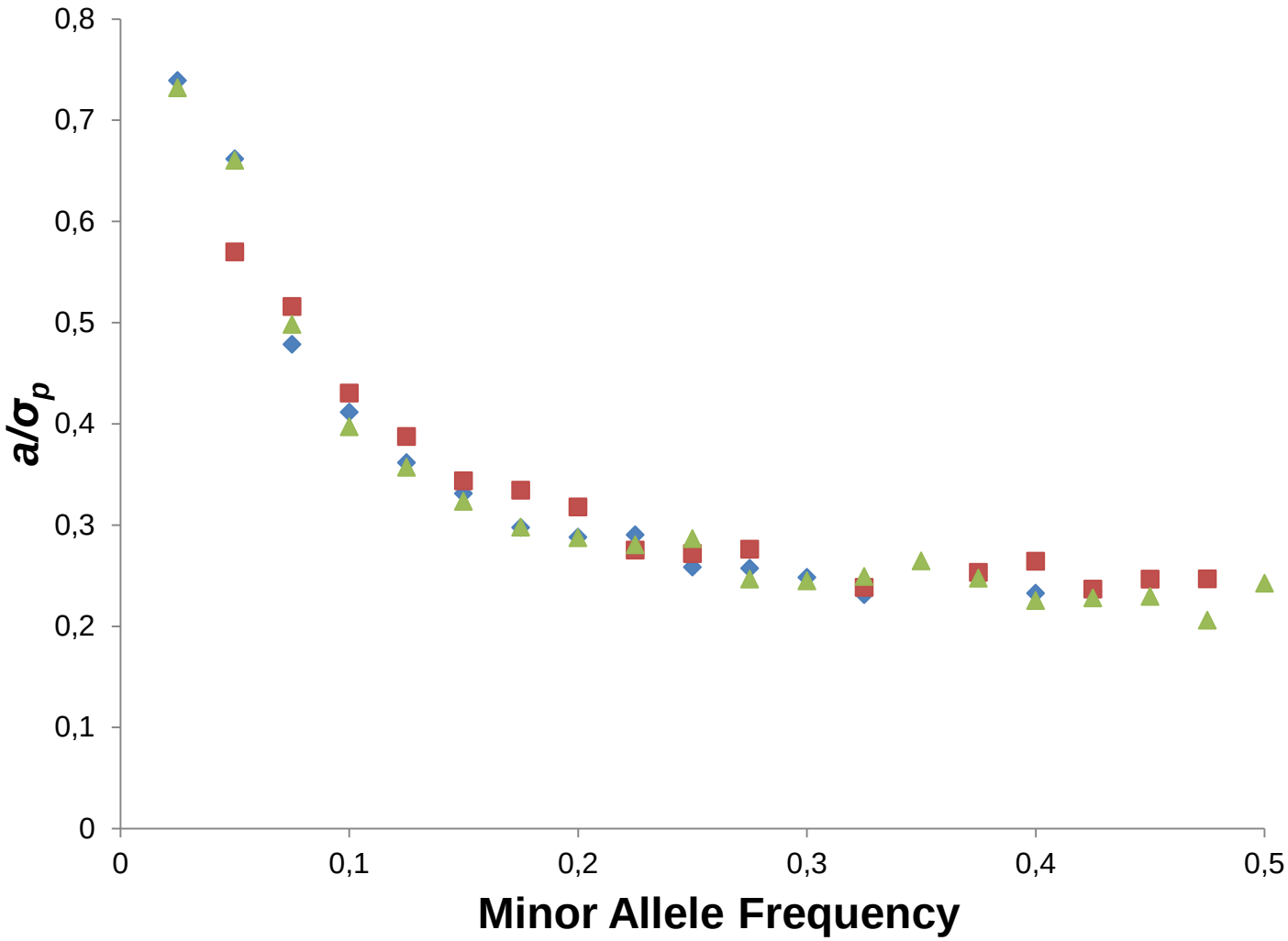
Standardized effects not small

Many significant SNPs (FDR < 0.1)
Many sex-specific effects
Novel genes

Local LD

Mackay *et al.* 2012. Nature 482:173

Relationship Between Effect Size and MAF

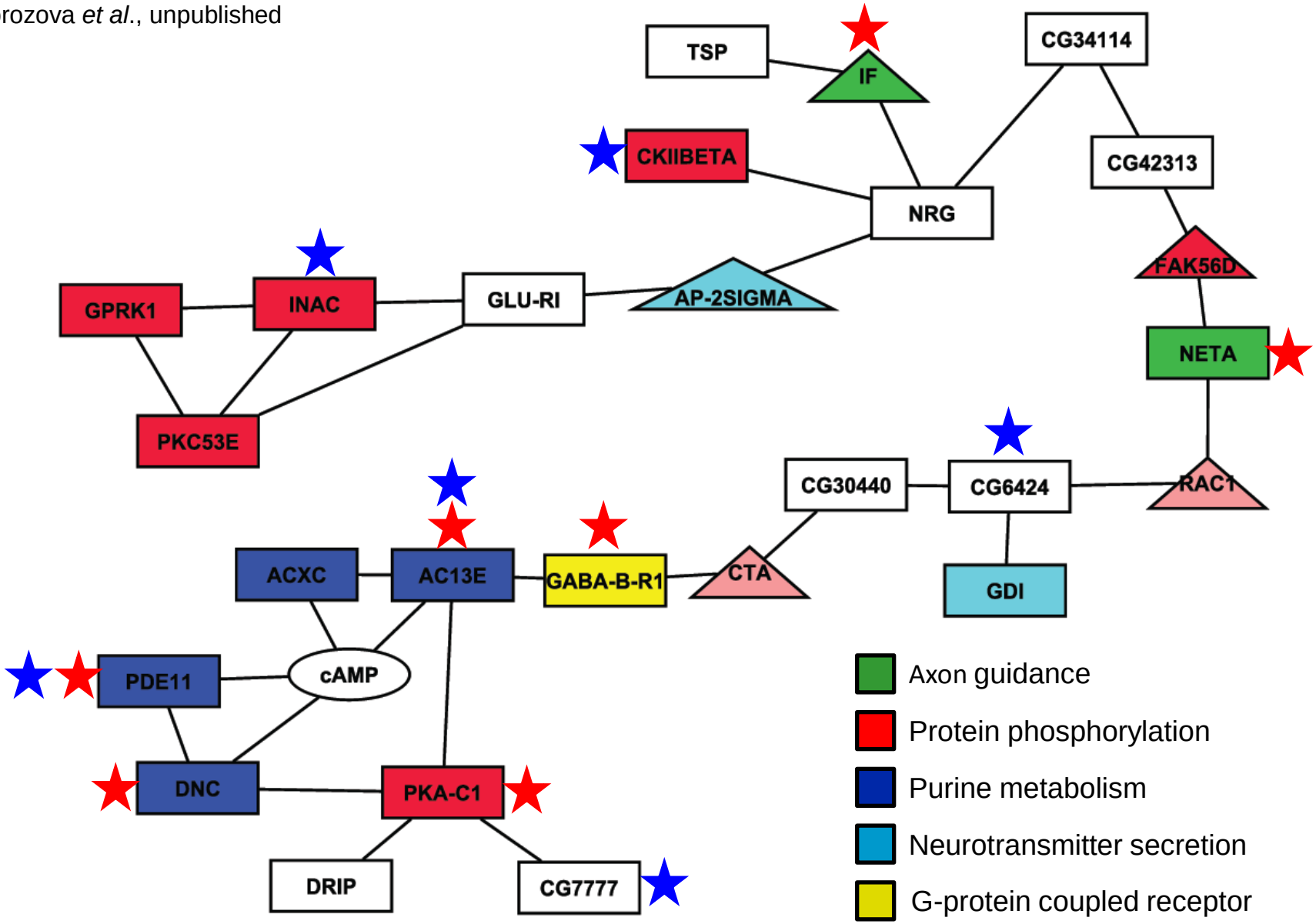


Mackay *et al.* 2012. Nature 482:173

R_Spider Analysis of Ethanol Sensitivity GWAS

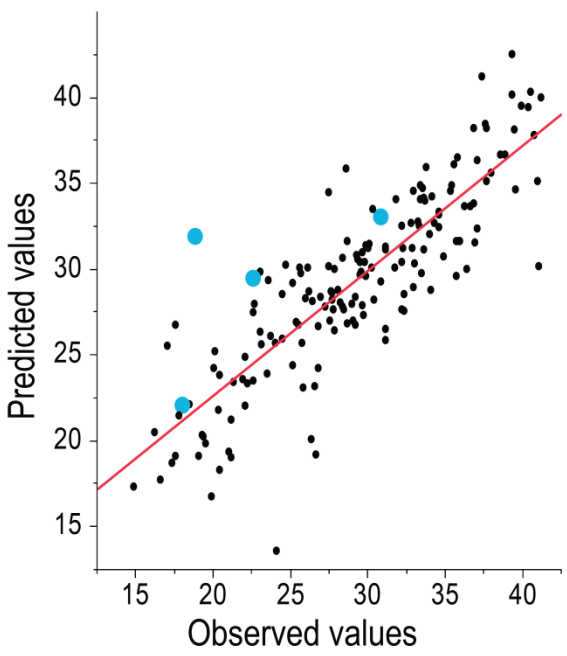
www.BioProfiling.de

TV Morozova *et al*., unpublished

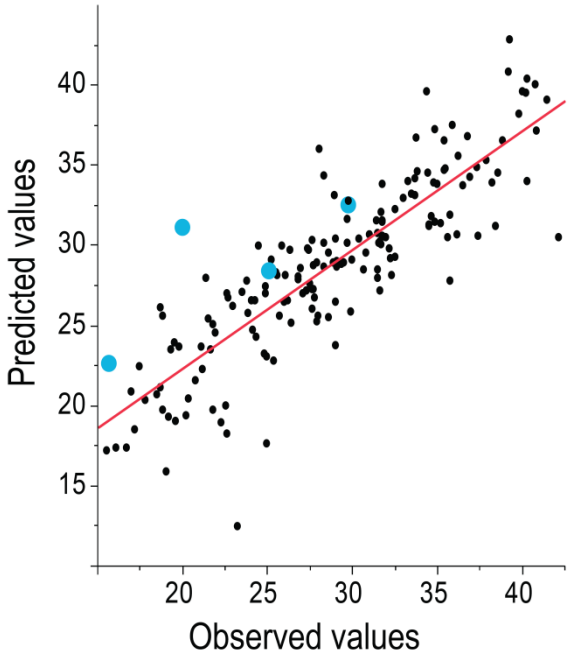


Additive Multi-SNP Model Explains Most Variation

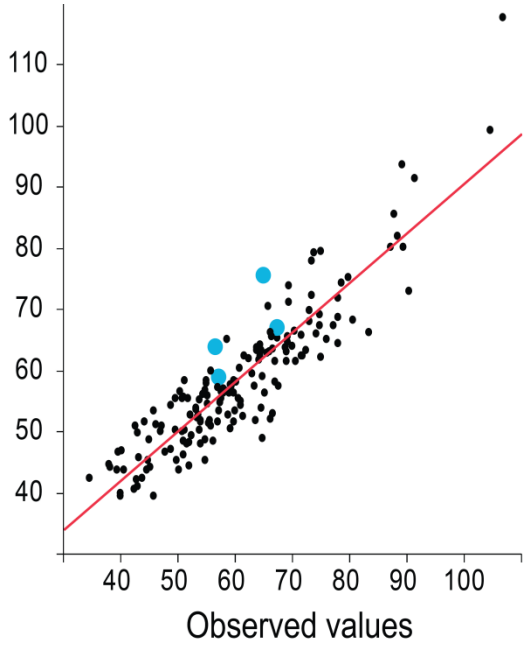
Startle Response
$r^2 = 0.72$



Chill Coma Recovery
$r^2 = 0.84$



Starvation Resistance
$r^2 = 0.81$



Mackay *et al.* 2012. Nature 482:173

But Predictive Ability is Low

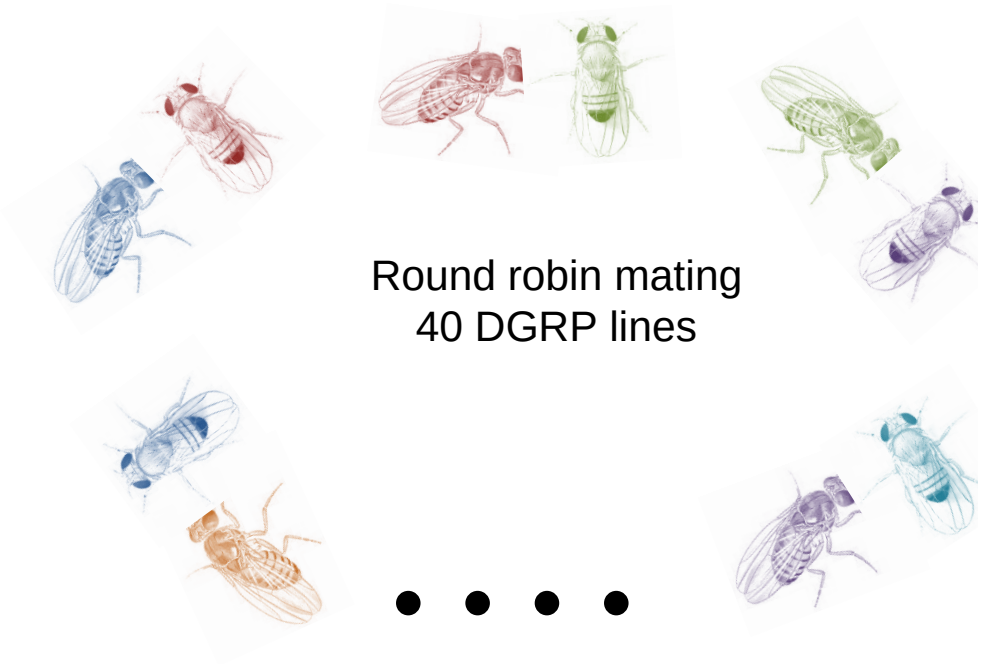
Trait	Broad Sense H^2	Predictive Ability*
Starvation Resistance	0.58	0.239 ± 0.008
Startle Response	0.56	0.230 ± 0.012
Chill Coma Recovery	0.43	-0.038 ± 0.008

*GBLUP, 5-fold cross validation; ~2.5 million common SNPs

- Small sample size?
- Rare alleles?
- Alleles with large effects?
- Non-additive genetic variation?

Ober *et al.* 2012. PLoS Genetics 8: e1002685

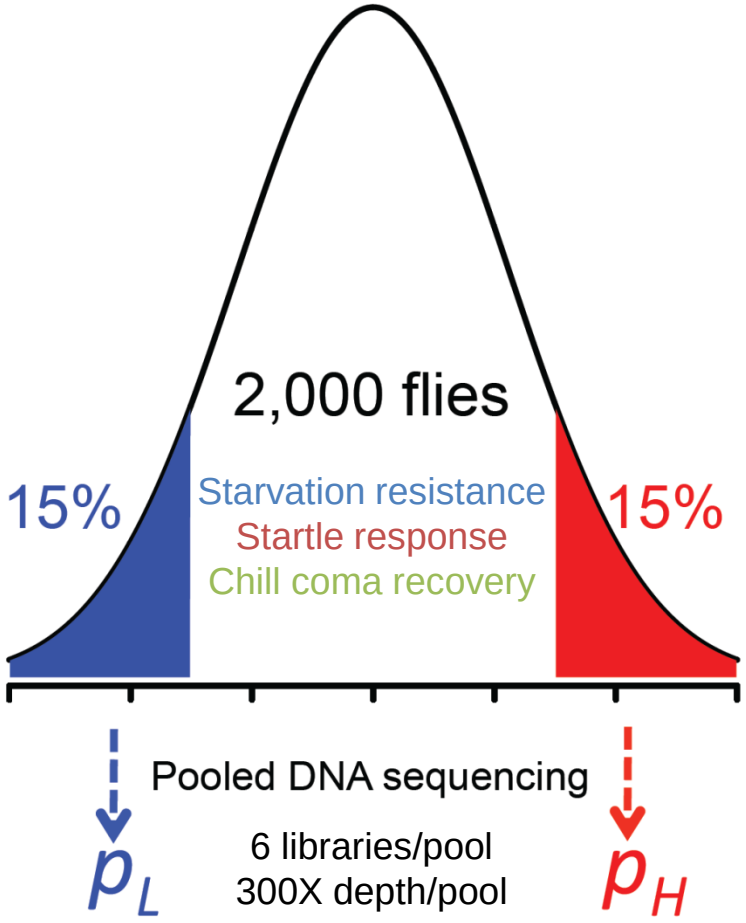
Testing Multi-Locus Predictions: Flyland



Round robin mating
40 DGRP lines

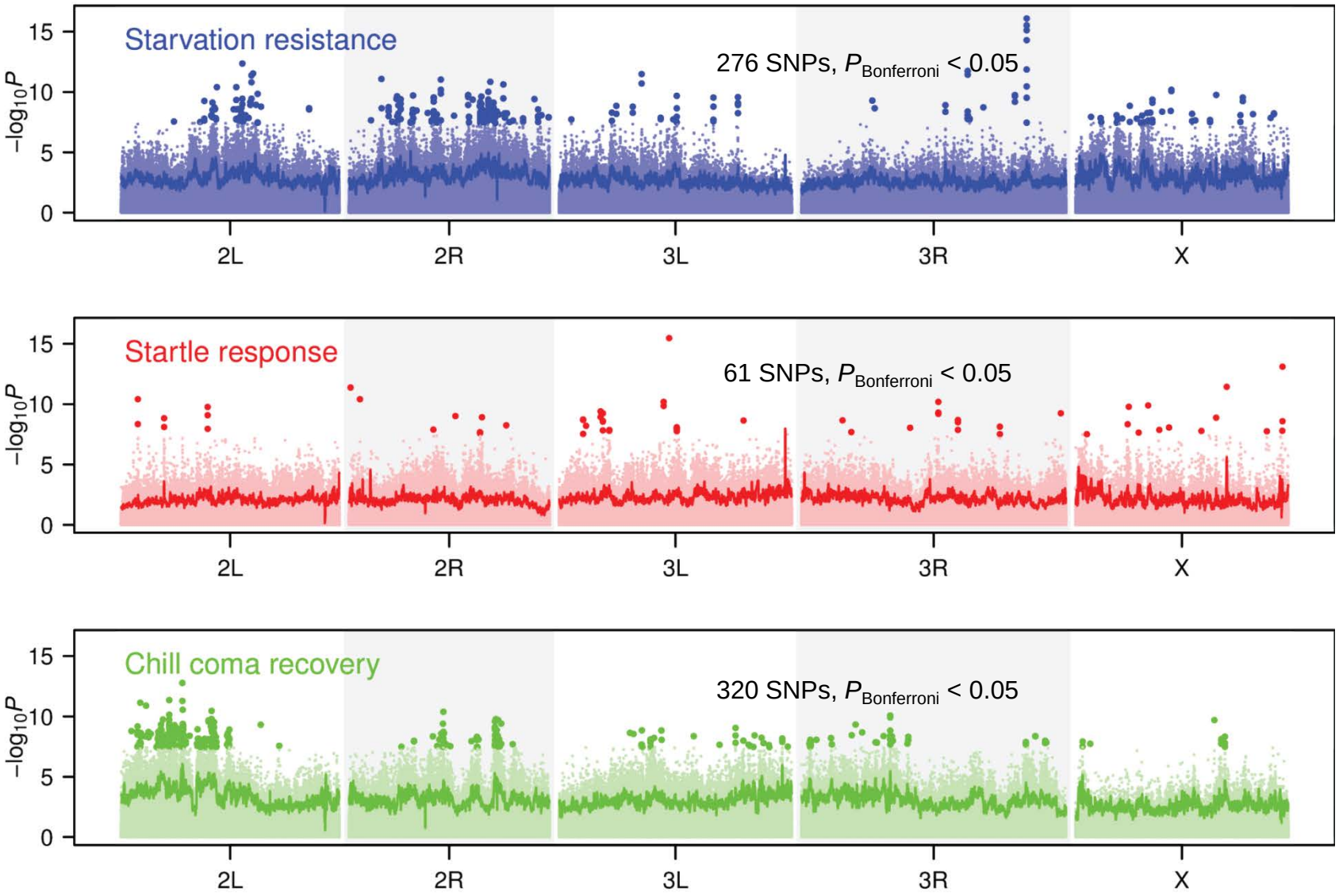
Random mating
70 generations
N= 800/generation

Flyland Population



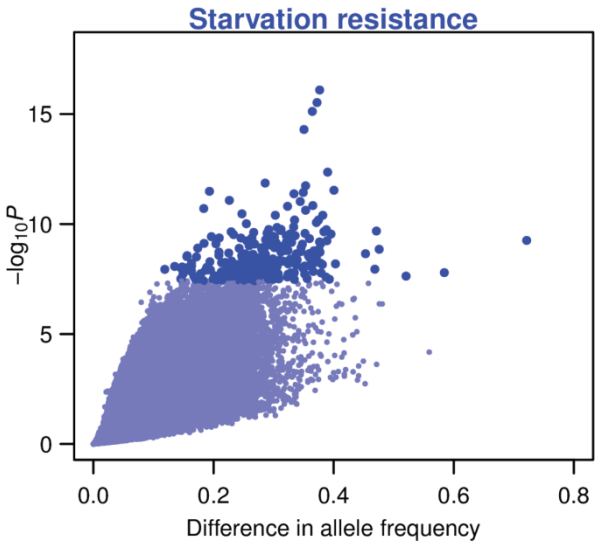
Huang *et al.*, PNAS 109: 15553

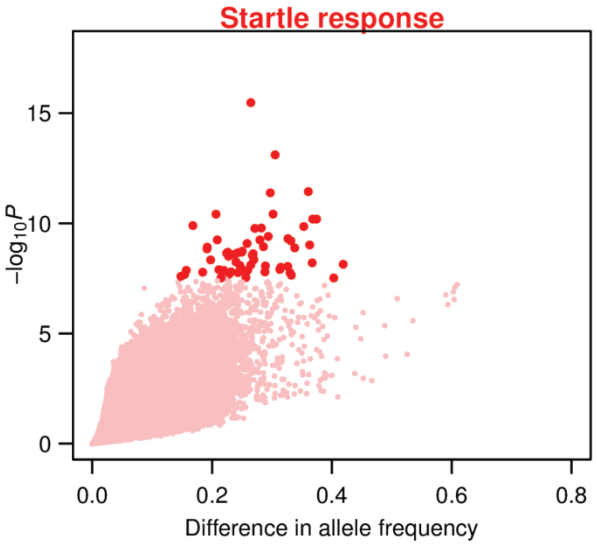
Complex Genetic Architecture in Flyland

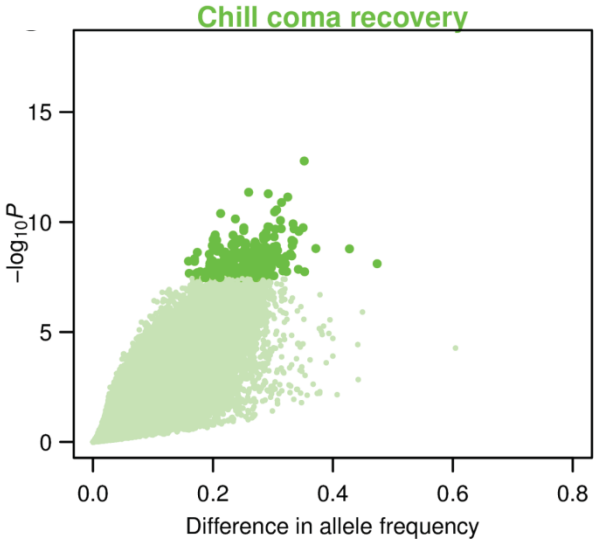


Huang *et al.*, PNAS 109: 15553

SNP Effects Not Small

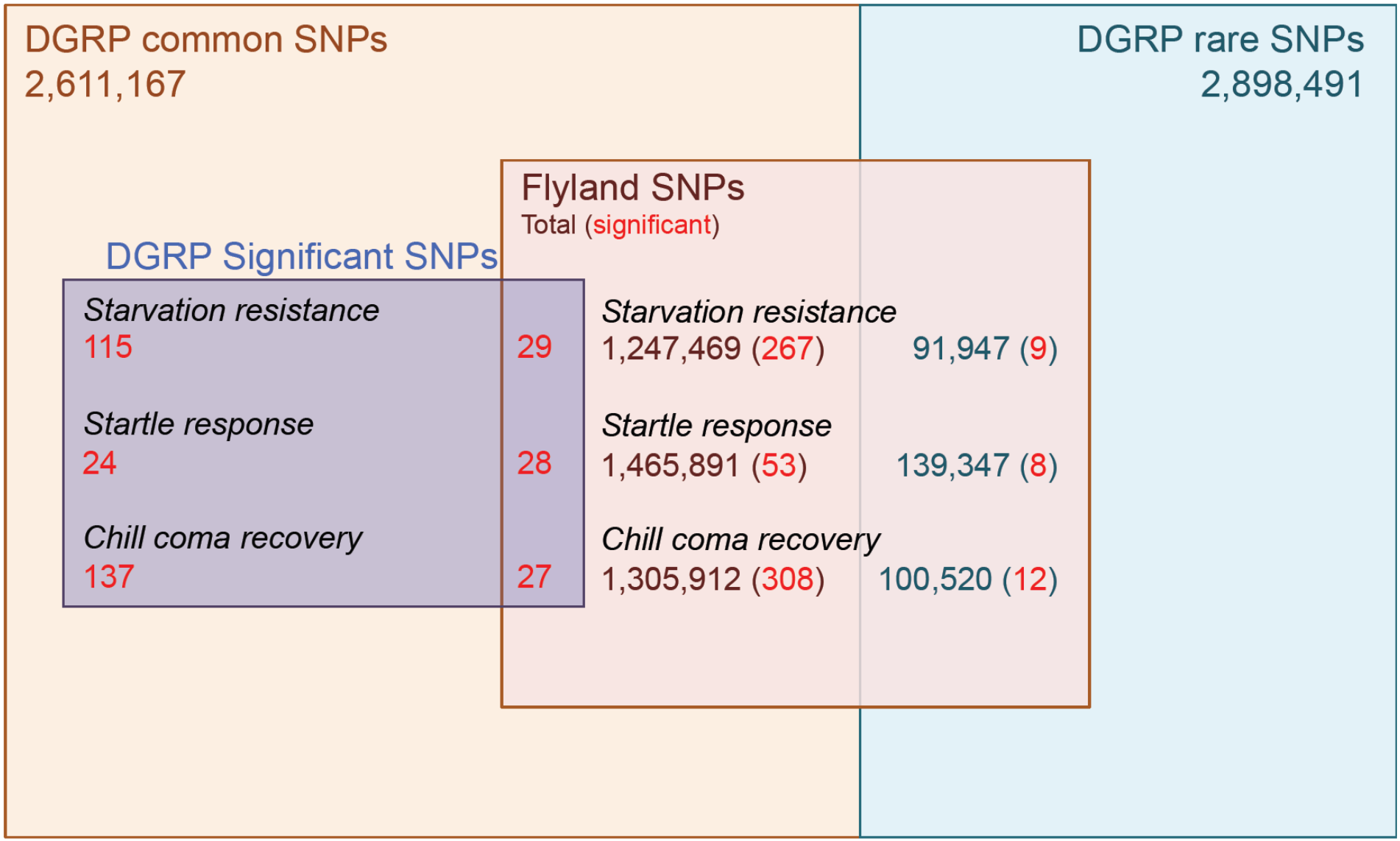






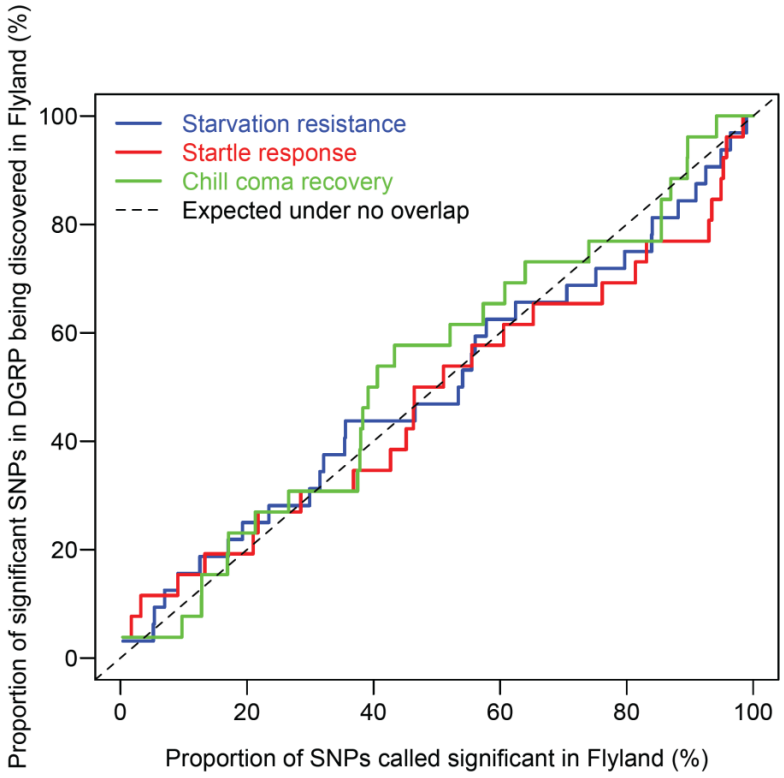
Huang *et al.*, PNAS 109: 15553

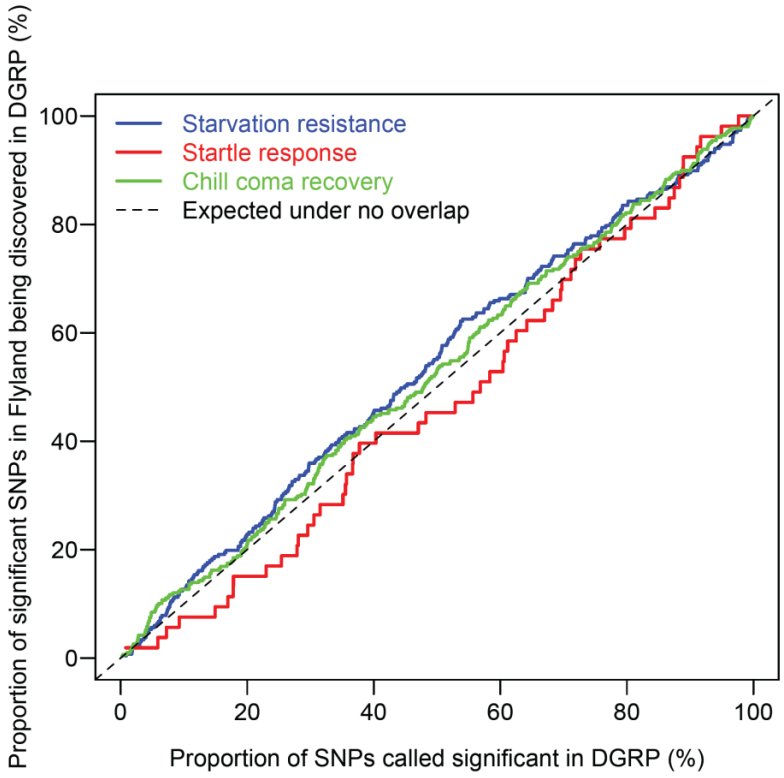
No Replication between Flyland and DGRP



Huang *et al.*, PNAS 109: 15553

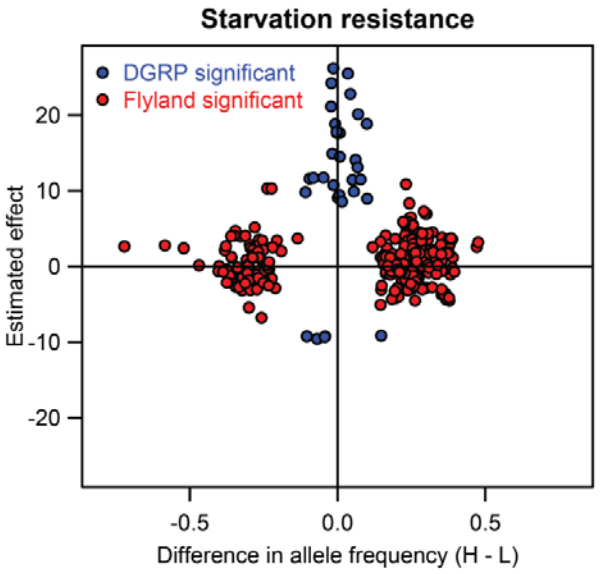
No Replication Between Flyland and DGRP:
Not From Different *P*-value Thresholds

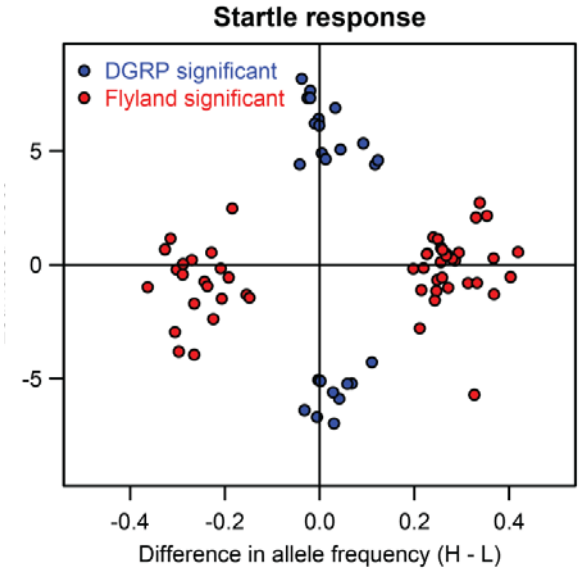


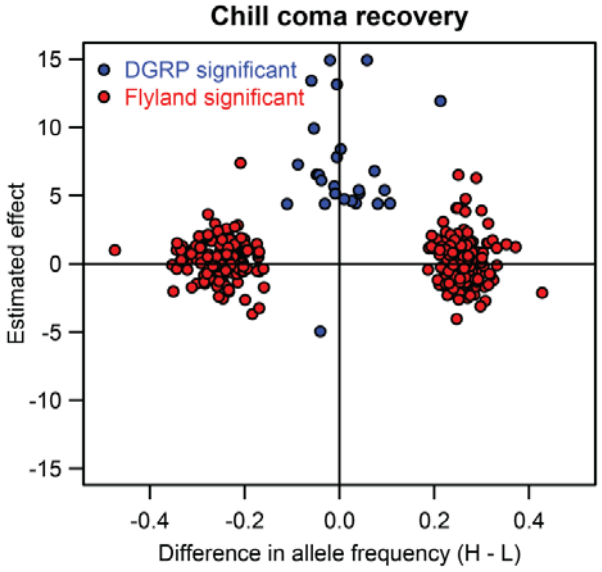


Huang *et al.*, PNAS 109: 15553

No Replication between Flyland and DGRP:
Effects Not Correlated

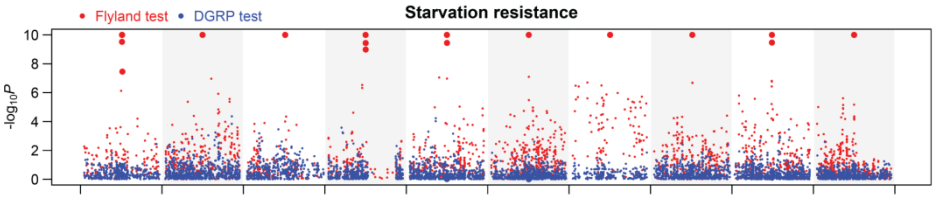


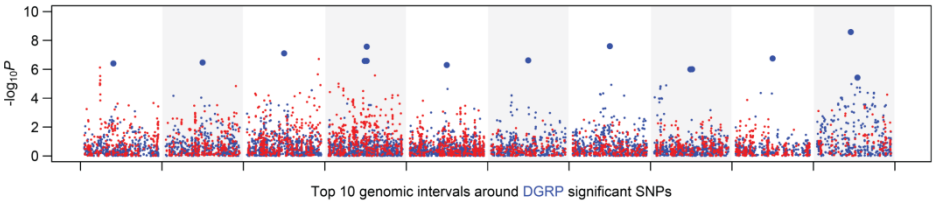


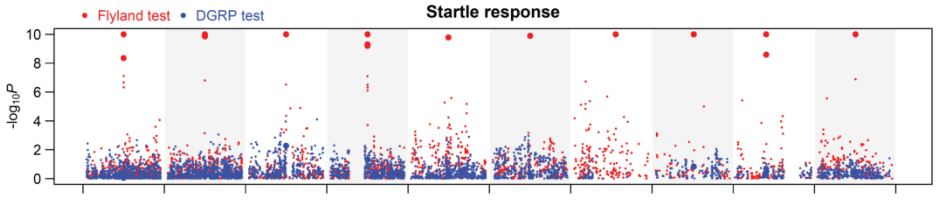


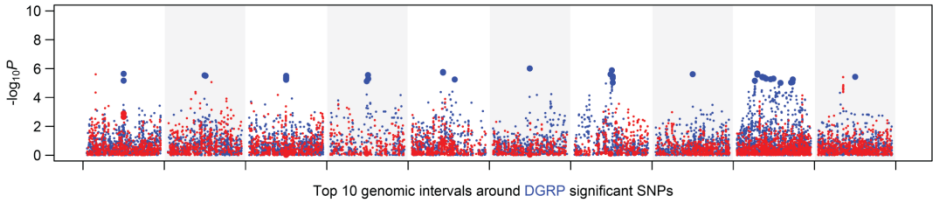
Huang *et al.*, PNAS 109: 15553

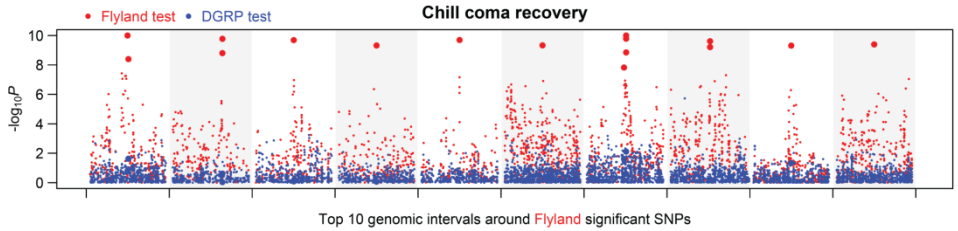
No Replication between Flyland and DGRP:
SNPs not in LD

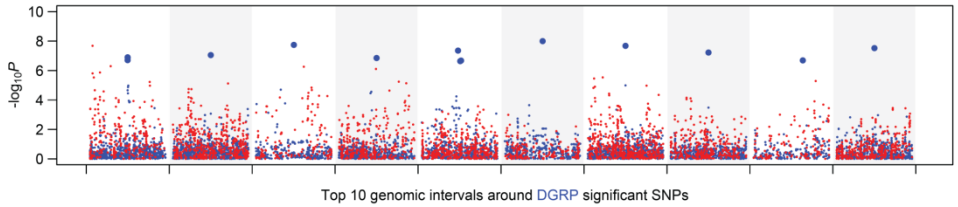












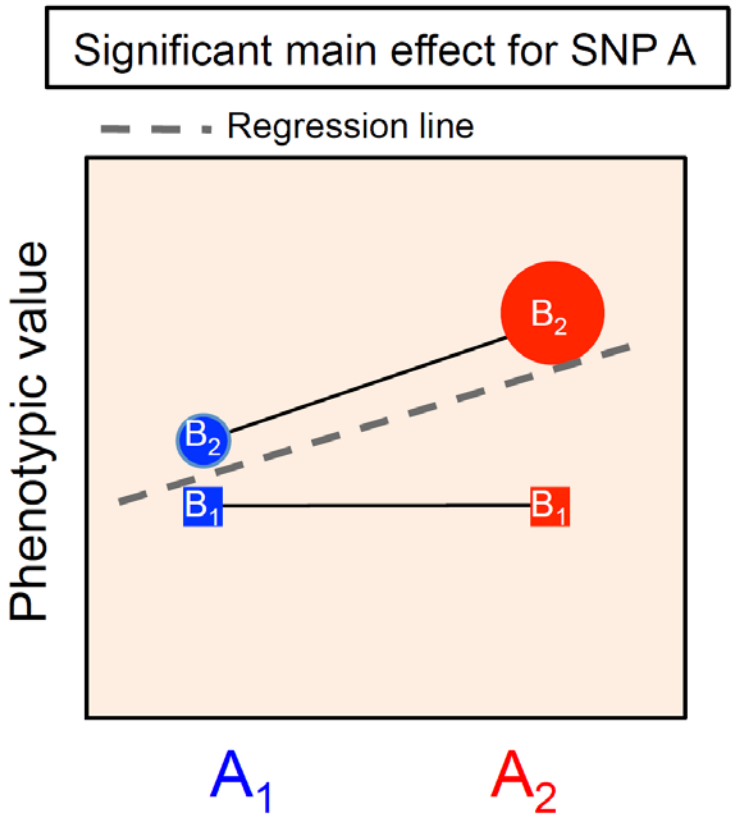
Huang *et al.*, PNAS 109: 15553

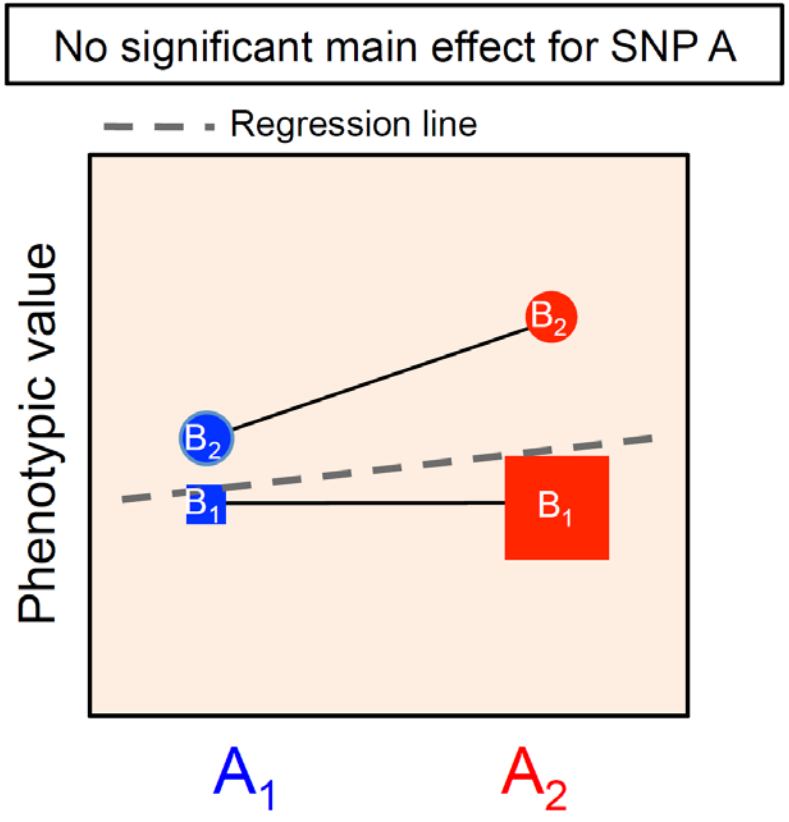
No Replication between Flyland and DGRP: Epistasis

Trait	Trait-associated common SNPs (DGRP + Flyland)	Significant epistatic pairs	Trait-associated SNPs with at least one significant interacting SNP (Flyland + DGRP)
Starvation resistance	411 (144 + 267)	19,629	288 (57 + 231)
Startle response	105 (52 + 53)	1,252	48 (11 + 37)
Chill coma recovery	472 (164 + 308)	57,863	314 (79 + 235)

Huang *et al.*, PNAS 109: 15553

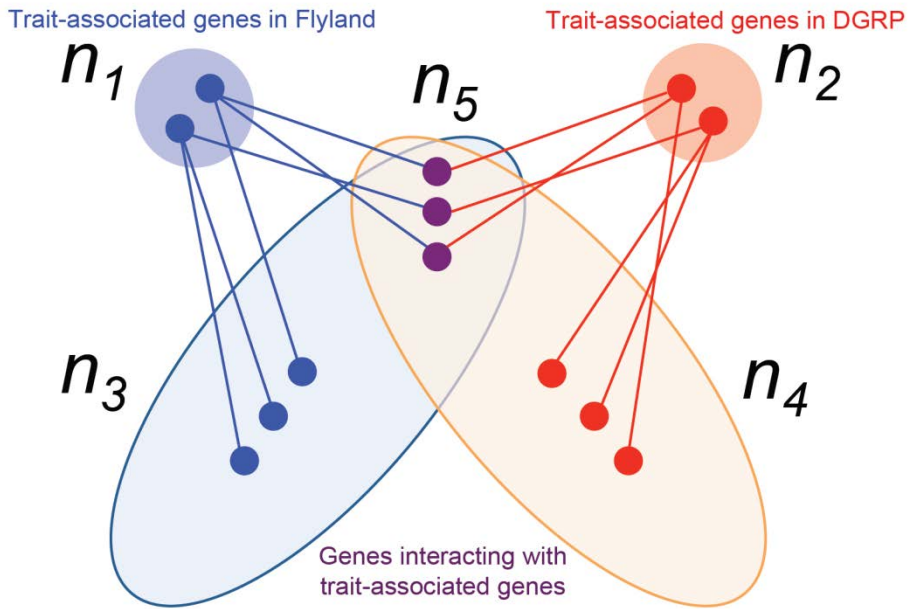
Context Dependence Due to Differences in Allele Frequency





Huang *et al.*, PNAS 109: 15553

Epistatic Networks Reveal Common Interacting Genes



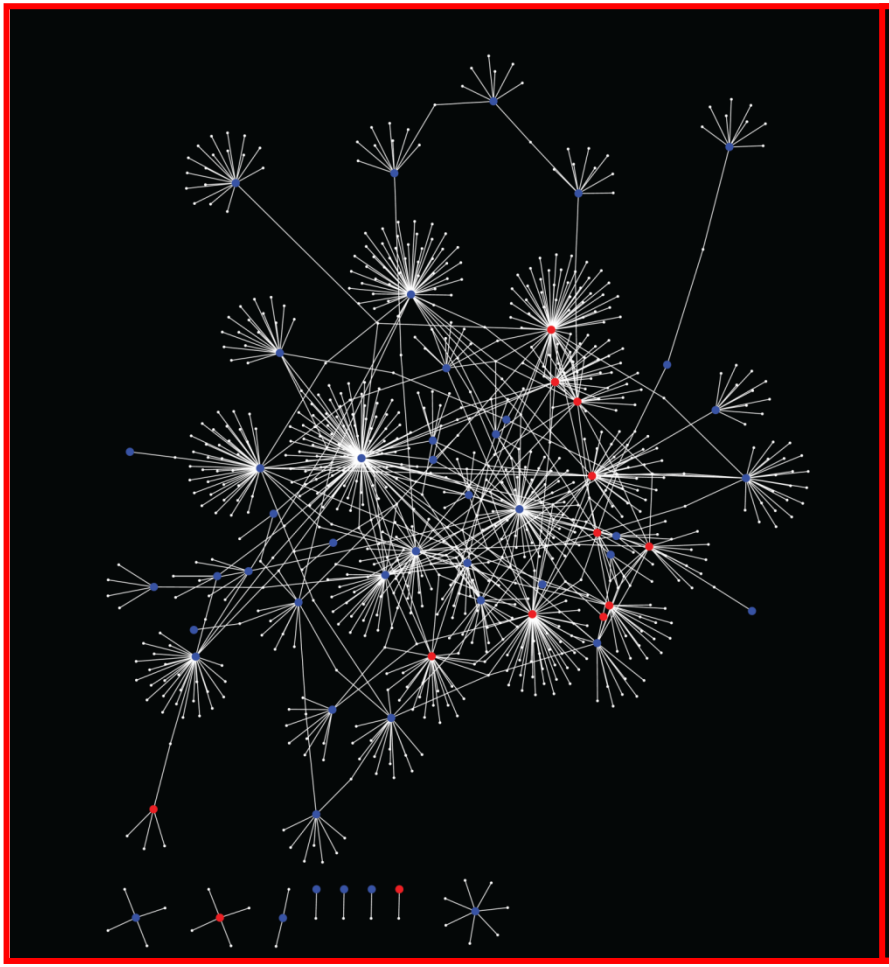
Trait	Trait-associated genes in Flyland (n_1)	Trait-associated genes in DGRP (n_2)	Genes interacting with trait-associated genes in Flyland (n_3)	Genes interacting with trait-associated genes in DGRP (n_4)	Genes interacting with trait-associated genes in both Flyland and DGRP (n_5)	Permutation *P* value for n_5
Starvation resistance	59	20	682	295	71	< 0.001
Startle response	6	3	53	5	1	0.021
Chill coma recovery	75	17	1723	659	181	< 0.001

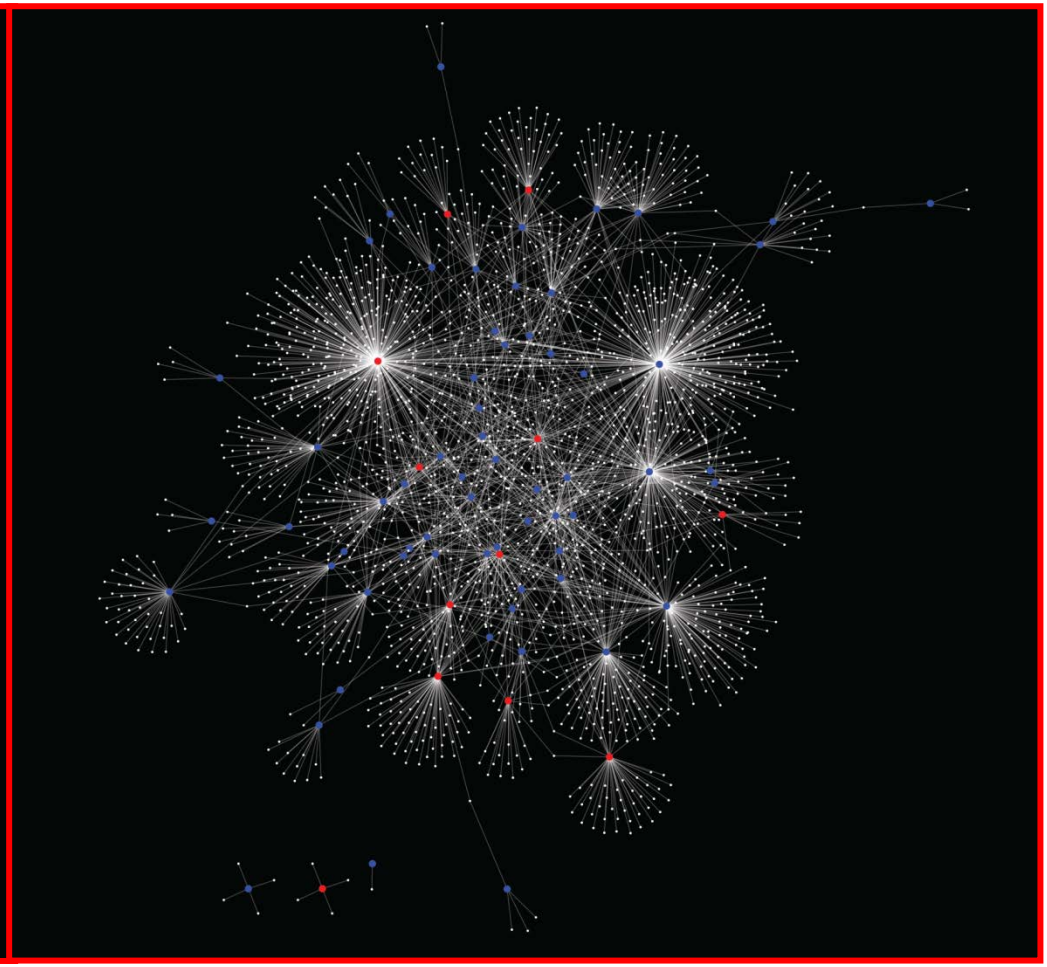
Huang *et al.*, PNAS 109: 15553

Interconnected Networks of Common Interacting Genes

Starvation Resistance

Chill Coma Recovery





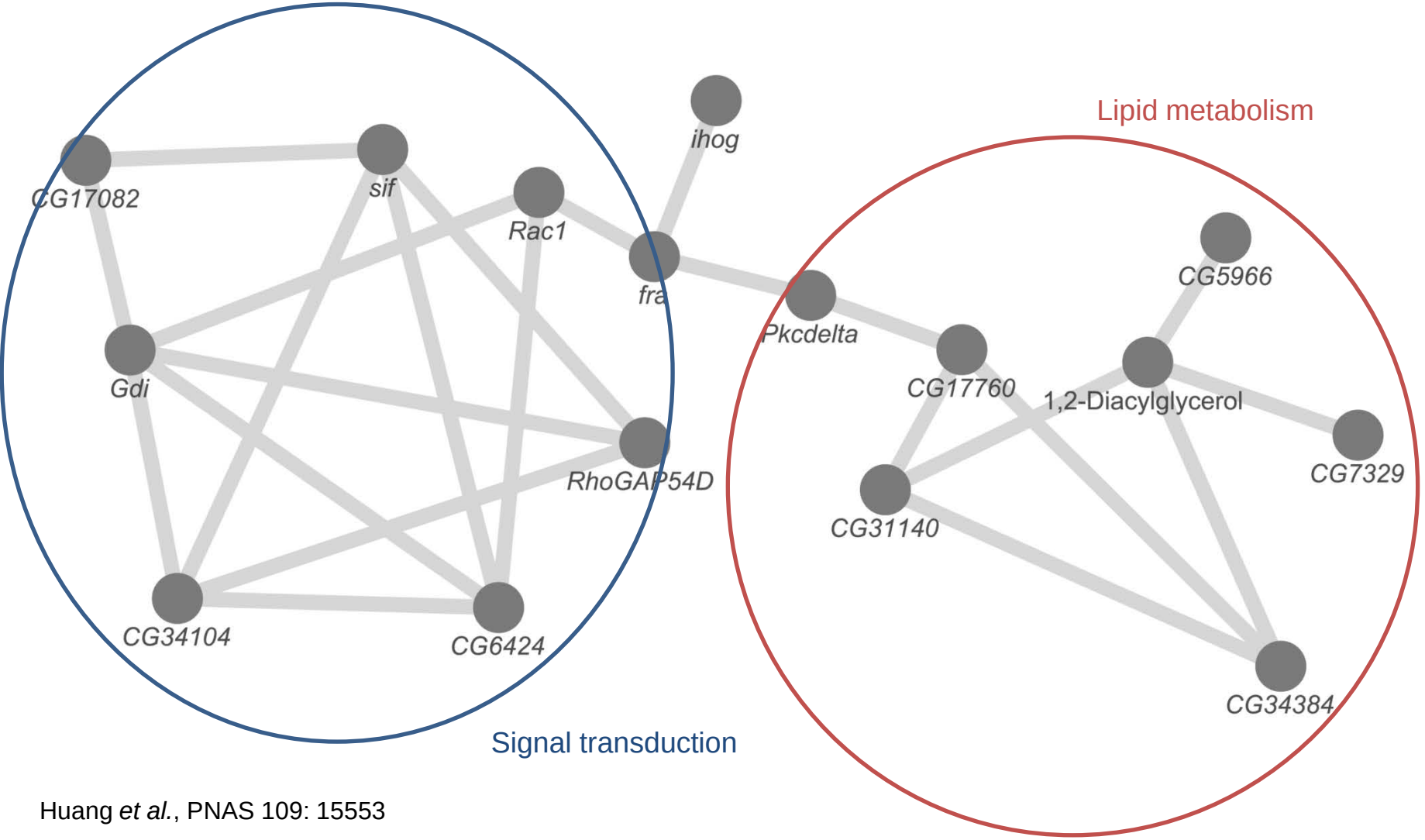
SNPs context-dependent, network architecture conserved

Huang *et al.*, PNAS 109: 15553

Epistatic Interaction Networks Enriched for Known Genetic Interactions

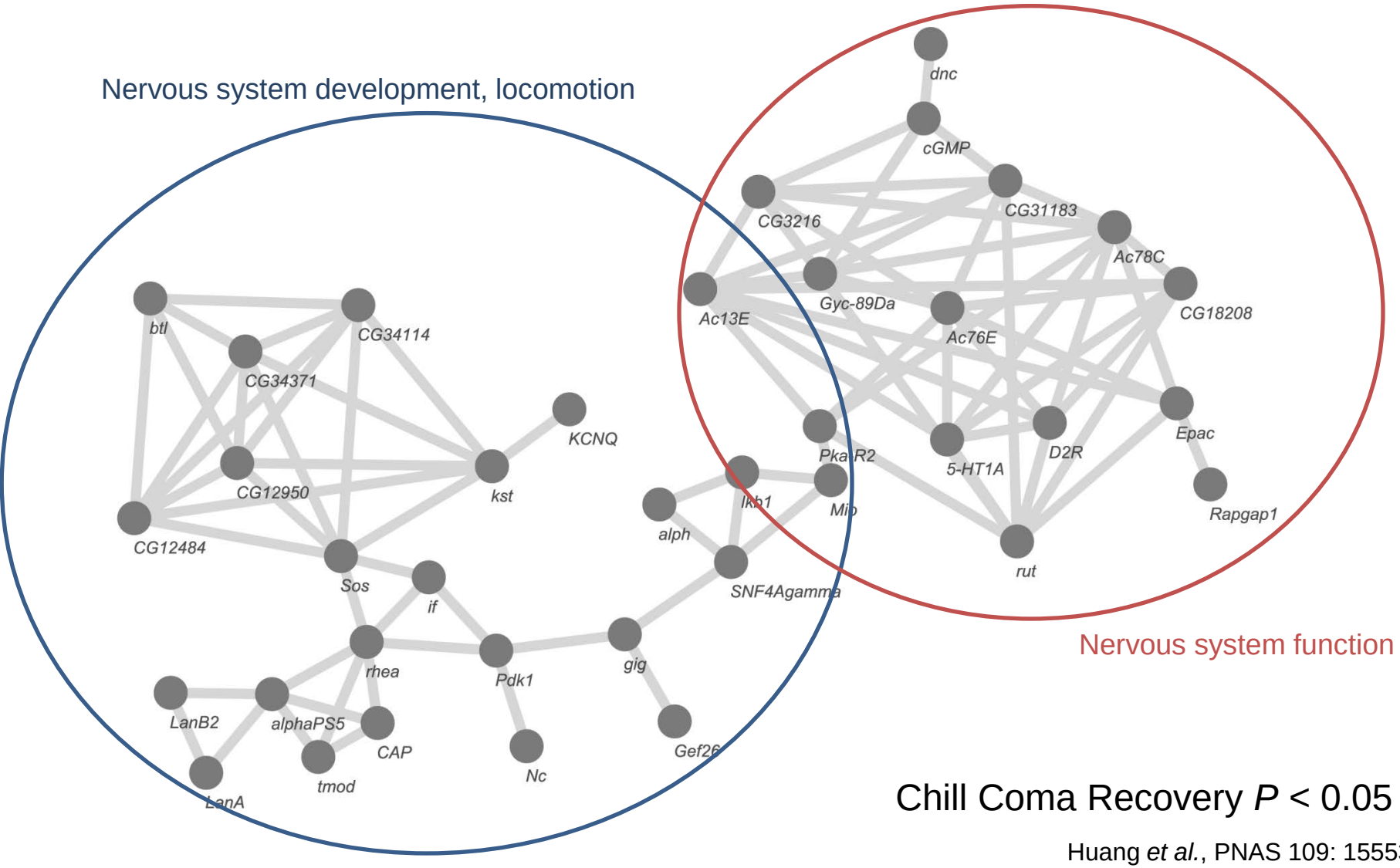
www.BioProfiling.de

Starvation Resistance $P < 0.005$



Huang *et al.*, PNAS 109: 15553

Epistatic Interaction Networks Enriched for Known Genetic Interactions
www.BioProfiling.de



Chill Coma Recovery $P < 0.05$

Huang *et al.*, PNAS 109: 15553

General Strategy

1. Test for interactions between SNPs that do not replicate in different populations, reciprocally in each population

2. Within a population: pairwise search for epistasis using GWAS SNPs as focal SNPs

3. Within a population: Identify SNPs with different allele frequencies between extremes of quantitative trait distribution; use these as focal SNPs for pairwise GWAS

A Paradigm Shift?

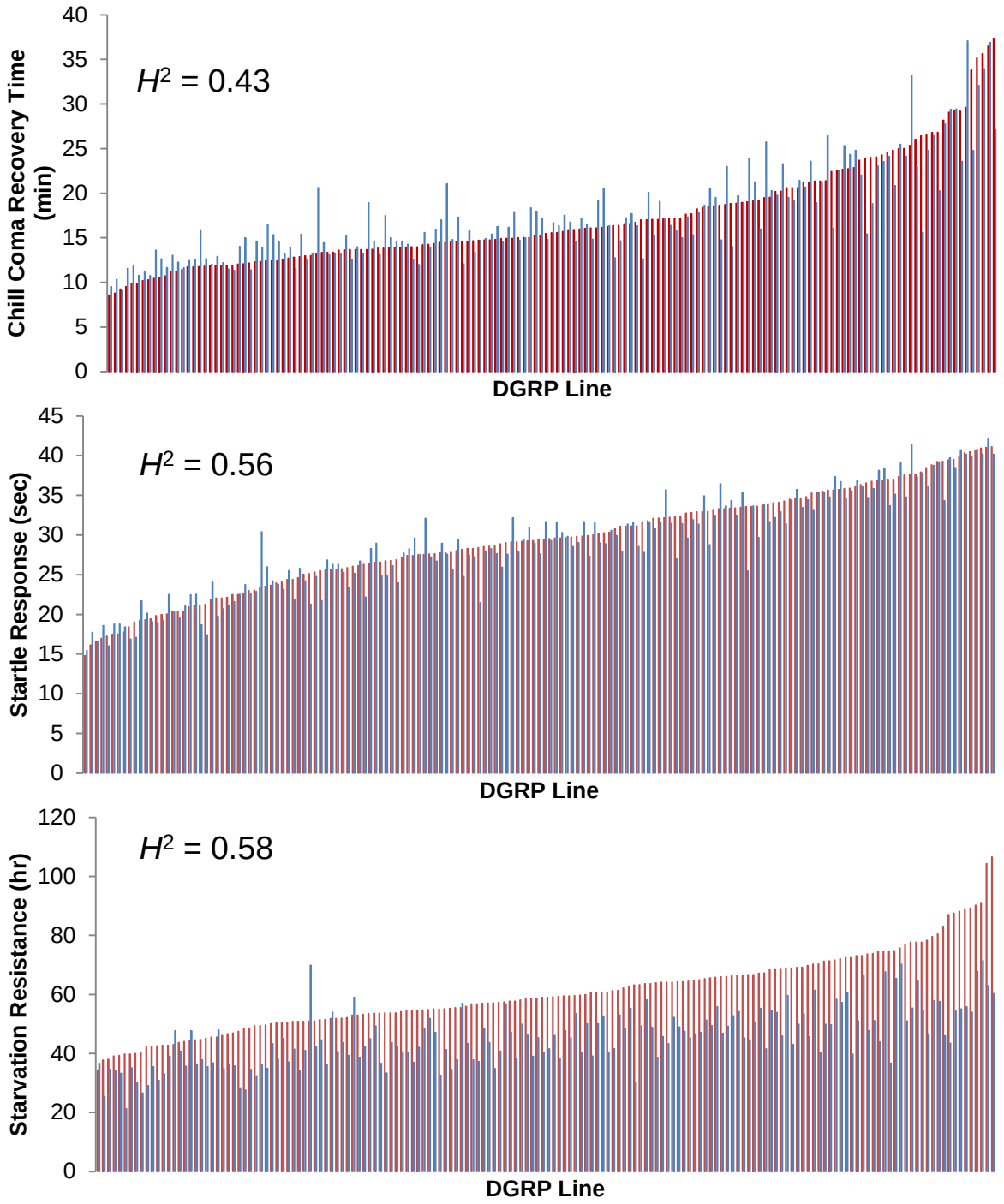
Is this a lot of variation?

Or too little given molecular diversity?
(~4 Million SNPs
+ non-SNP variants)

Additive SNP effects may be underestimated if they interact epistatically but the model does not include epistasis

Has the genome evolved to suppress effects of deleterious mutations?

Epistatic interactions = genetic basis of canalization (homeostasis)

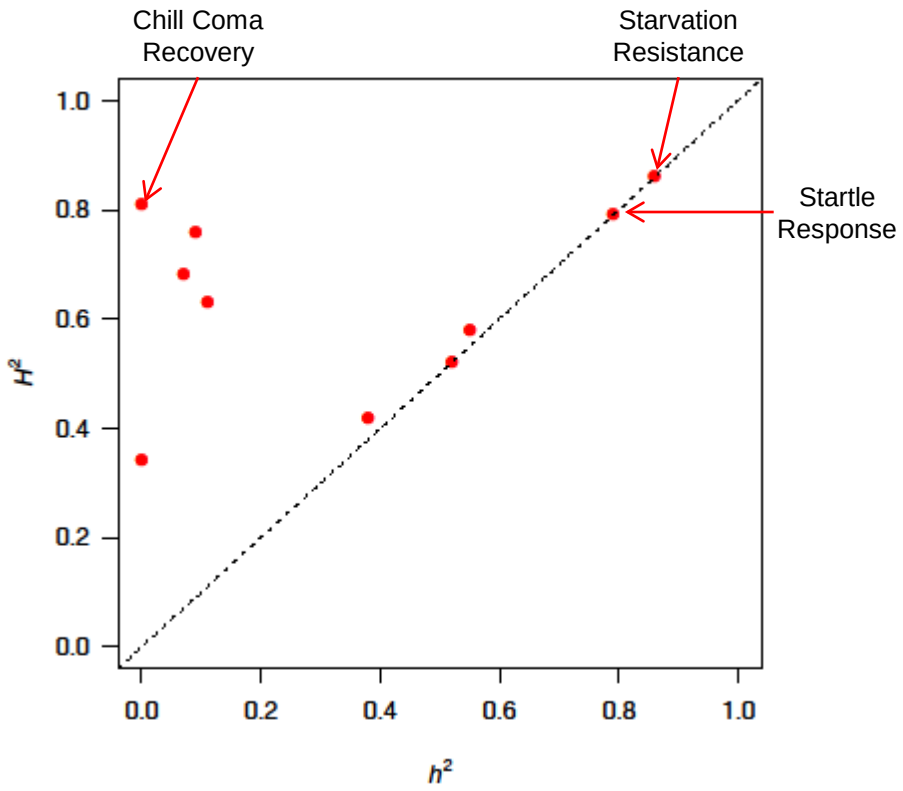


Evidence for Epistasis From h^2 (Realized) and H^2

	Observed		Expected
Trait	Realized h^2	H^2 (Inbred Lines)	$H^2 = 2h^2/(1+h^2)$
Copulation latency	0.07	0.25	0.13
Startle response	0.16	0.58	0.28
Aggressive behavior	0.09	0.78	0.17
Ethanol knock-down time	0.08	0.24	0.15

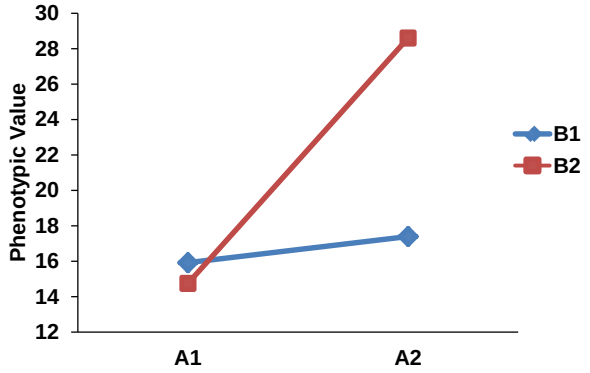
Mackay *et al*. 2005, PNAS 102: 6622-6629; Edwards *et al*. 2006, PLoS Genetics 2: 3154; Jordan *et al*. 2007, Genome Biol. 8: R172; Morozova *et al*. 2009, Genetics 83: 733-745; Ayroles *et al*. 2009, Nat. Genet. 41: 299-307; Edwards *et al*. 2009, Genome Biol. 10: R76; Morozova *et al*. 2009, Genetics 83: 733-745

Evidence for Epistasis From h^2 (GBLUP) and H^2



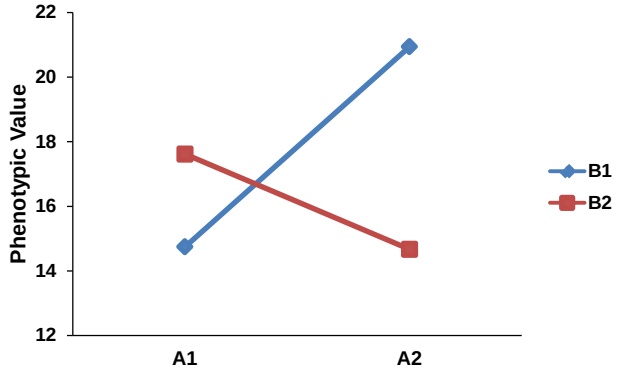
Apparent additivity

- Estimates of V_A include V_{AA}
- Epistasis from change of magnitude of effects in different backgrounds



Largely epistastic

- Epistasis from change of direction of effects in different backgrounds



Hypothesis

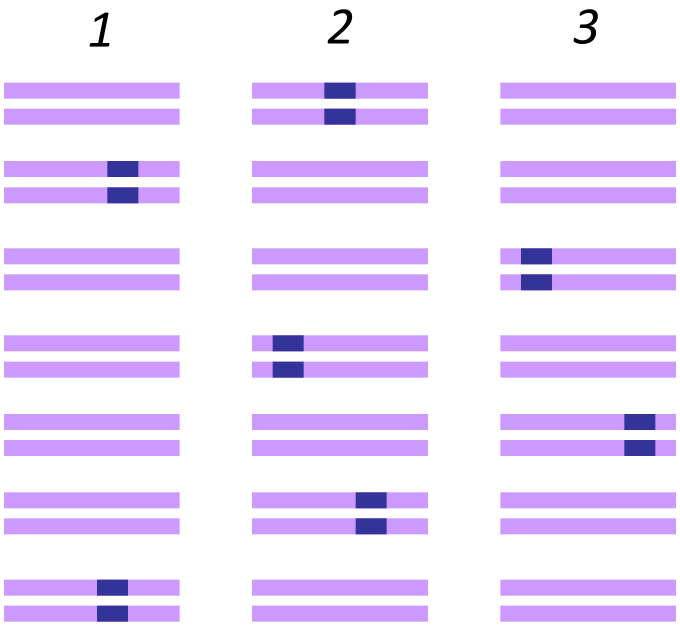
IF

- Extensive epistatic interaction networks suppress allelic effects in outbred populations

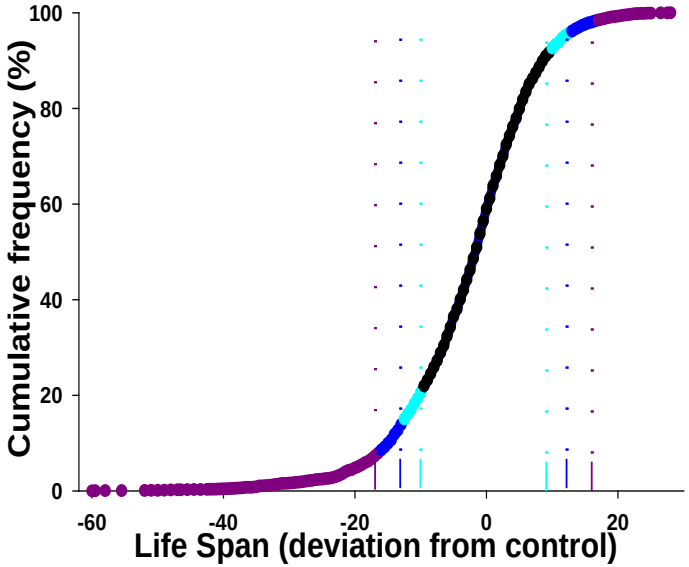
THEN EXPECT

- Large numbers of induced mutations in a single inbred background affect quantitative traits (decanalization)
- The same mutations have different effects in different genetic backgrounds
- Mutations affecting the same trait perturb the same underlying network and interact epistatically
- Response to selection from mutations affecting the same trait will be less than expected based on homozygous and heterozygous effects
- Natural variants suppress (buffer) effects of newly induced mutations

P-Element Mutagenesis in *Drosophila*

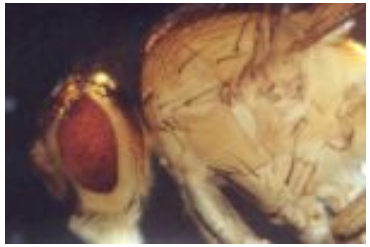


Magwire *et al*. 2010, PLoS Genetics 6: e1001037



Large Numbers of Induced Mutations Affect Quantitative Traits?

21.9%



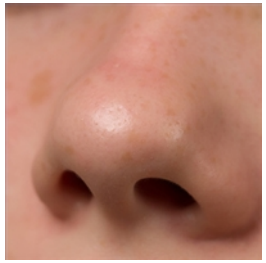
23.3%

34.7%

24.8%

41.1%

37.1%

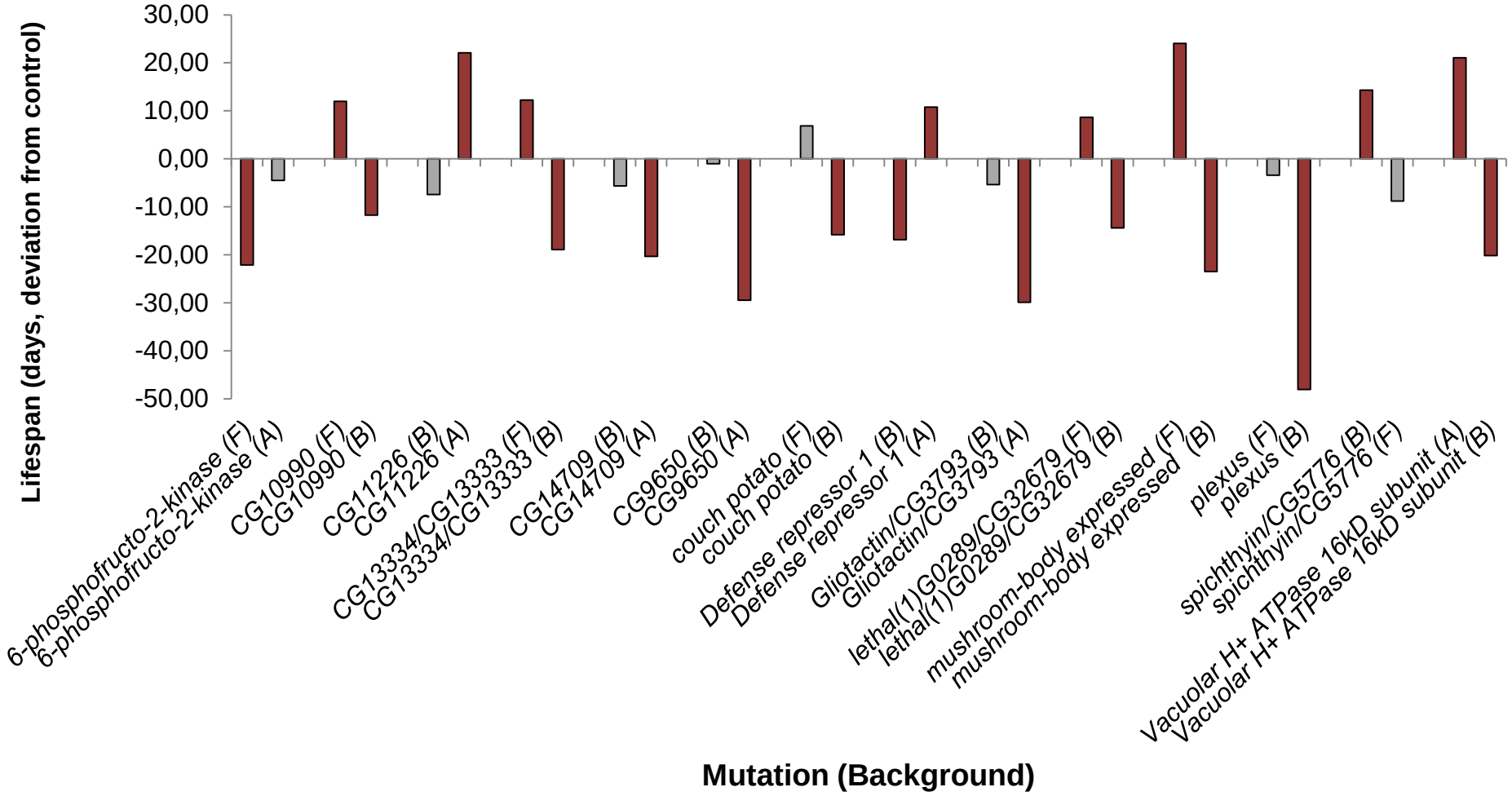


6.2%

32.7%

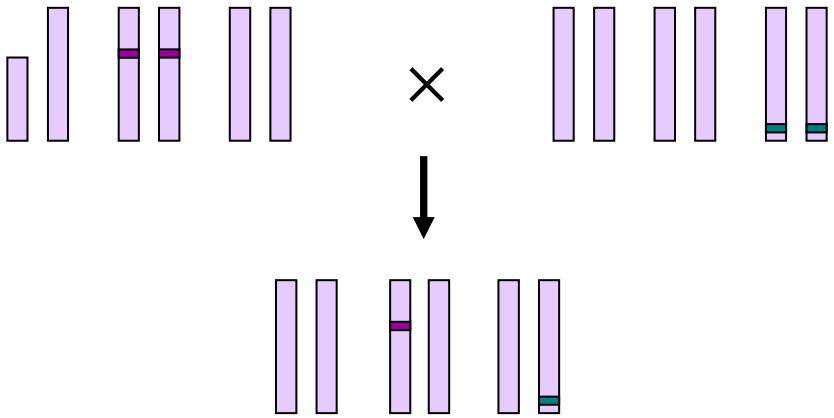
Dilda & Mackay 2002, Genetics 162: 1655; Norga *et al*. 2003, Curr. Biol. 13: 1388-1397; Harbison *et al*. 2004, Genetics 166: 1807-1823; Sambandan *et al*. 2006, Genetics 174: 1349-1363; Yamamoto *et al*. 2008, PNAS 105: 12393-12398; Edwards *et al*. 2009, BMC Biology 7: 29; Magwire *et al*. 2010, PLoS Genetics 6: e1001037; Morozova *et al*. 2011, Genetics 187: 1193-1205

Same Mutations Have Different Effects in Different Genetic Backgrounds?



Magwire *et al*. 2010, PLoS Genetics 6: e1001037

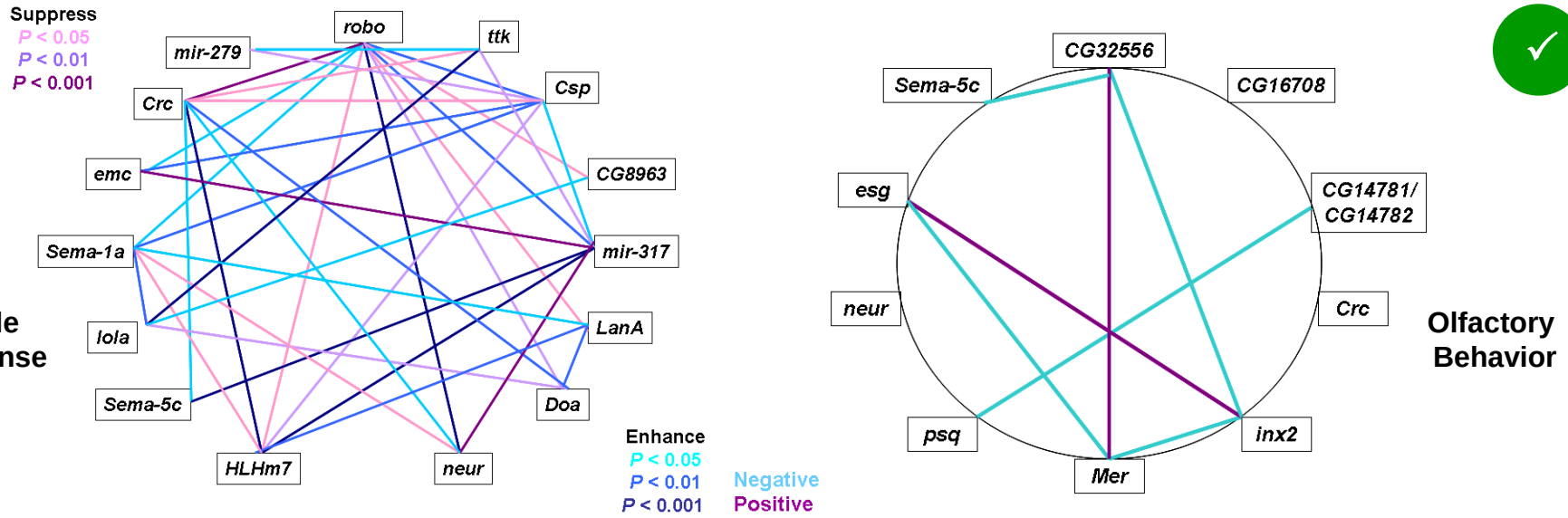
Mutations Affecting the Same Trait Interact Epistatically?



1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	1,10
	2,3	2,4	2,5	2,6	2,7	2,8	2,9	2,10
		3,4	3,5	3,6	3,7	3,8	3,9	3,10
			4,5	4,6	4,7	4,8	4,9	4,10
				5,6	5,7	5,8	5,9	5,10
					6.7	6,8	6,9	6,10
						7,8	7,79	7,10
							8,9	8,10
								9,10

Federowicz *et al*. 1998, Genetics 148: 1885-1891

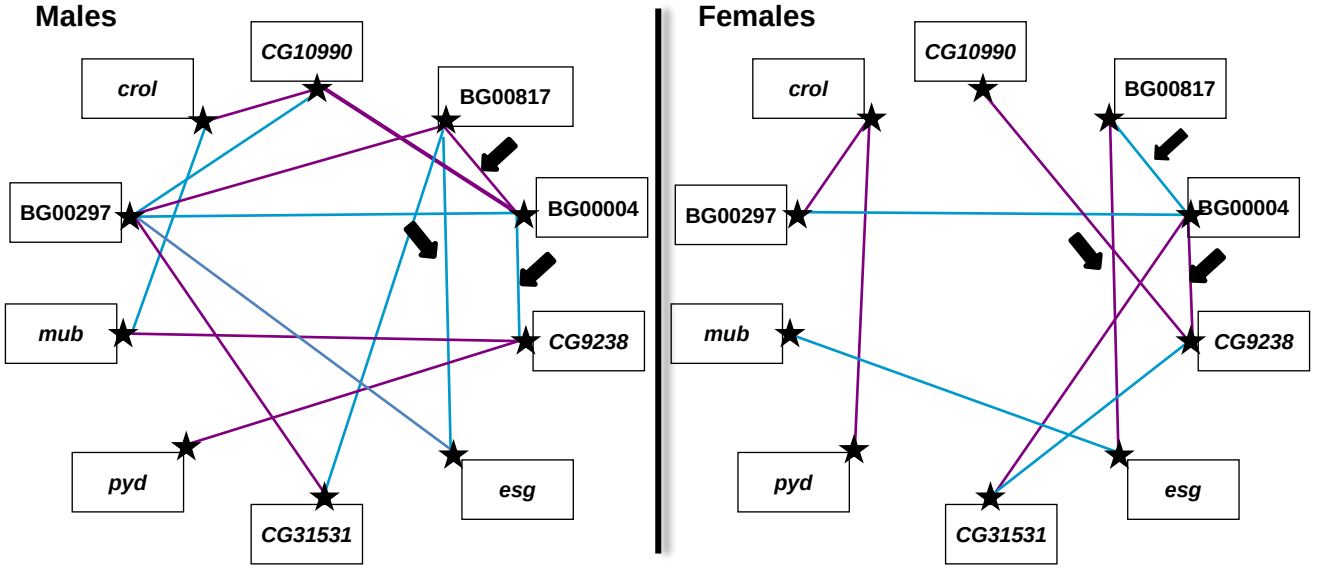
Mutations Affecting the Same Trait Interact Epistatically?



Startle Response

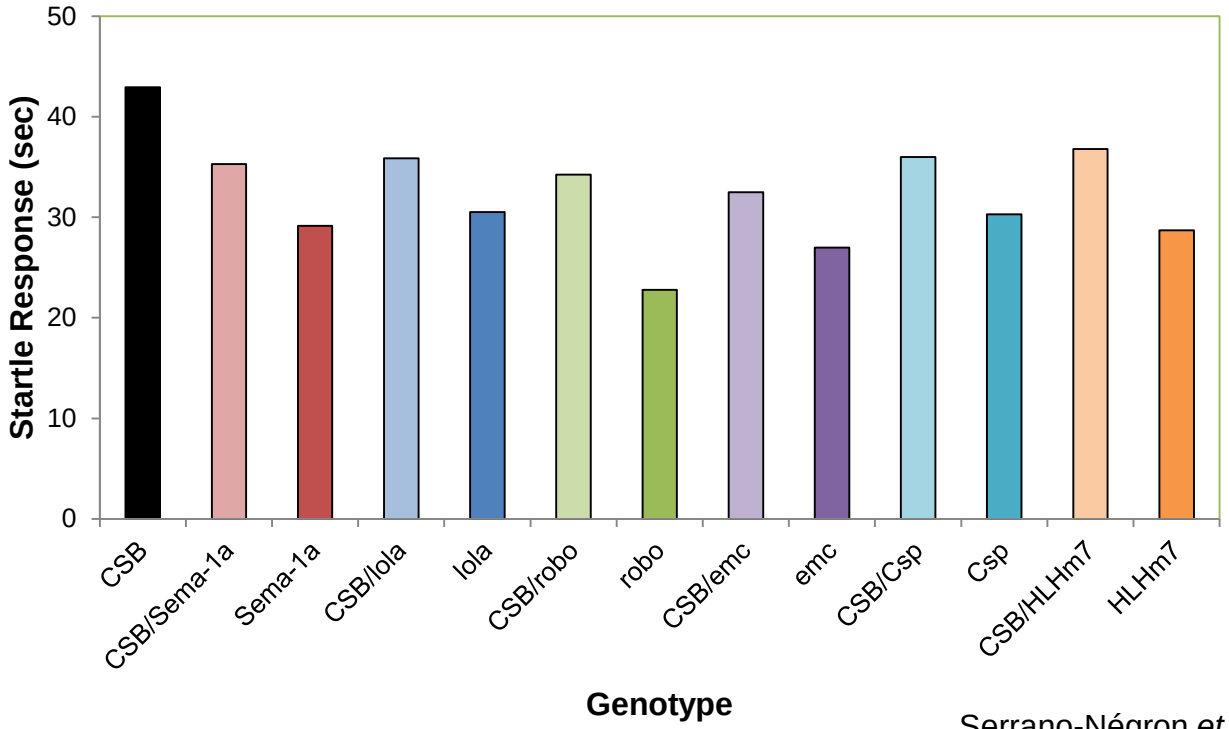
Olfactory Behavior

Life Span



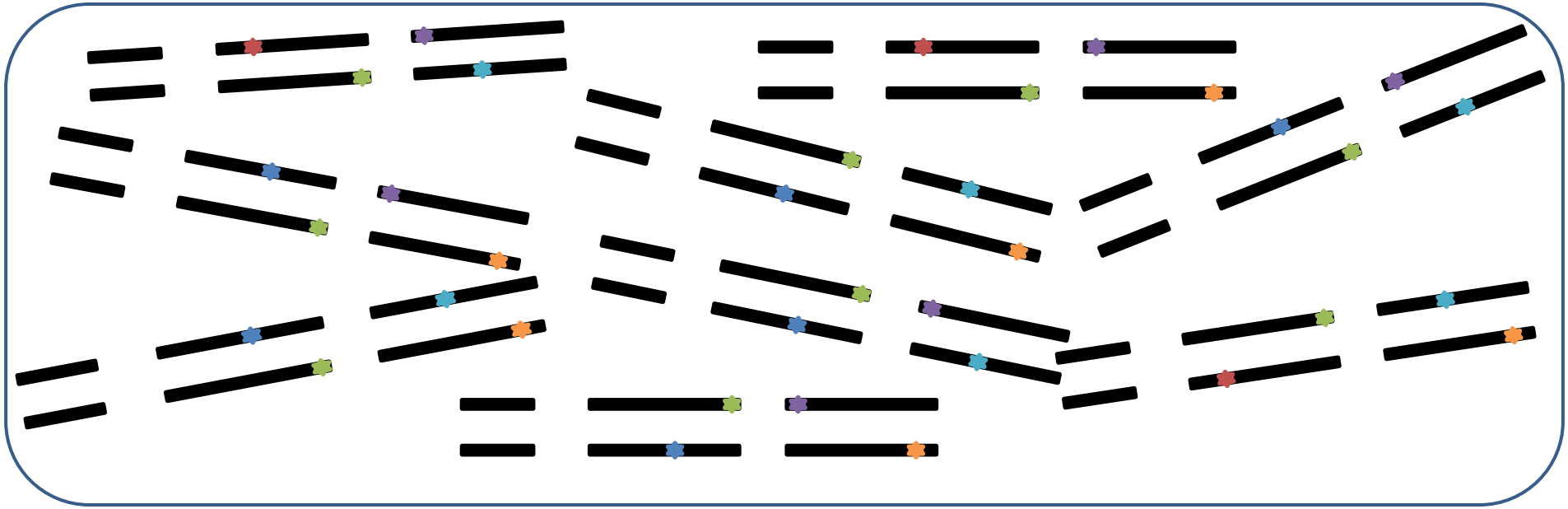
Sambandan *et al*. 2006, Genetics 174: 1349-1363; Yamamoto *et al*. 2008, PNAS 105: 12393-12398; Magwire *et al*. 2010, PLoS Genetics 6: e1001037

Response to Selection From Mutations Less Than Additive?



Serrano-Négron *et al.*, unpublished data

Response to Selection From Mutations Less Than Additive?



$$p = 2/3;\ q = 1/3$$

$$V_A = 2pq\Sigma[a+d(q-p)]^2$$

$$V_D = \Sigma(2pqd)^2$$

$$h^2 = V_A/(V_A+V_D+V_E)$$

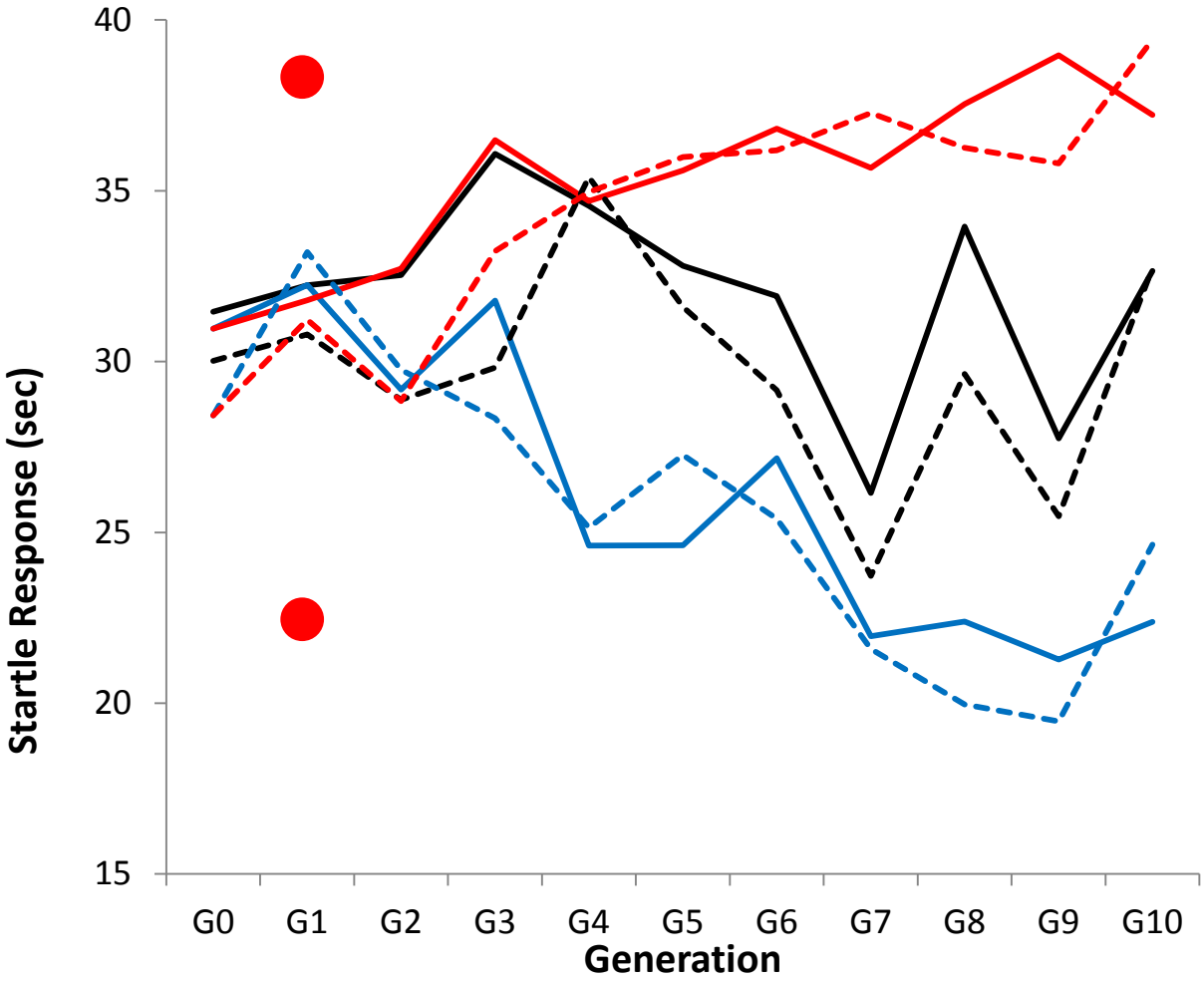
$$h^2_{EXP} = 0.951$$

$$R_1 = ih^2\sigma_P$$

$$R_{EXP} = 15.5$$

Serrano-Négron *et al*., unpublished data

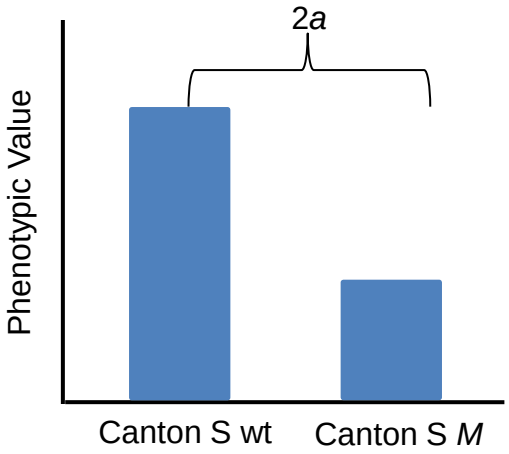
Response to Selection From Mutations Less Than Additive?

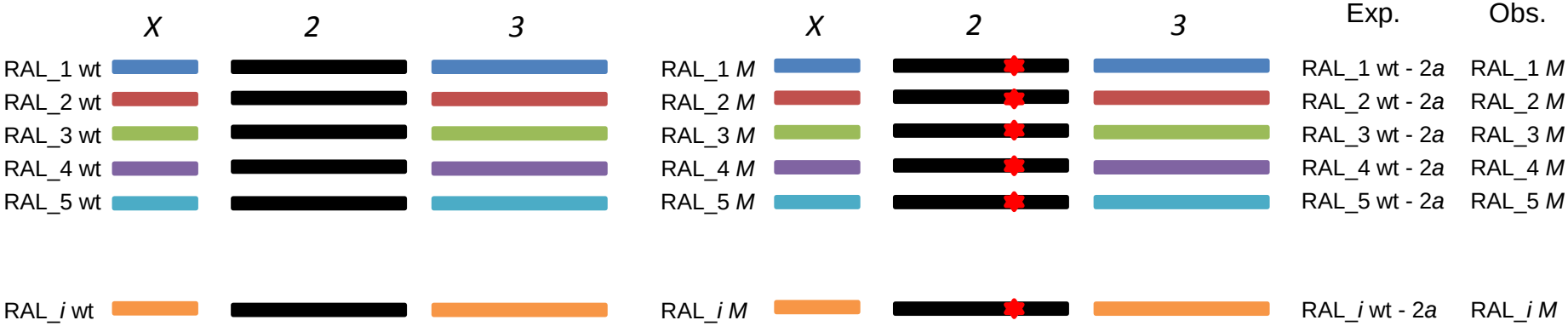


$h^2 = 0.078$

Serrano-Négron *et al*., unpublished data

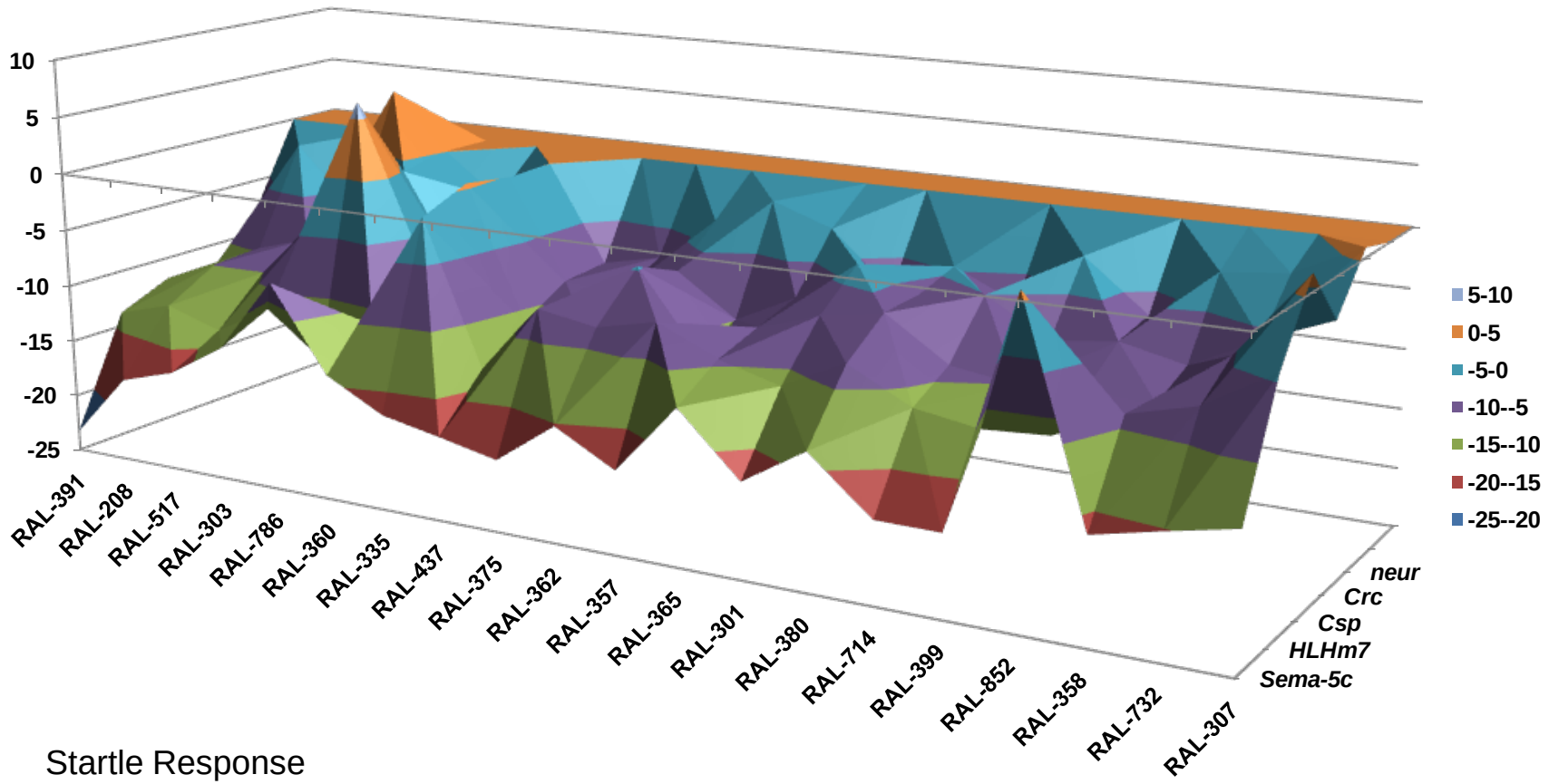
Natural Variants Suppress Effects of Newly Induced Mutations?





Expected – Observed = Estimate of Interaction Effect (*I*)

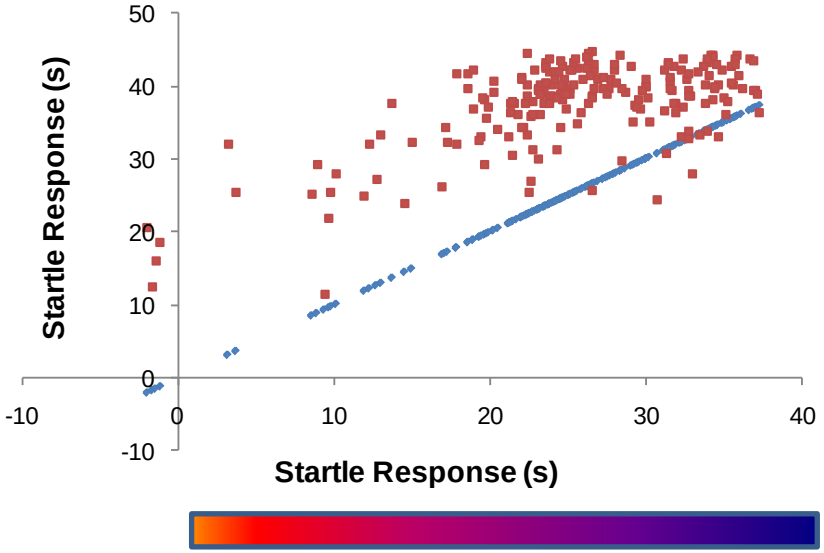
Natural Variants Suppress Effects of Newly Induced Mutations?

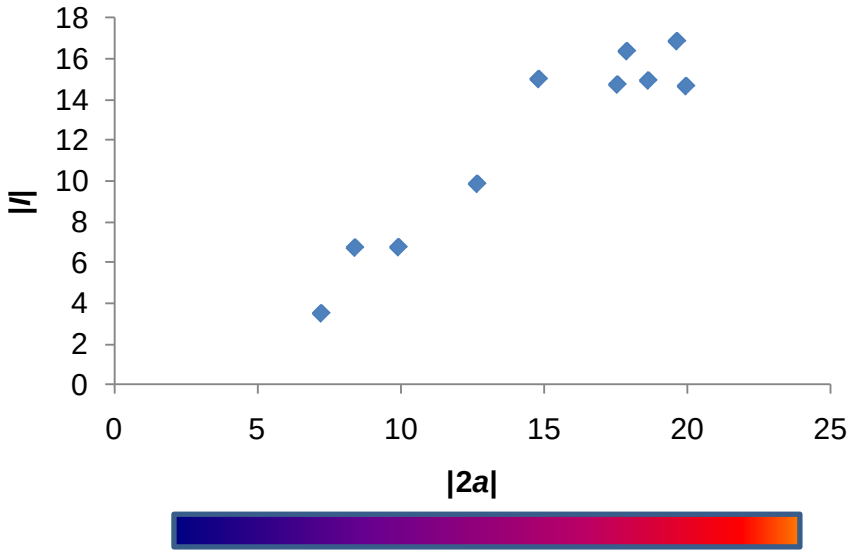


Startle Response

Yamamoto *et al*. Genet. Res. 91: 373-382

Natural Variants Suppress Effects of Newly Induced Mutations?





Yamamoto *et al*. Genet. Res. 91: 373-382

Lessons Learned

- Extensive epistasis for *Drosophila* quantitative traits
- Epistatic interactions define genetic networks that are:
  - Highly interconnected
  - Enriched for biologically plausible GO categories, metabolic and cellular pathways
  - Novel
- Implications of epistasis:
  - Additivity an emergent property of underlying epistatic networks
  - Population-specific genetic architecture for individual molecular variants, but common network
  - Missing heritability – small additive effects even though interactions highly significant
  - Individual risk prediction requires knowledge of interacting partners
  - (De)canalization and human disease
  - Genomic basis of selection response: population-specific
  - Potential for speciation

Team DGRP Freeze 2.0

Robert Anholt
Wen Huang
Yutaka Inoue
Richard Lyman
Mike Magwire
Tanya Morozova
Jason Peiffer
Yazmin Serrano-Négron
Eric Stone
Lavanya Turlipati
Aki Yamamoto

Bart Deplancke
Andreas Massouras

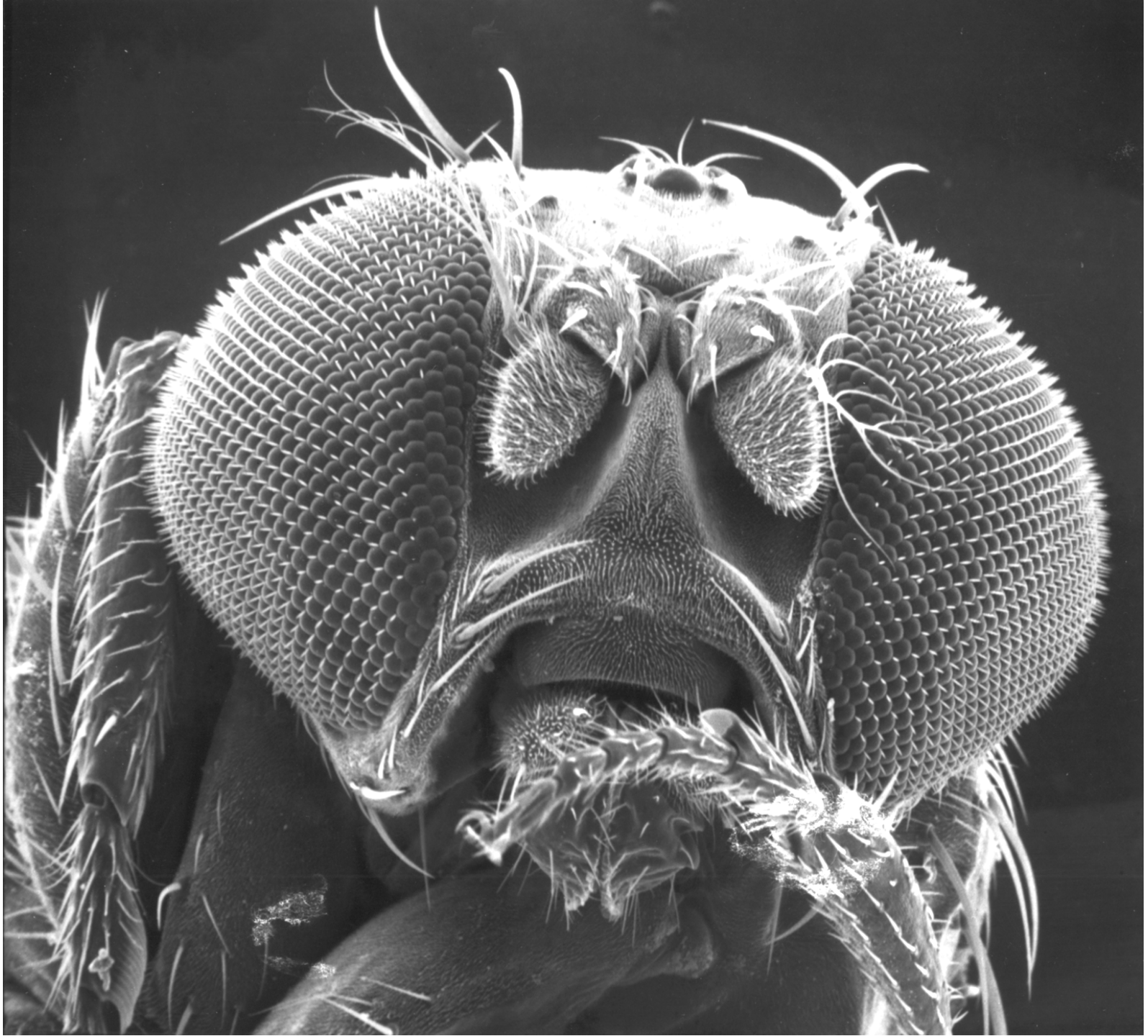
Eileen Furlong
Jan Korbel
Thomas Zichner

David Mittelman

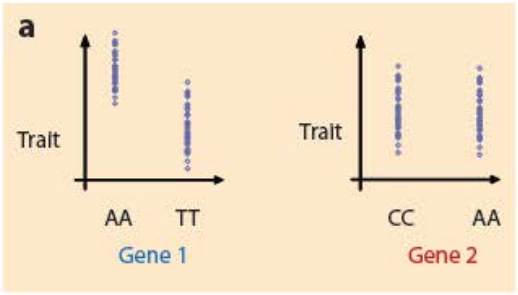
Spencer Johnson
Aaron Tarone

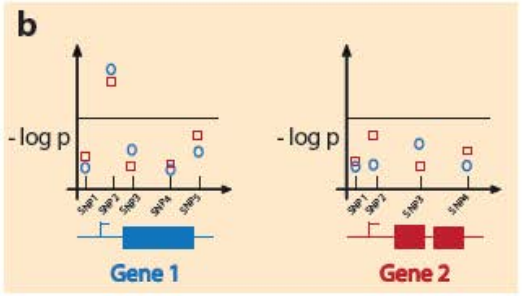
Stephen Richards
Richard Gibbs
Yi Han
Jeff Reid

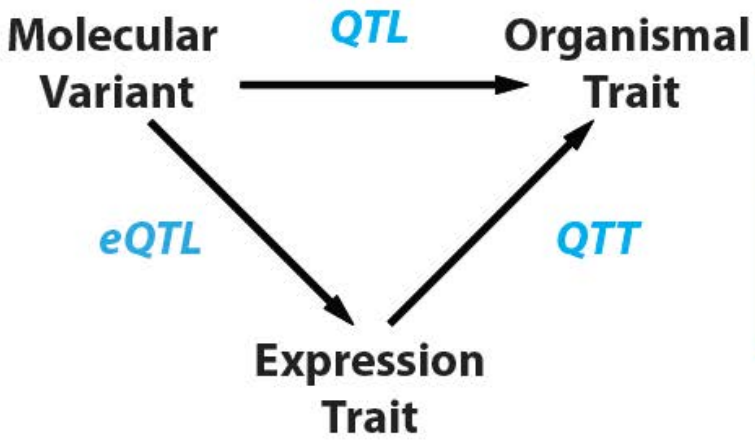
Funding

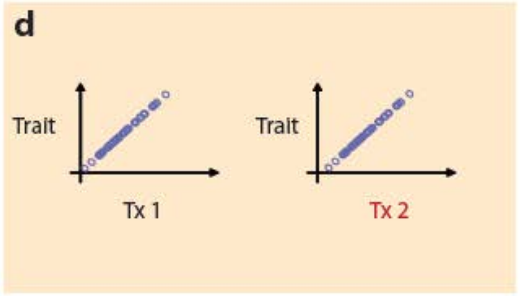


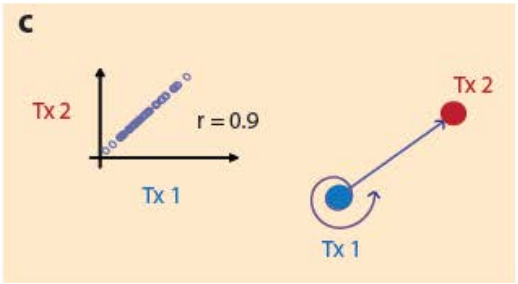
Drosophila Systems Genetics







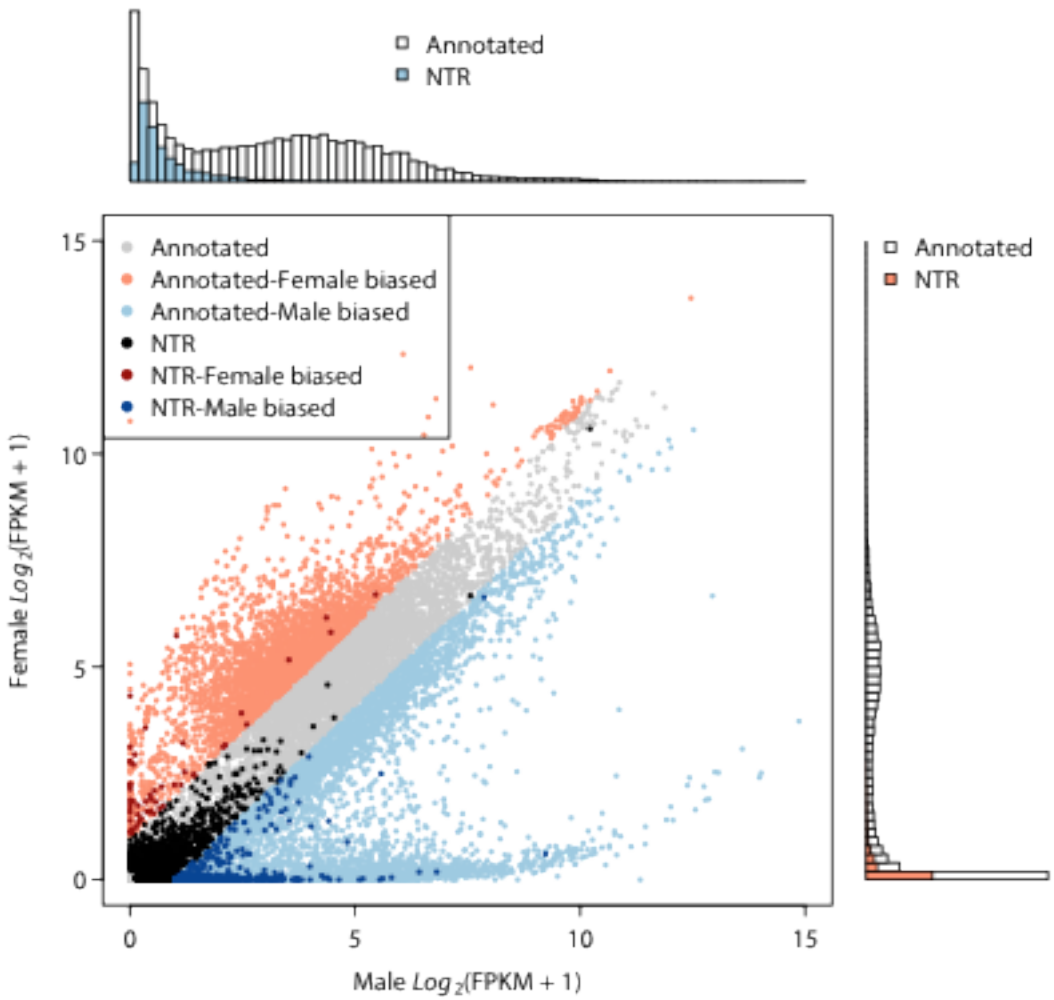




Mackay *et al*. Nat Rev Genet. (2009) 10:565

DGRP Transcriptomes

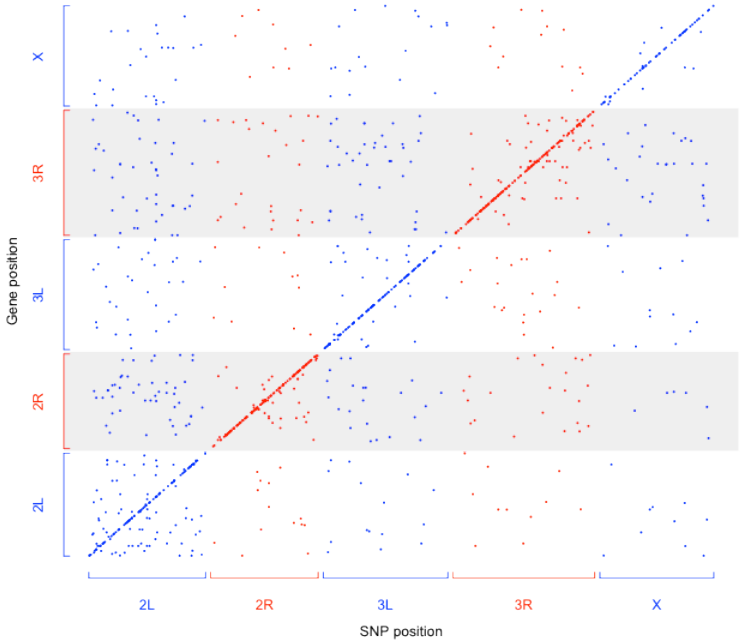
- Illumina RNA-Seq, pooled RNA from 192 DGRP lines (3-5 day old mated flies, whole bodies)
- ~ 100M fragments (100bp paired end) for females and males separately
- Aligned to R5.49 of the Flybase annotation and the reference genome
- Transcript models assembled based on RNA-Seq alignments
- 2,807 novel splice isoforms
- 3,504 (3.6M bp) Novel Transcribed Regions (NTRs)
- 427 and 228 NTRs genetically variable in males and females, respectively



Enabling Systems Genetics in the DGRP

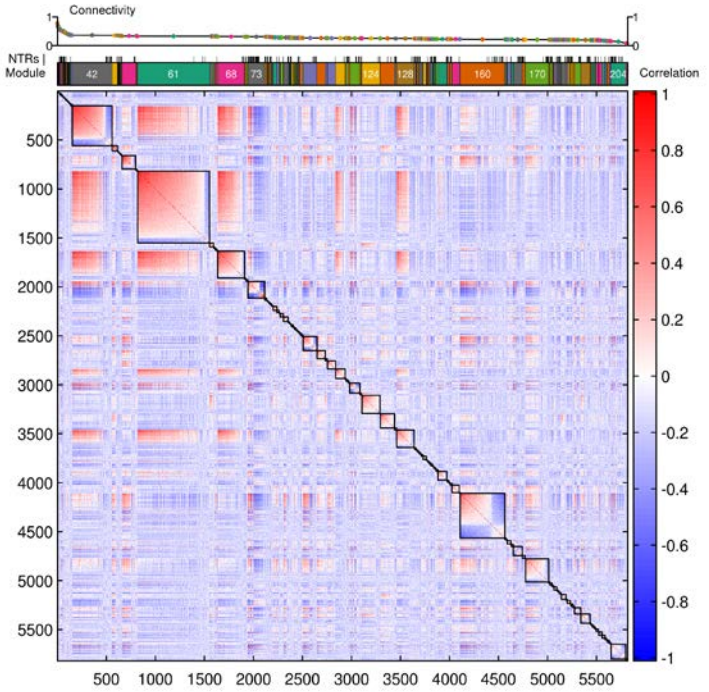
- Affymetrix 2.0 tiling arrays
- 192 DGRP lines
- 3-5 day old, mated males and females
- Whole bodies

eQTL Mapping



- Old, noisy technology
- Whole bodies?
- One developmental stage?
- *Drosophila* Variome Project (DVP) based on DGRP?

Co-Expression Modules



NTR positions, indicated by vertical bars, often negatively correlated with protein-coding genes