

# The role of deleterious substitutions in crop genomes

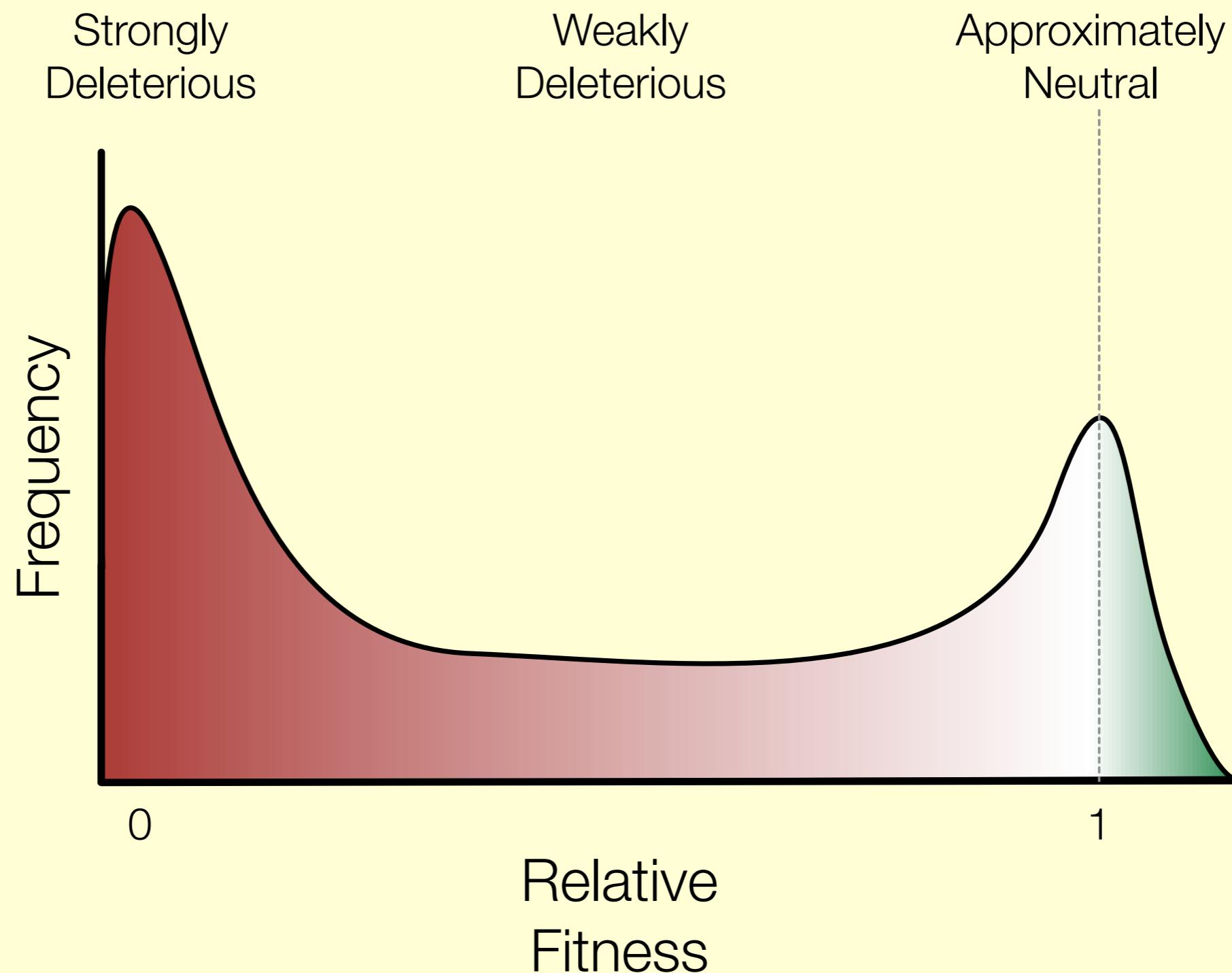
Thomas Kono  
2016-09-23

# A Classic Issue

Discussion of deleterious alleles goes back 85 years

party in the adaptation. The case of large mutations to the organism may first be considered, since their consequences in this connexion are of an extremely simple character. A considerable number of such mutations have now been observed, and these are, I believe, without exception, either definitely pathological (most often lethal) in their effects, or with high probability to be regarded as deleterious in the wild state. This is merely what would be expected on the view, which

# Hypothetical Distribution of Fitness Effects



# Genetic Load

- Reduction in mean fitness of population with respect to a hypothetical optimal genotype

$$L_m = (2pq)(hs) = (m/hs)(hs) = m. \quad (7)$$

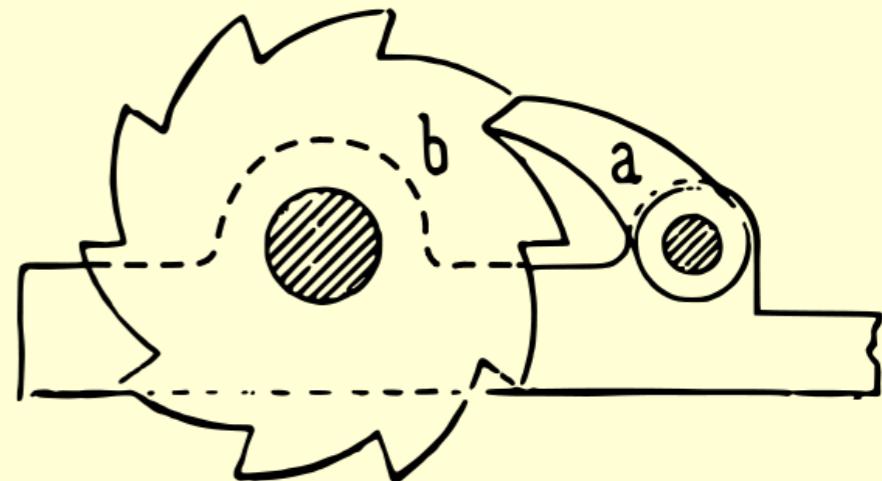
Mutation

$$L_s = sp^2 + ta^2 = st/(s + t). \quad (9)$$

Segregation

# Why deleterious alleles segregate

- Purgling effects - Finite  $N_e$  limits the effectiveness of purifying selection<sup>1</sup>
- “Muller’s Ratchet”<sup>2</sup> - Deleterious mutations fix in low recombination regions
- Linked selection effects - genetic hitchhiking<sup>3</sup>



1: Takebayashi and Morrell 2001

2: Muller 1964

3: Hill and Robertson 1966

# Millions of Sequence Variants

1000 Genomes			
Release	Variants	Individuals	Populations
Phase 3	84.4 million	2504	26
Phase 1	37.9 million	1092	14
Pilot	14.8 million	179	4

# A Classic Issue, Revisited

## **Contamination of the Genome by Very Slightly Deleterious Mutations: Why Have We Not Died 100 Times Over?**

ALEXEY S. KONDRAKHOV

*Section of Ecology and Systematics, Cornell University, Ithaca, NY 14853, U.S.A.*

doi:10.1111/evo.12195

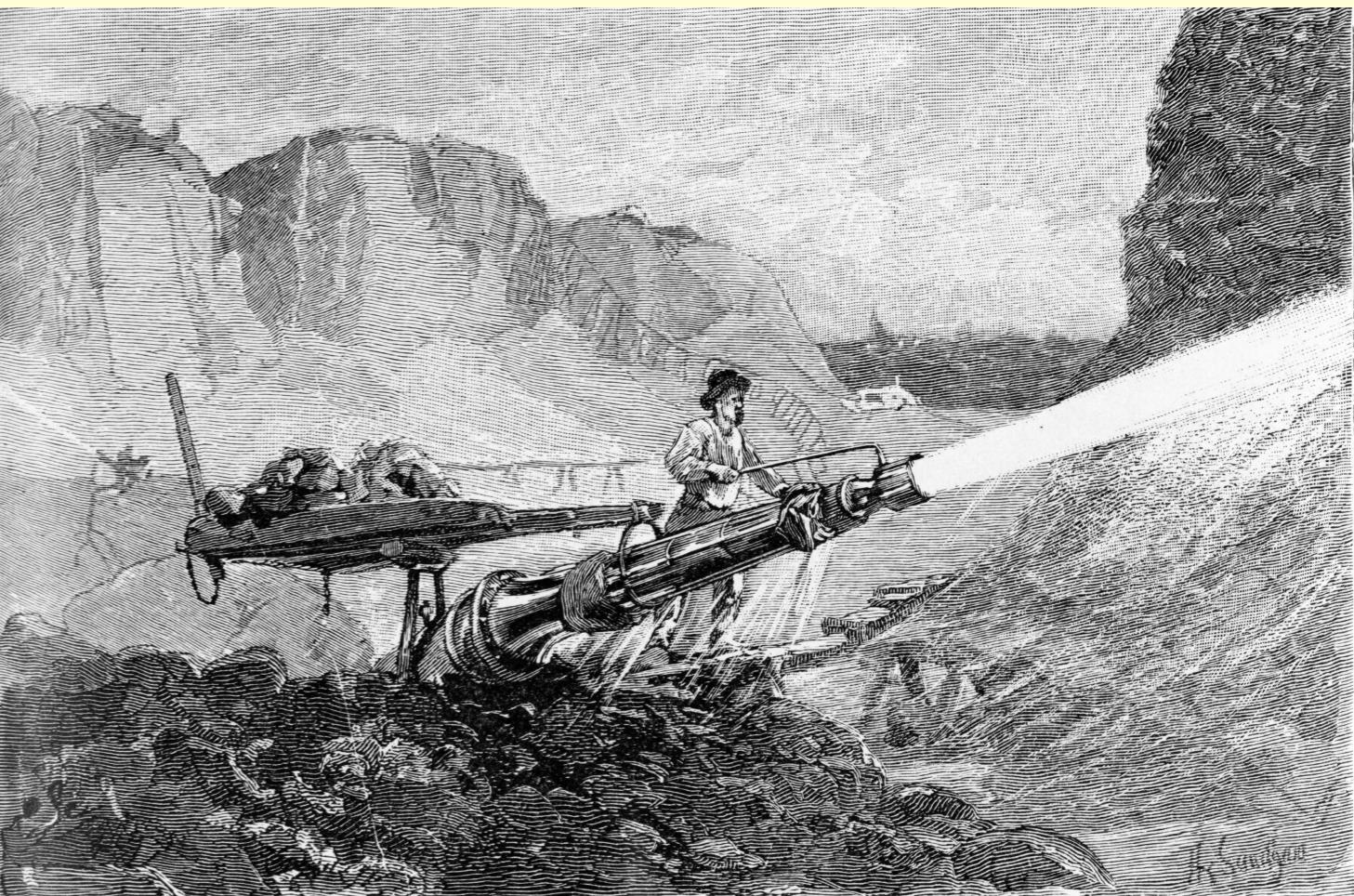
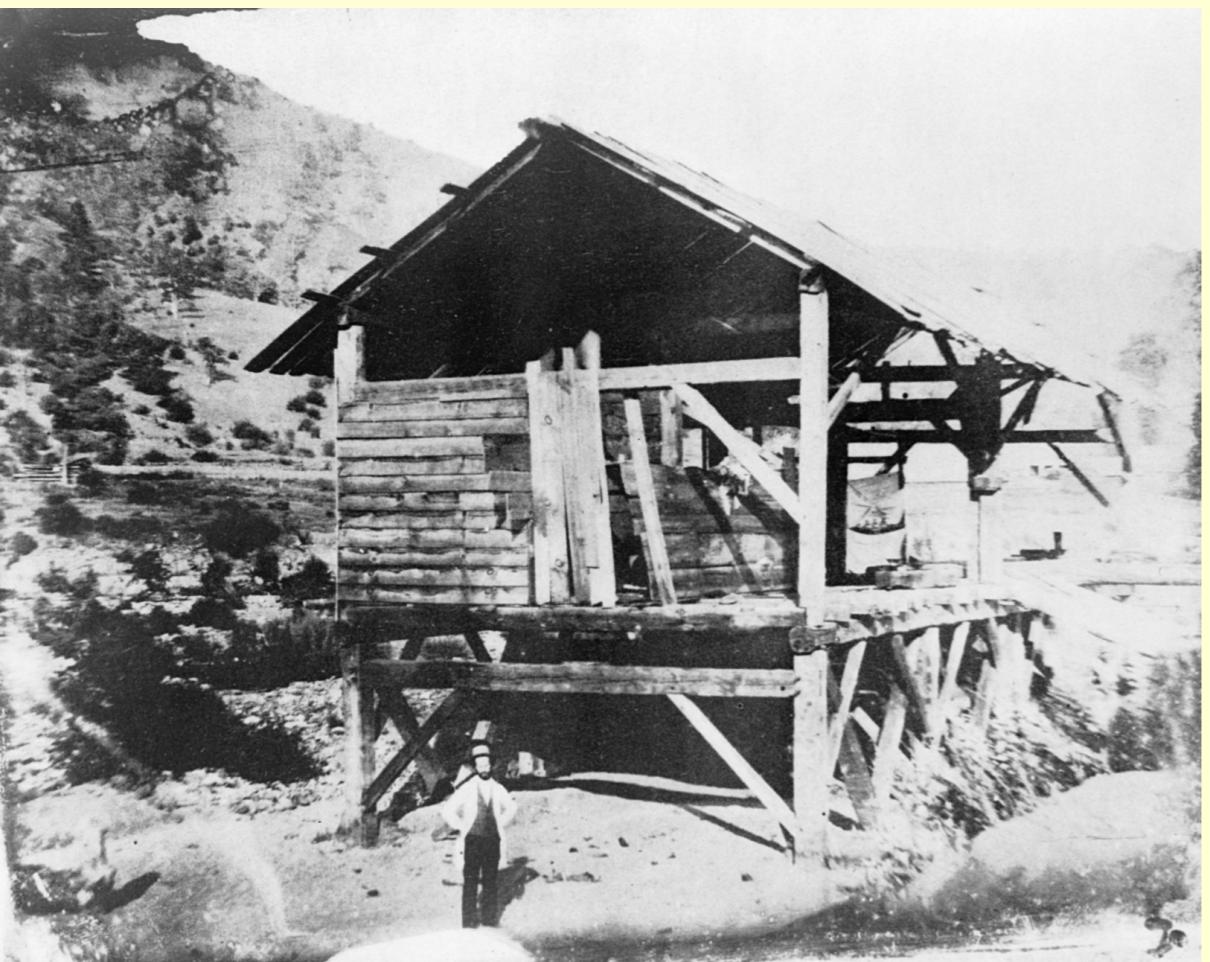


## **WHY WE ARE NOT DEAD ONE HUNDRED TIMES OVER**

Brian Charlesworth<sup>1,2</sup>

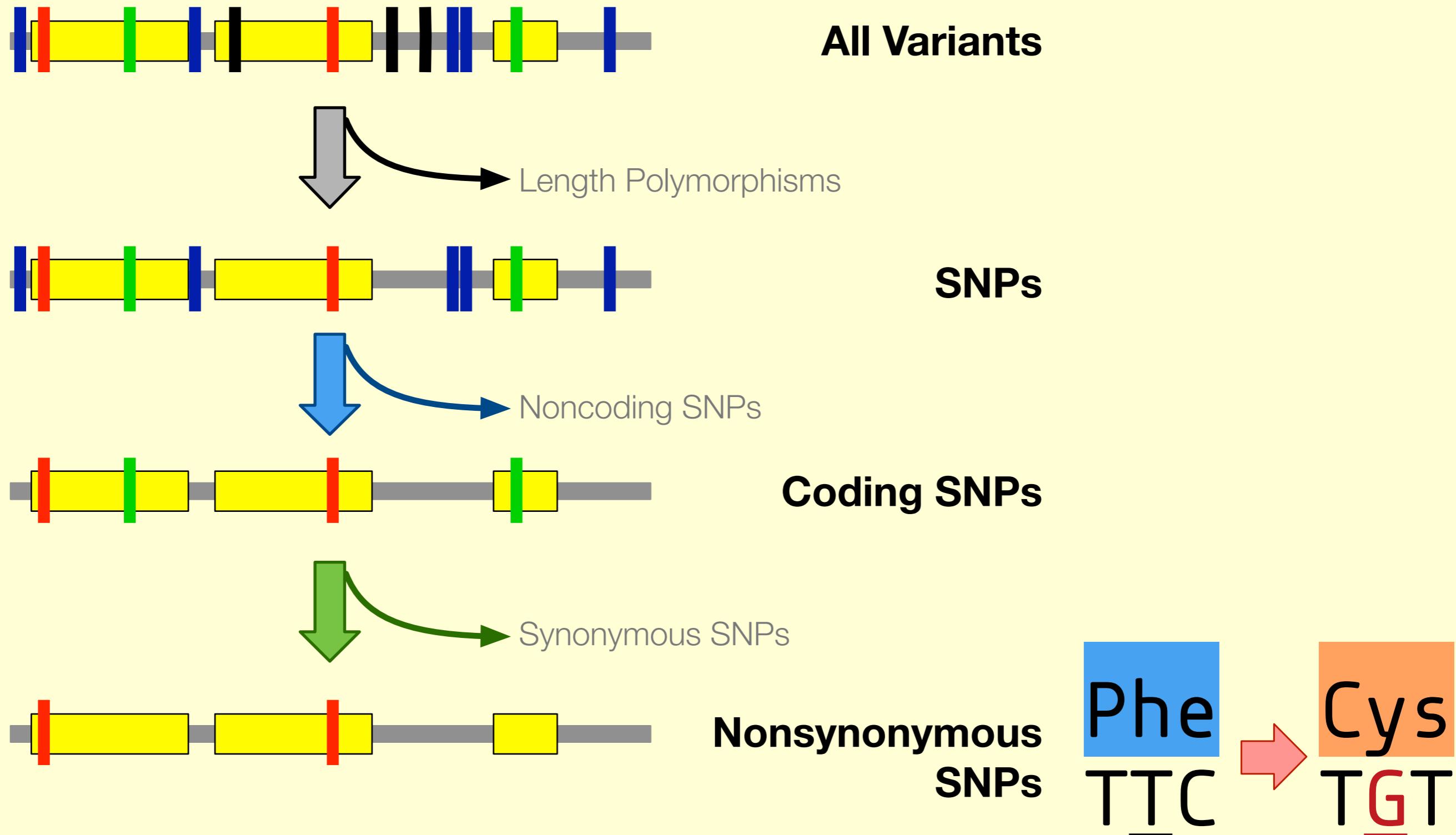
<sup>1</sup>*Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT,  
United Kingdom*

# Why deleterious alleles?

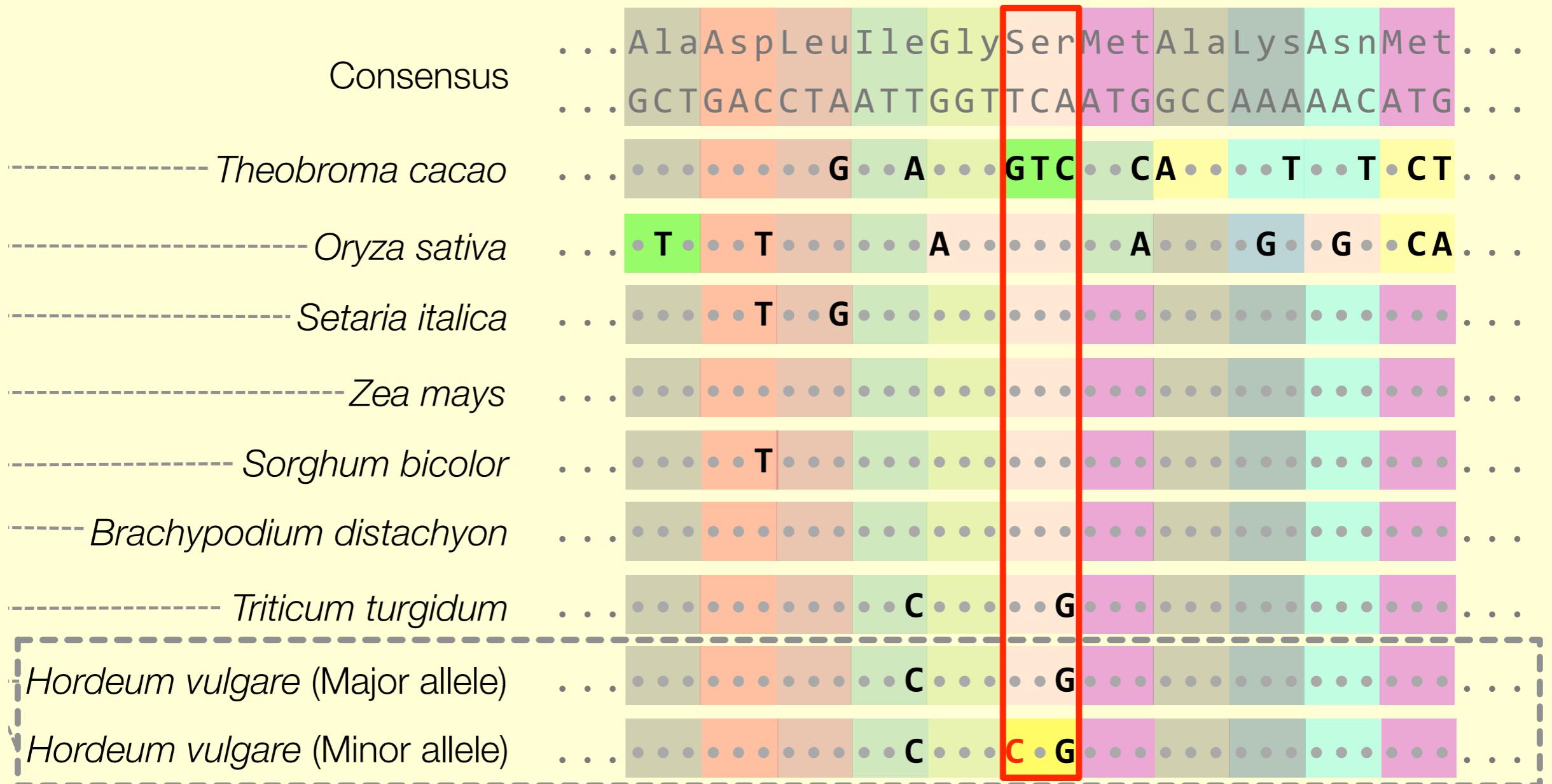


Lewontin 1974  
Rockman 2012

# How to Identify Potentially Deleterious Variants



# How to Identify Potentially Deleterious Variants



# How to Identify Potentially Deleterious Variants

- Example: C/G SNP segregating in barley

Ancestral:     ...Ala...

               ...CCT...

Derived:      ...Pro...

               ...GCT...

Alignment:       
EKSKEEEKEENNAEENEEDDAAPAAAEEKKEEGGEEEEEE

**Tolerated**, does not impact protein function

# How to Identify Potentially Deleterious Variants

- Example: A different C/G SNP segregating in barley

Ancestral:     ...Thr...

              ...ACA...

Derived:     ...Arg...

              ...AGA... → Not in alignment

Alignment:    

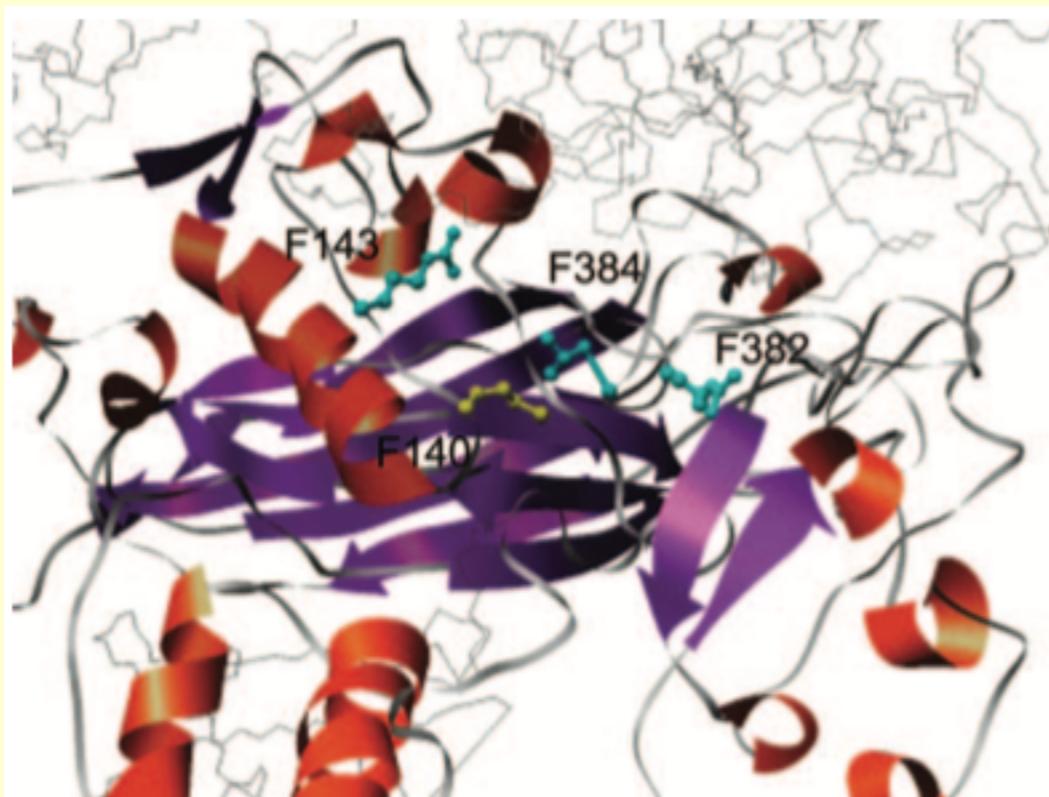
**Deleterious**, does impact protein function

# Some Caveats: Non-additive Effects

- Within a locus:
  - Deleterious alleles expected to be recessive<sup>1</sup> but most approaches estimate only additive effects
  - Less important in inbreeding species, but still important for predicting phenotypes<sup>2</sup>

# Some Caveats: Non-additive Effects

- Between loci:
  - Compensatory mutations - epistatic effects



# Some Caveats: Conditional Fitness Effects

- Programs predict loss of protein function
  - Assume that function is conserved across alignment, and loss of function is deleterious
  - Biased toward strong loss of function
  - Positive selection on new function - including local adaptation - violates assumption

# Outline

- Collecting metadata on SNPs
- Survey of potentially deleterious SNPs in crop genomes
- Potentially deleterious SNPs in a barley breeding program

# SNP Metadata

- What information about SNP markers is available in public databases?
- How do annotations from GenBank compare to those from a reference genome?

# SNP Metadata: A Photo Analogy



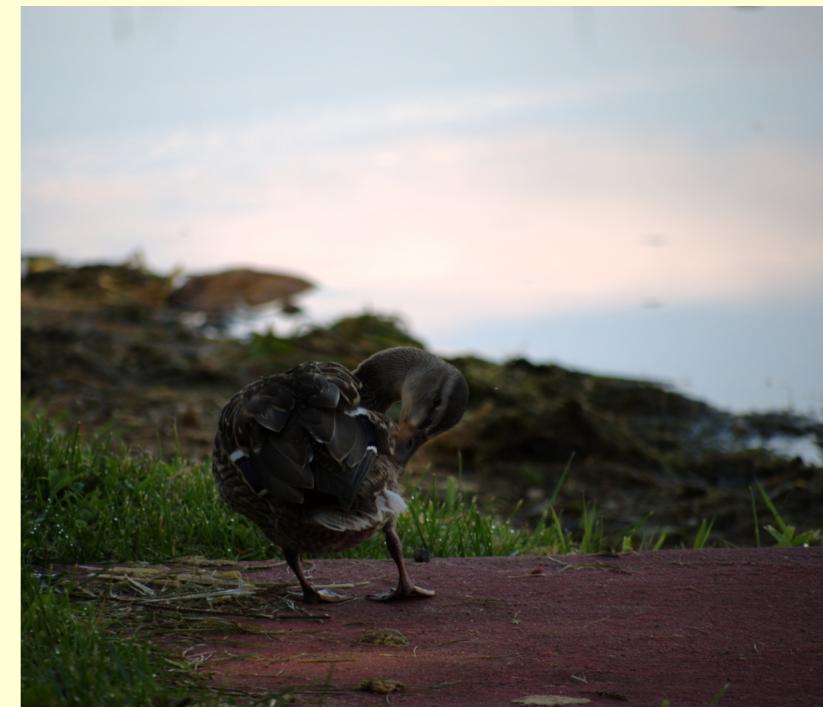
# SNP Metadata: A Photo Analogy

Kind: JPEG image
Size: 732,671 bytes (733 KB on disk)
Where: Data_Disk ▶ Dropbox ▶ Pictures ▶ Photos ▶ Itasca Photos
Created: August 20, 2011 at 17:02:32
Modified: August 20, 2011 at 17:02:32
<input type="checkbox"/> Stationery pad
<input type="checkbox"/> Locked
▼ More Info:
Dimensions: 1200 × 1057
Device make: Panasonic
Device model: DMC-G1
Color space: RGB
Color profile: sRGB IEC61966-2.1
Focal length: 200
Alpha channel: No
Red eye: No
F number: 5.6
Exposure program: 1
Exposure time: 1/125

Date info

Camera info

Exposure info



# SNP Metadata: A Photo Analogy

Kind:	JPEG image
Size:	732,671 bytes (733 KB on disk)
Where:	Data_Disk ▶ Dropbox ▶ Pictures ▶ Photos ▶ Itasca Photos
Created:	August 20, 2011 at 17:02:32
Modified:	August 20, 2011 at 17:02:32
<input type="checkbox"/> Stationery pad	
<input type="checkbox"/> Locked	
▼ More Info:	
Dimensions:	1200 × 1057
Device make:	Panasonic
Device model:	DMC-G1
Color space:	RGB
Color profile:	sRGB IEC61966-2.1
Focal length:	200
Alpha channel:	No
Red eye:	No
F number:	5.6
Exposure program:	1
Exposure time:	1/125

- Can apply same information to SNPs
  - Gene name
  - Coding/Noncoding
  - Synonymous/  
Nonsynonymous

# SNP Metadata: Pipeline Development

BLAST  
Hits

GenBank  
Record

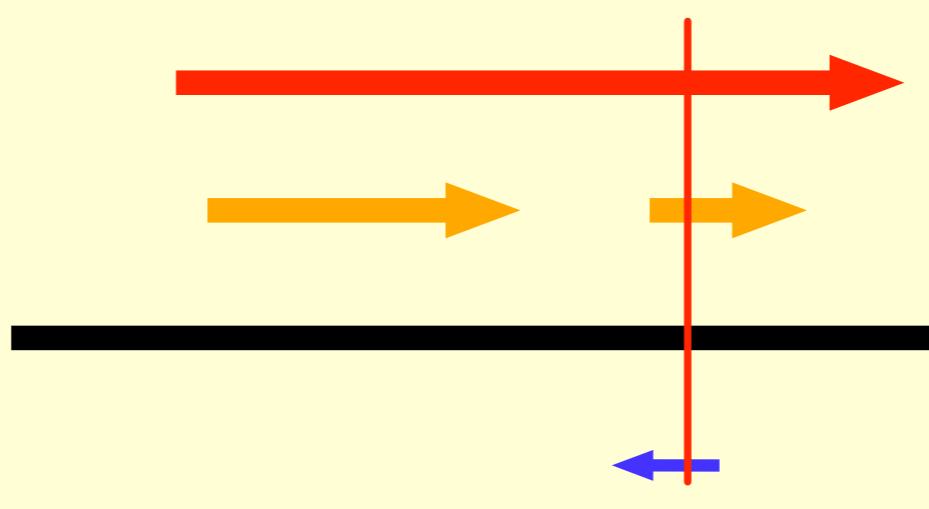
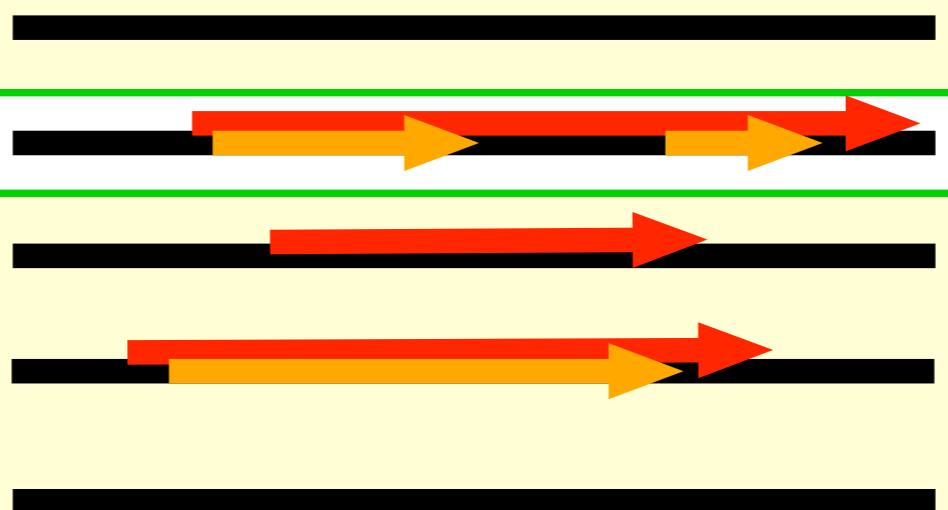
SNP Contextual Sequence

Gene

CDS

Sequence

C / G  
Arg / Gly



Amber  
Eule-Nashoba



Kiran Seth



# Chapter 1: Output to dbSNP

Gene Name	GenBank Number	Protein Impact
SNP: RIL01_12001_39 GENENAME: CG4629 ACCESSION: AE014134 COMMENT: nonsynonymous I S SAMPLESIZE: LENGTH: ? 5'_FLANK: ATTCACTCCATTTCAGAAGAGAGCGGCAGGGACAGCGACGCCGGCTGTAGCGGCGCCA OBSERVED: T/G 3'_FLANK: TGGCAGCGGCAAAAAGCCCAGTGAGAGAGCCGTCATTTGAGAGGTCGCCGGAGCGCC 		

SNP States

# SNP Metadata: Testing in Several Species

- Can annotate genotyping SNPs or GBS SNPs
- Better annotations from more well developed genomic resources

Organism (Source)	Annotated	Syn and Non-coding	Nonsyn	Total
Fruit Fly (GBS)	7,636	6,269	1,362	9,766
Goatgrass (GBS)	199	135	49	5,000
Oat (GBS)	51	30	11	5,000
Cotton (dbSNP)	168	131	32	5,000
Wheat (Illumina)	5,377	3,398	1,636	8,632
Barley (Illumina)	7,774	6,095	1,593	9,606

# SNP Metadata: Summary

- Gene name and protein sequence impact are available from annotation against public resources.
- Annotations from GenBank are largely concordant with those from a reference genome, but also include annotations from related species.

# Outline

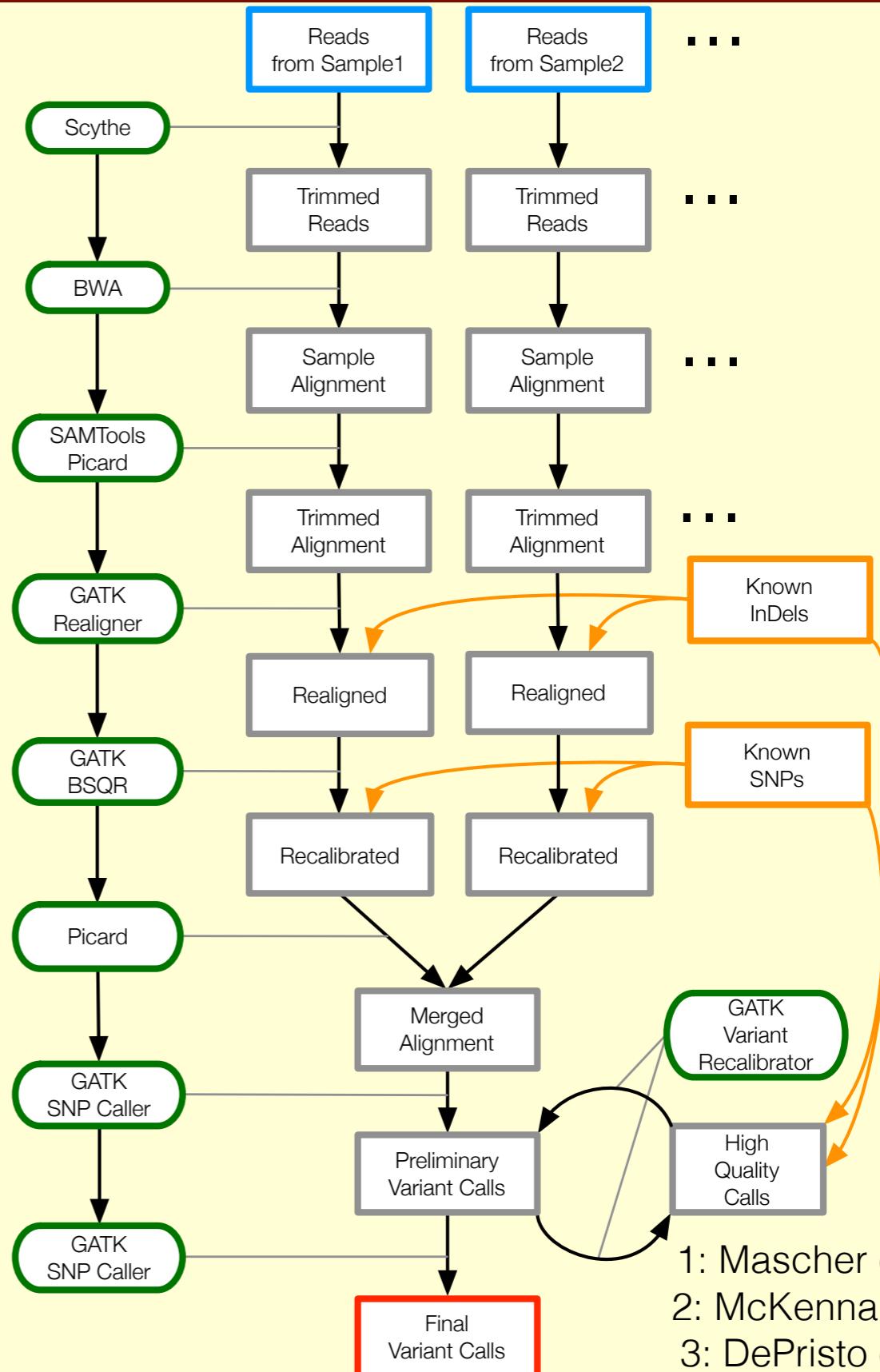
- Collecting metadata on SNPs
- Survey of potentially deleterious SNPs in crop genomes
- Potentially deleterious SNPs in a barley breeding program

# Deleterious SNP Survey

- How many putatively deleterious SNPs segregate in two crop species?
- Are SNPs that are causative for a phenotypic variant more likely to be annotated as deleterious, as compared to those without known phenotypic effects?

# Deleterious SNP Survey: DNA Resequencing Data

- Exome resequencing<sup>1</sup> of 15 barley accessions
- Whole genome resequencing of 8 soybean accessions
- Workflow based on GATK best practices<sup>2,3</sup>



# Deleterious SNP Survey: Addressing Reference Bias

	Morex (Reference)	Steptoe	Harrington	Kindred
SNP 1	A	G	G	A
SNP 2	T	C	T	T
SNP 3	C	C	T	T
Diff. From Reference	0	2	2	1

# Deleterious SNP Survey: Addressing Reference Bias

	Ancestral	Morex (Reference)	Steptoe	Harrington	Kindred	
SNP 1	A	A	G	G	A	Ancestral
SNP 2	C	T	C	T	T	Derived
SNP 3	T	C	C	T	T	$0 \rightarrow 1$
Diff. From Ancestral	0	2	2	2	1	

Outgroup Species

Cultivated Species

# Deleterious SNP Survey: BAD\_Mutations

- Implements a likelihood ratio test<sup>1</sup> for neutral v. constrained evolution
- Uses genome alignments from Ensembl Plants<sup>2</sup> and Phytozome<sup>3</sup>
- Corrects for reference bias
- [https://github.com/MorrellLAB/BAD\\_Mutations](https://github.com/MorrellLAB/BAD_Mutations)

1: Chun and Fay 2009

2: <http://plants.ensembl.org/>

3: <https://phytozome.jgi.doe.gov>

# Deleterious SNP Survey: Comparison of Programs

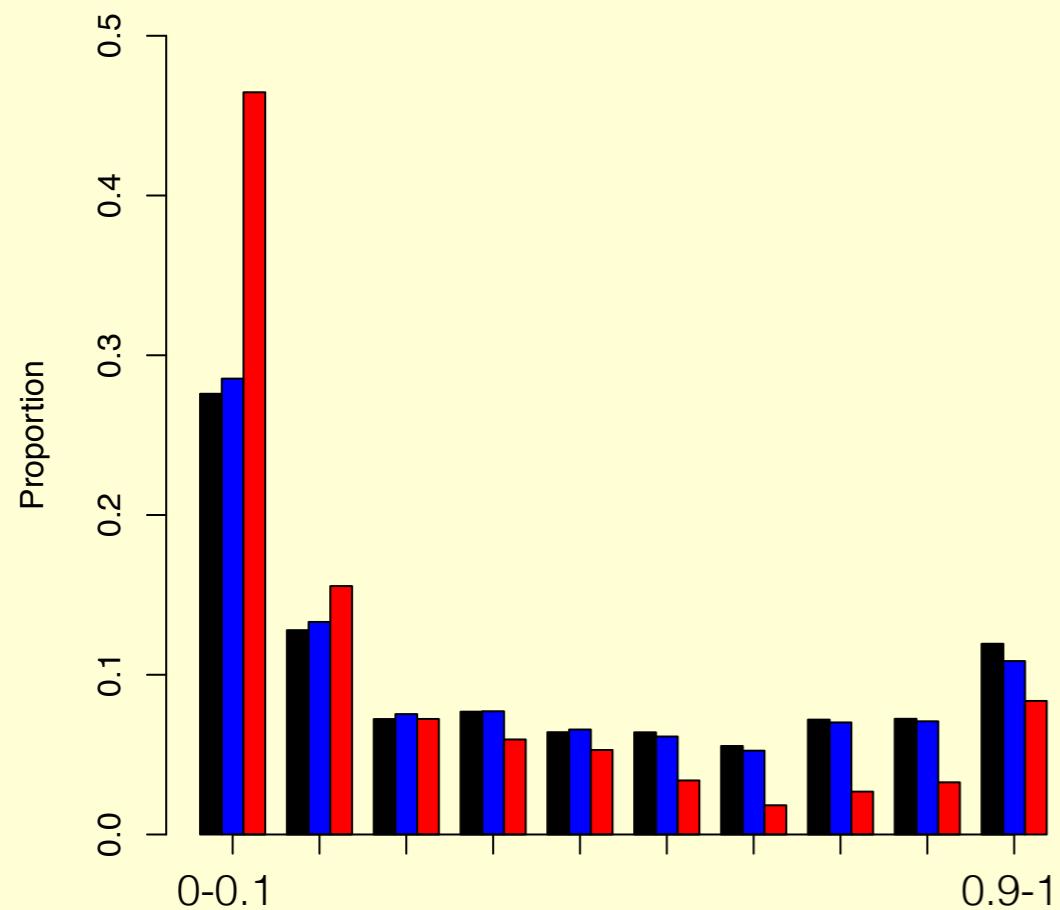
- Identify hundreds of putatively deleterious SNPs in protein coding regions.
  - Different approaches identify mostly non-overlapping set of SNPs as deleterious

Species	SIFT	PPH2	LRT	Intersect
Barley	3,400	3,295	3,221	<b>1,006</b>
Soybean	1,972	3,881	3,135	<b>784</b>

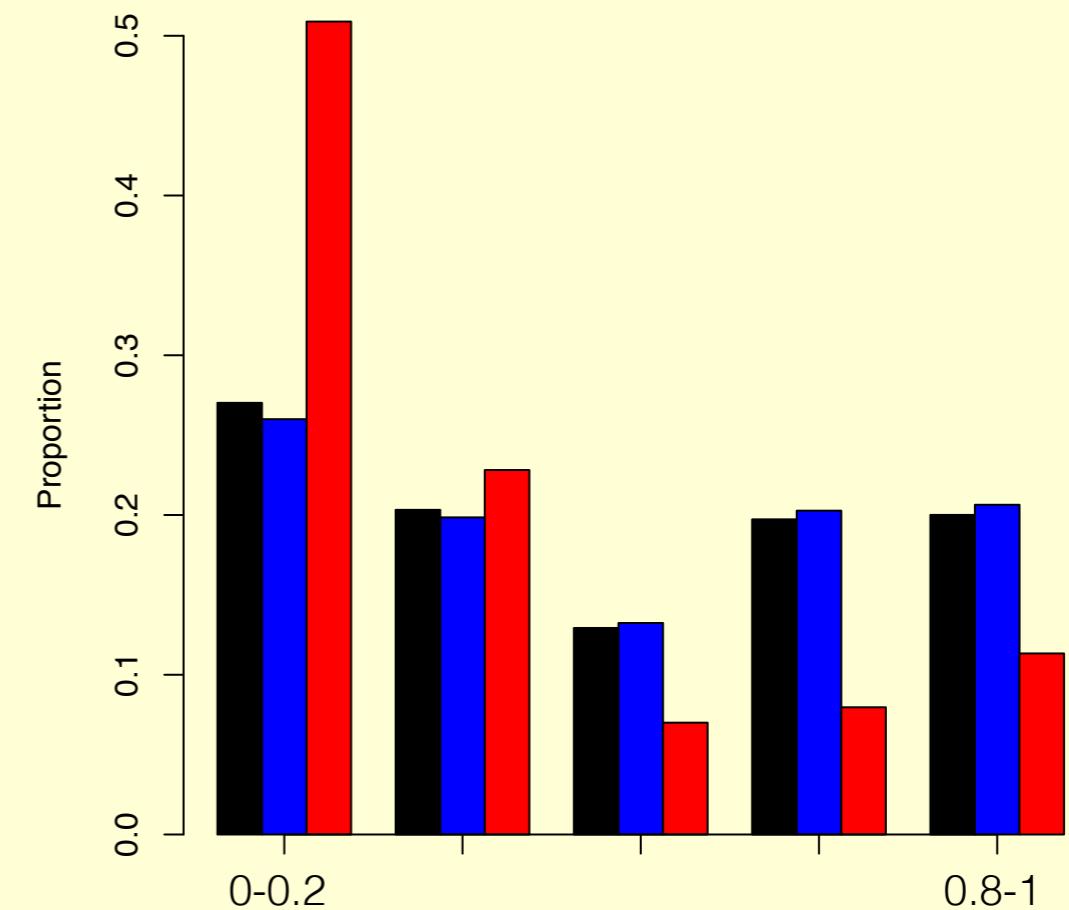
# Deleterious SNP Survey: Enrichment for Low Frequency

- Tend to be at lower frequencies than tolerated variants

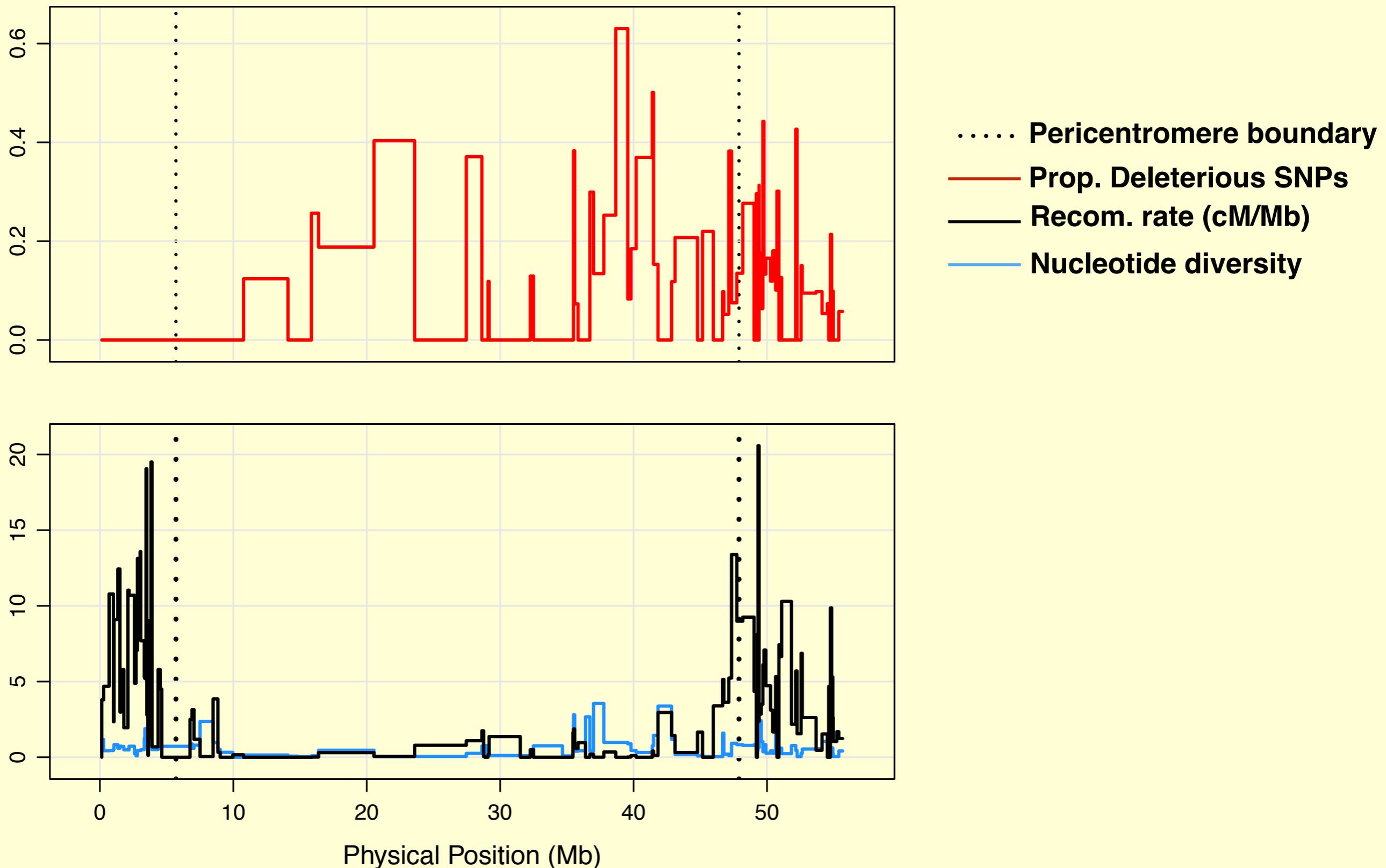
Barley



Soybean



# Deleterious SNP Survey: Enrichment in Low Recombination Regions



# Deleterious SNP Survey: Causative SNPs are "Deleterious"

- Causative variants tend to be called deleterious more frequently than those without *a priori* known phenotypic impacts

	Tolerated	Deleterious	Total
Causative	23 <b>(67.6%)</b>	11 <b>(32.4%)</b>	34
No Known Phenotype	29,259 <b>(94.2%)</b>	1,790 <b>(5.8%)</b>	31,046

# Chapter 2: Summary

- Approximately 800-1,000 deleterious SNPs segregate in this sample of two crop species.
  - Putatively deleterious SNPs accumulate in low-recombination regions.
- SNPs with *a priori* known phenotypic effects are more likely to be predicted deleterious than those without.

# Outline

- Collecting metadata on SNPs
- Survey of potentially deleterious SNPs in crop genomes
- Potentially deleterious SNPs in a barley breeding program

# Delete有害 SNPs and Phenotypes

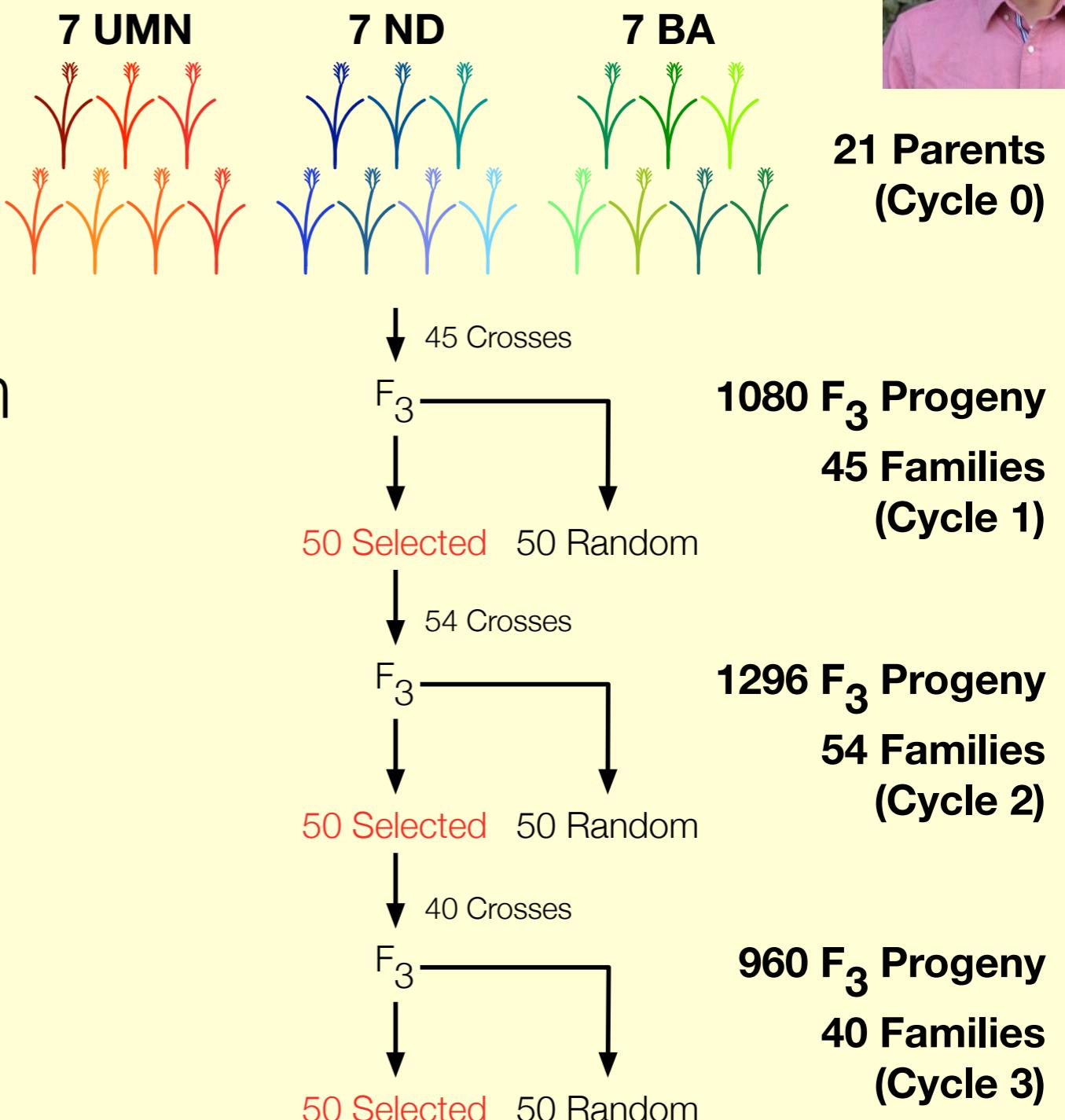
- Questions:
  - What proportion of variance for agronomic phenotypes do deleterious variants explain in barley, relative to tolerated variants?
  - Do deleterious variants decrease in frequency with selection for increased yield?

# Deleterious SNPs and Phenotypes: Barley Breeding Population

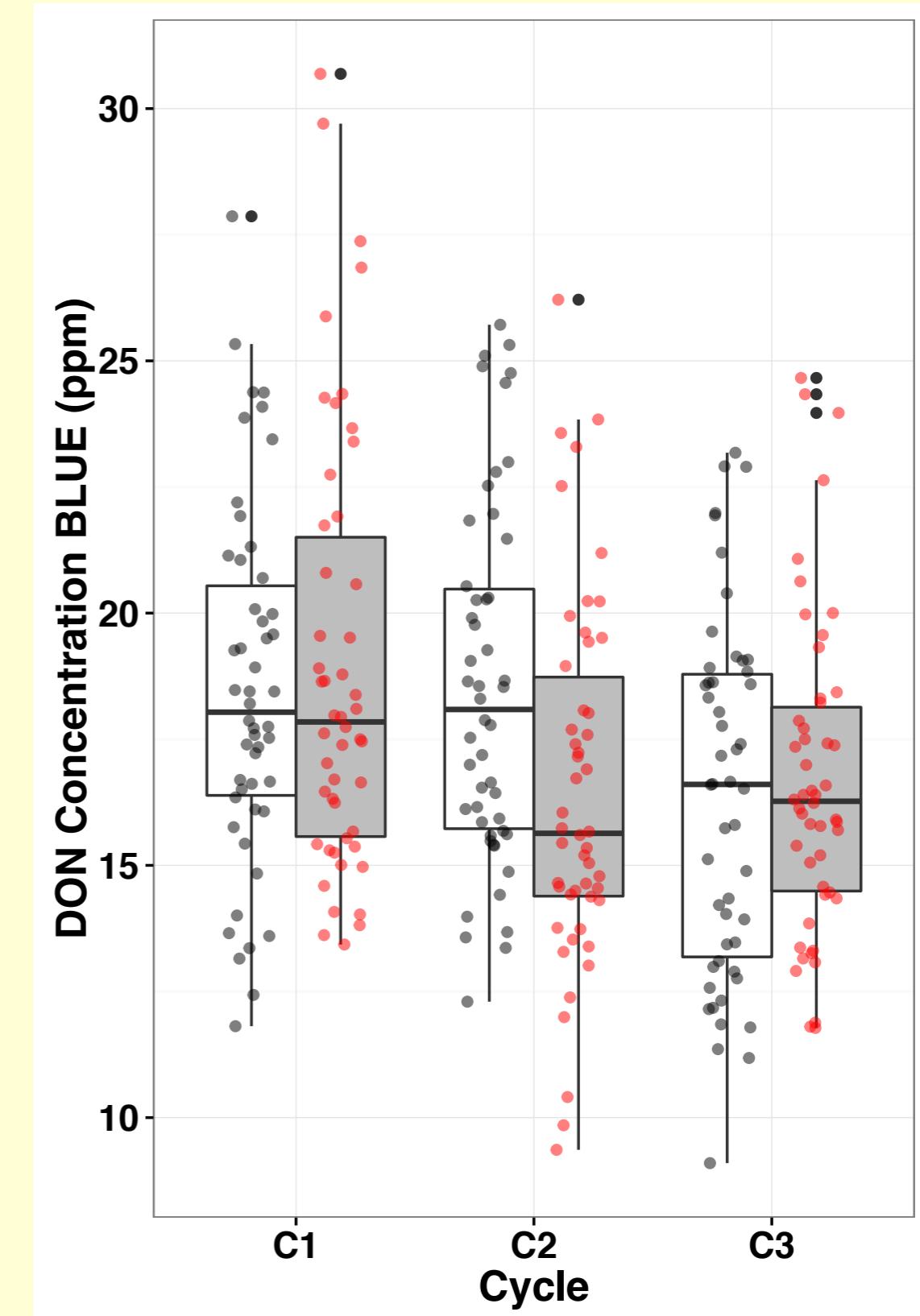
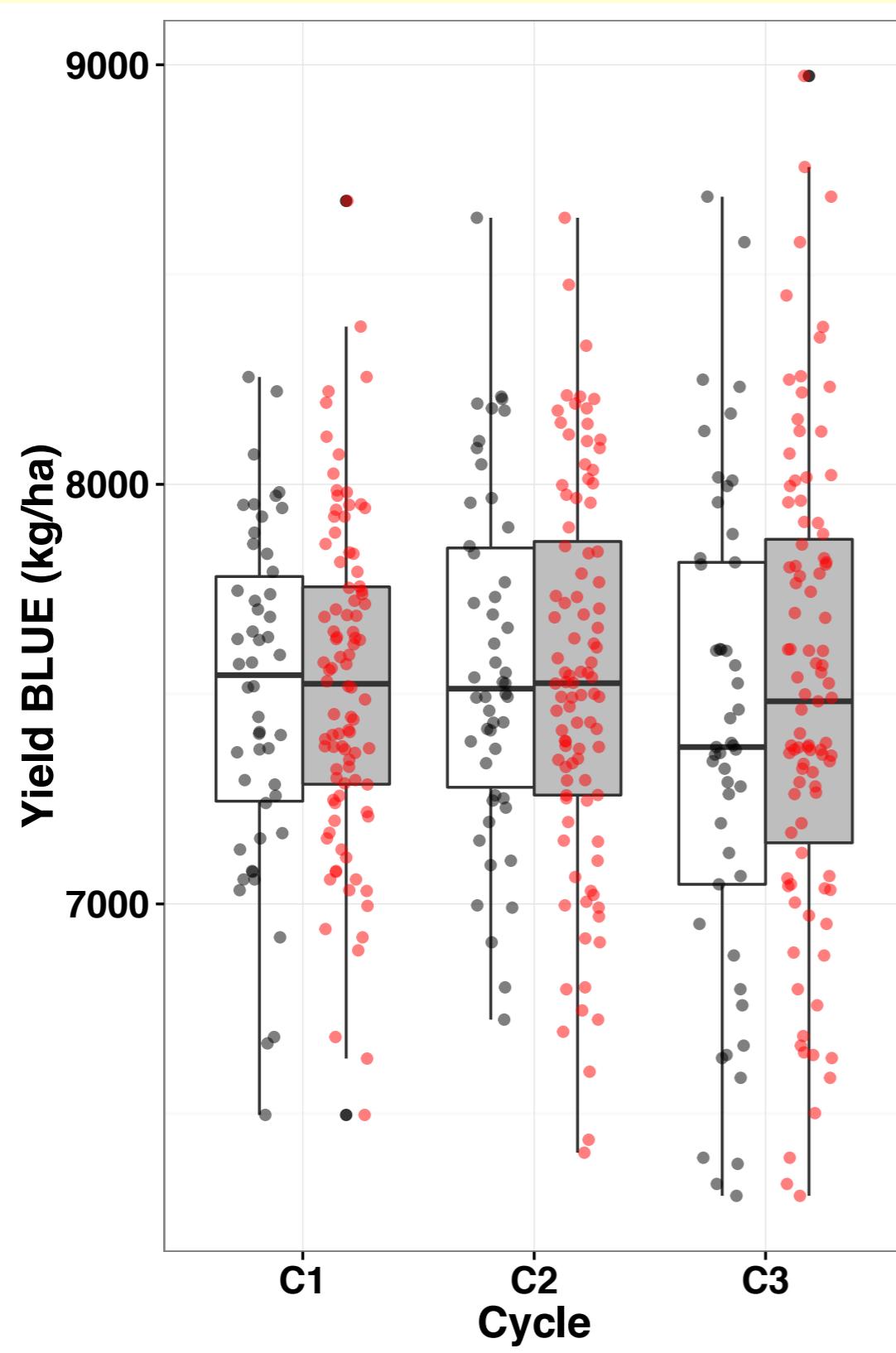
Tyler Tiede



- Parental exome sequenced
- Progeny genotyped with 384 SNP platform
- Evaluated for DON and grain yield

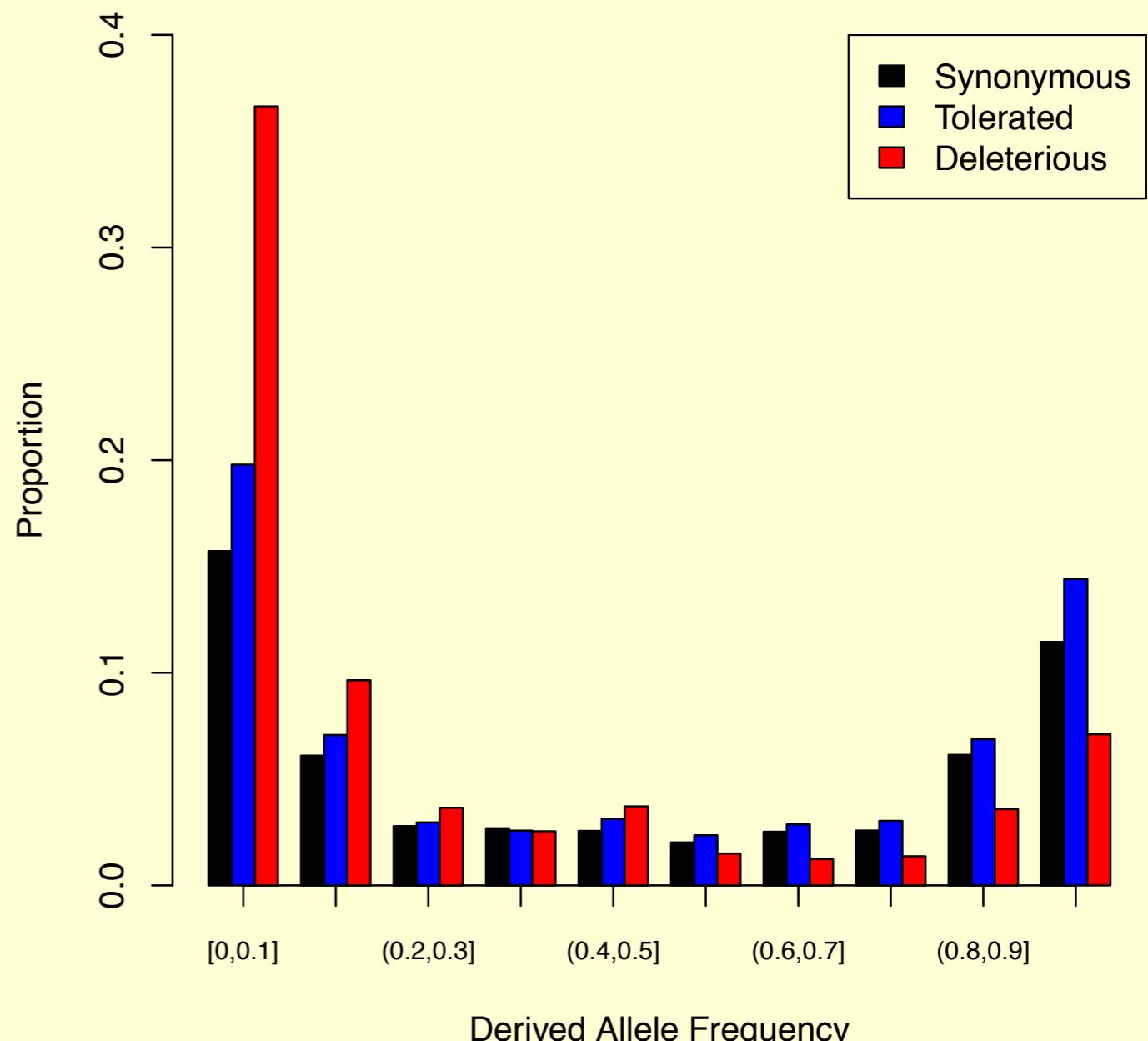


# Deleterious SNPs and Phenotypes: Phenotypes Respond to Selection



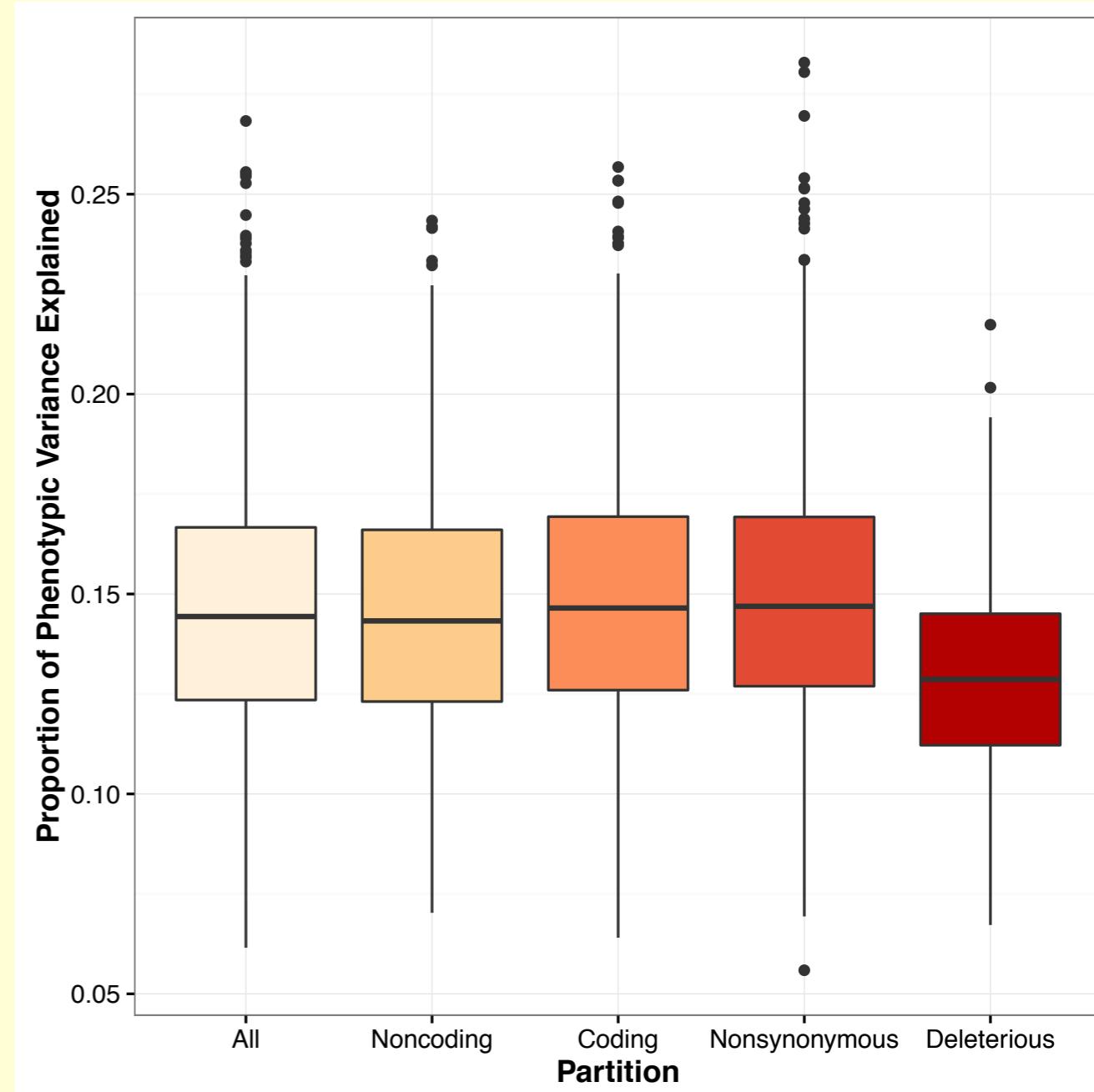
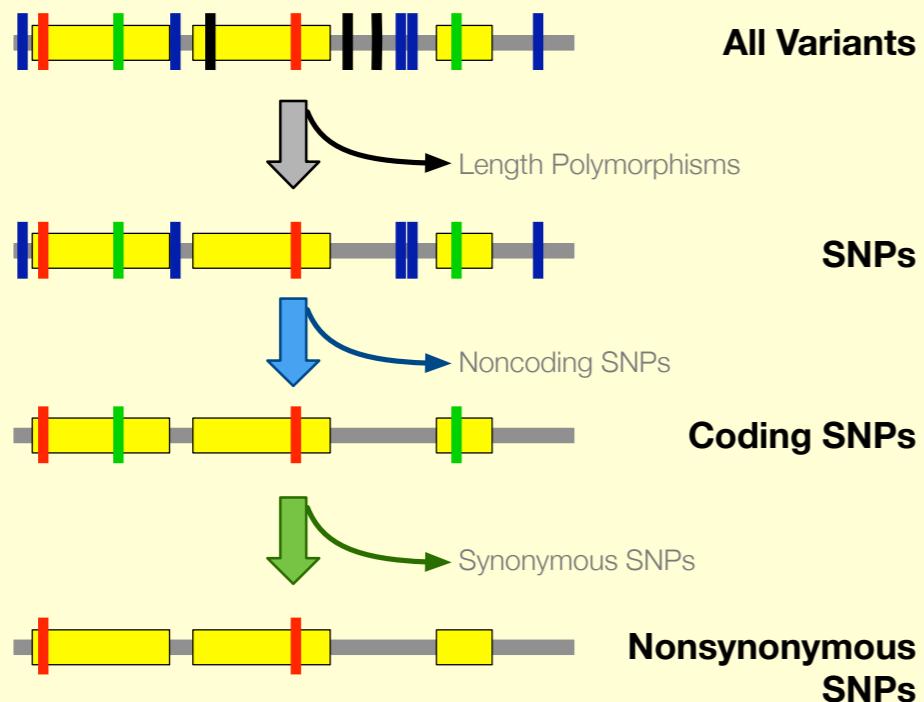
# Deleterious SNPs and Phenotypes: Polymorphic Deleterious SNPs

- Identify ~2,000 (including early stop) deleterious SNPs segregating in this population
- Tend to be at lower frequency than other classes of SNPs



# Deleteleterious SNPs and Phenotypes: Explaining Phenotypic Variance

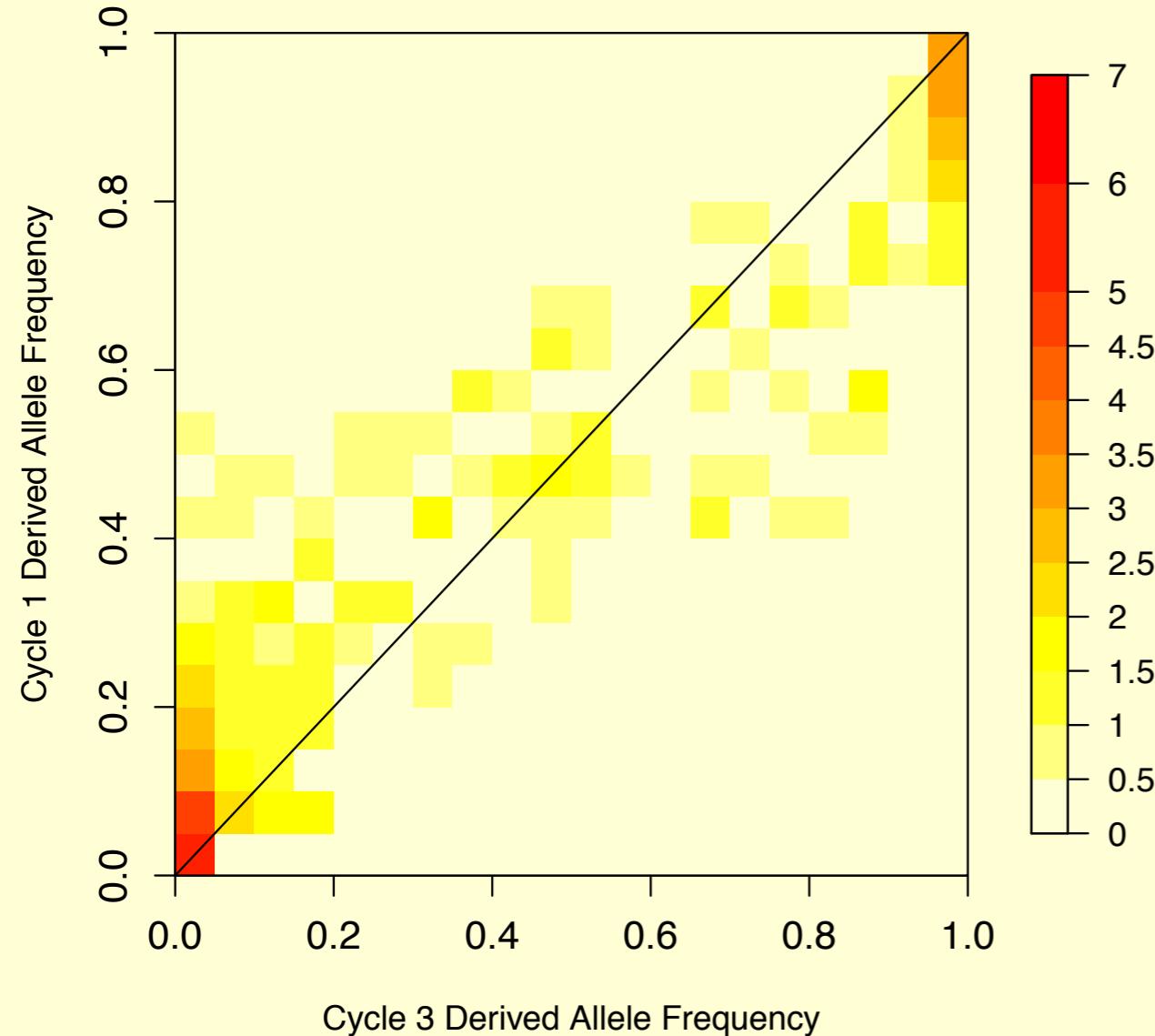
- Nonsynonymous SNPs explain more phenotypic variation on average than other classes of SNPs



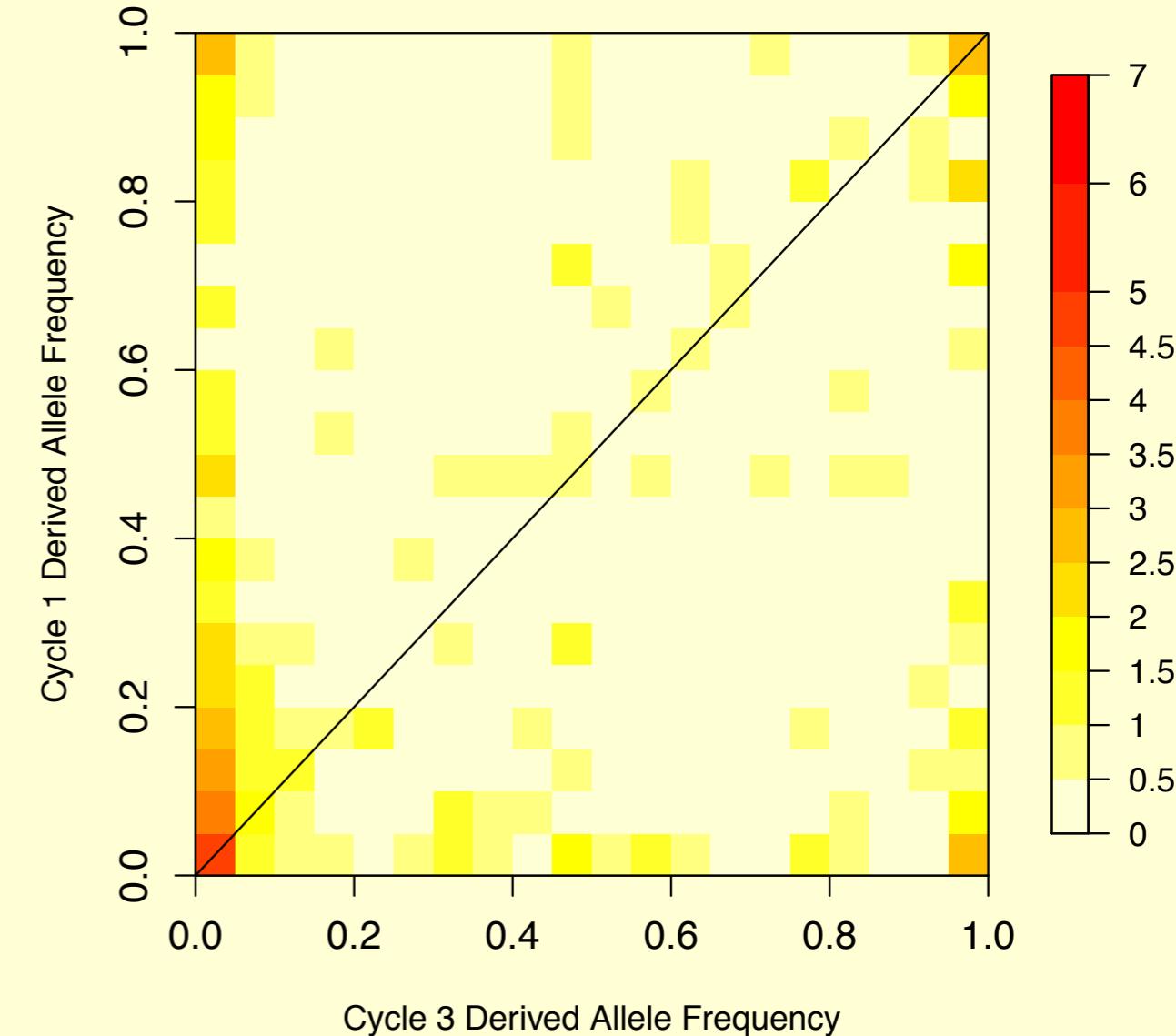
# Deleterious SNPs and Phenotypes: Frequency in Random and Selected Panels

## Deleterious SNPs

**Random Panel**

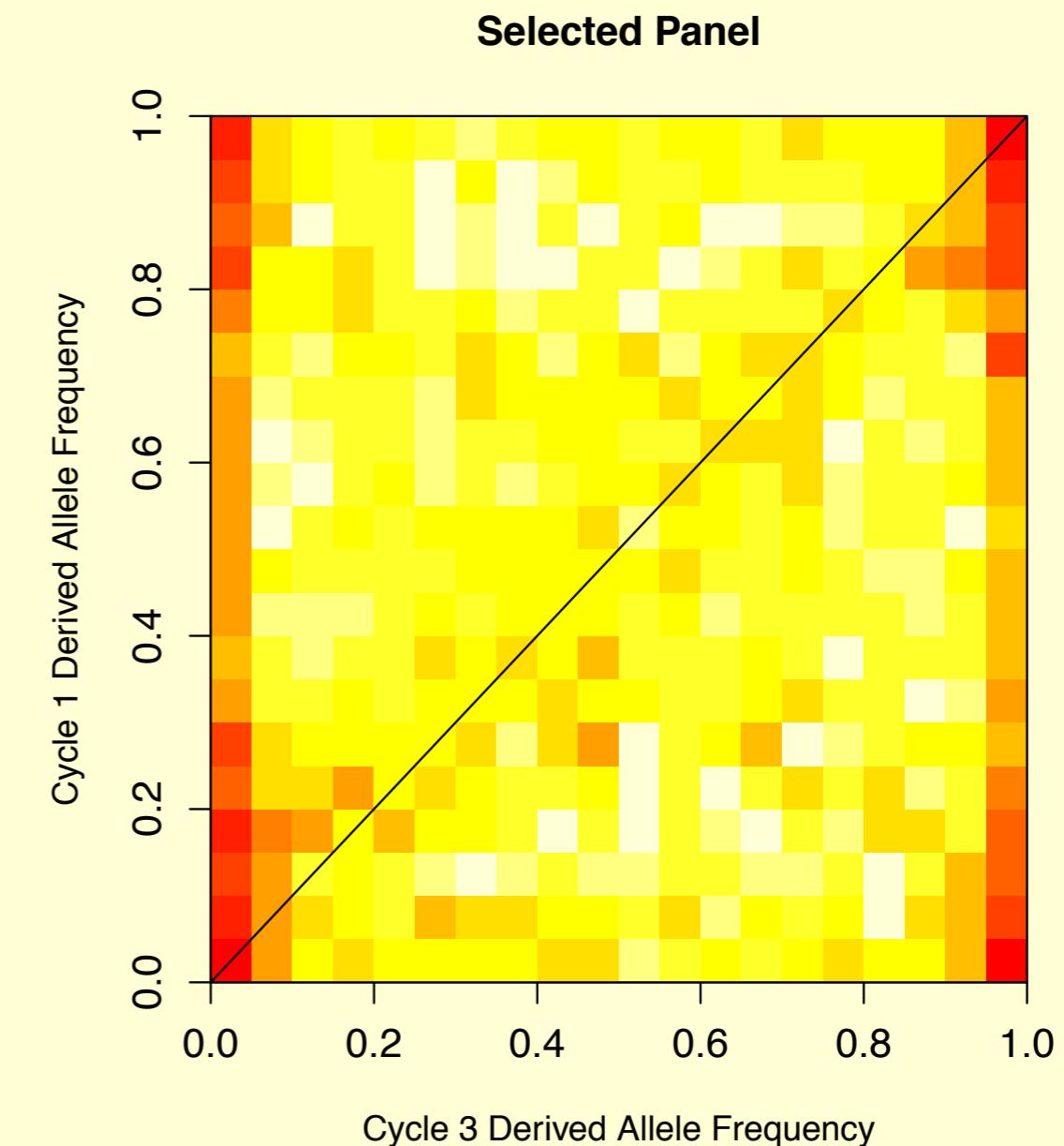
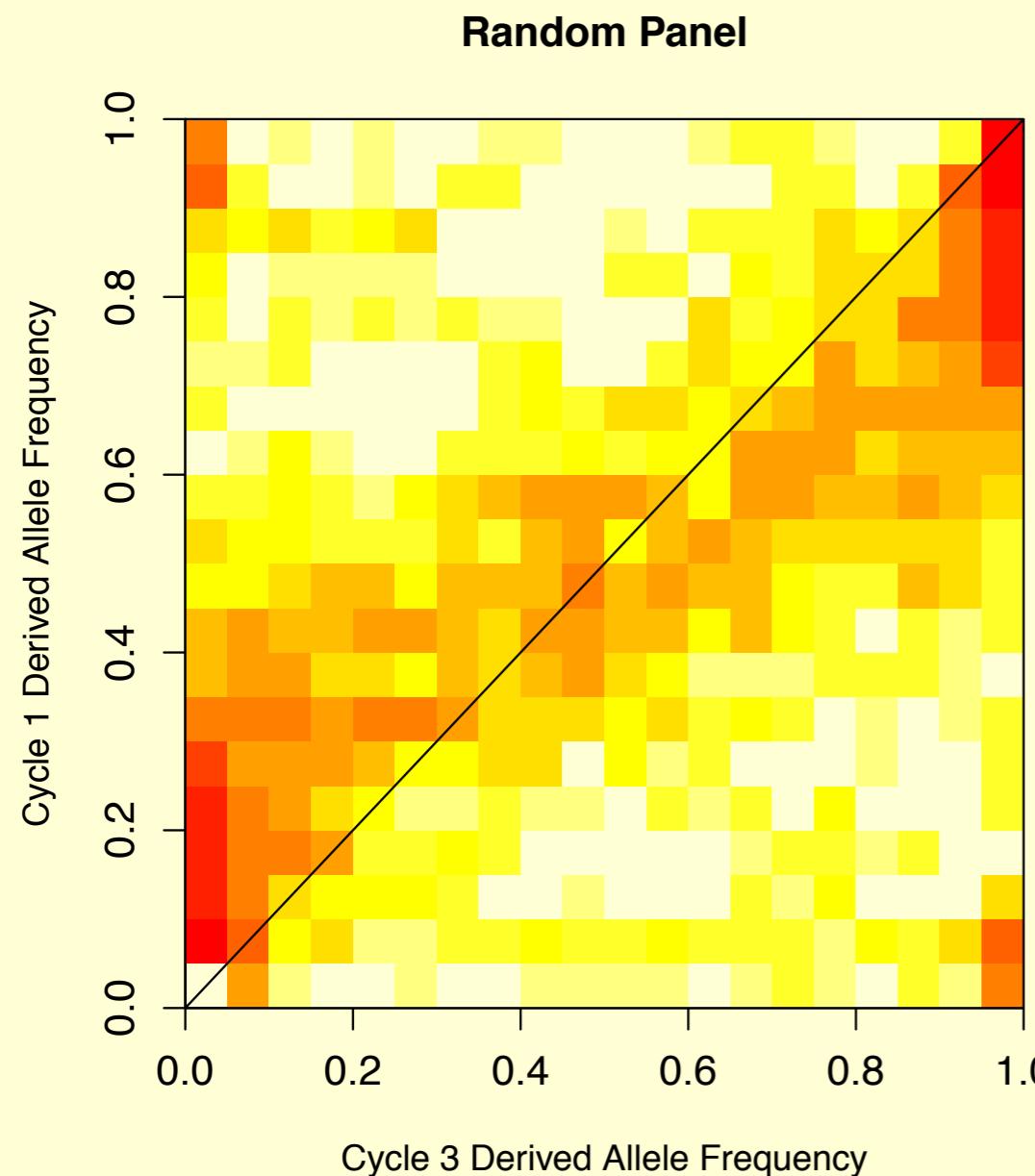


**Selected Panel**



# Deleterious SNPs and Phenotypes: Frequency in Random and Selected Panels

## Synonymous SNPs



# Deleterious SNPs and Phenotypes: Summary

- Deleterious SNPs explain less phenotypic variance than other categories of SNP, likely due to lower average minor allele frequency.
- Deleterious SNPs do decrease in frequency with selection for agronomic phenotypes, and it appears as though they decrease faster than synonymous SNPs.

# Overall Summary

- Querying public databases for SNP metadata in the absence of a reference genome
- New tool for predicting whether a SNP may be deleterious that addresses data quality and reference bias issues. Large effect SNPs more likely to be predicted as deleterious.
- Putatively deleterious SNPs do not explain more phenotypic variance than tolerated SNPs, but informative subsets may improve predictions.

# Acknowledgements

## Committee:

Peter Morrell  
Robert Stupar  
Ruth Shaw  
Kevin Smith  
Peter Tiffin

## Collaborators:

Justin Fay  
Mohsen Mohammadi  
Jesse Poland  
Amber Eule-Nashoba  
Kiran Seth

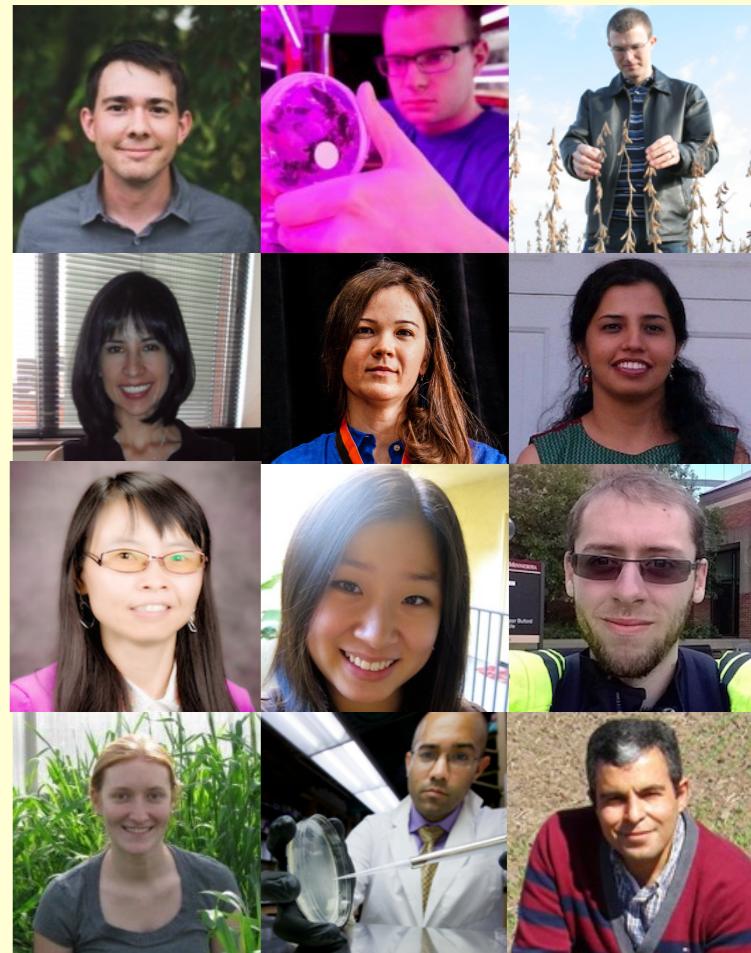
## Funding:

USDA NIFA National Needs Fellowship  
MnDRIVE Global Food Ventures  
UMN Doctoral Dissertation Fellowship



## Lab Members:

Li Lei  
Ana Poets  
Paul Hoffman  
Chaochih Liu  
Ashley Rozmarin  
Fengli Fu  
Michael Kantar  
Justin Anderson  
Benjamin Campbell  
Jean-Michel Michno



# Resources

## SNP Metadata:

SNPMeta program code

<https://github.com/MorrellLAB/SNPMeta>

Annotations:

<http://conservancy.umn.edu/handle/11299/181367>

### MOLECULAR ECOLOGY RESOURCES

Molecular Ecology Resources (2014) 14, 419–425

doi: 10.1111/1755-0998.12183

## SNPMeta: SNP annotation and SNP metadata collection without a reference genome

THOMAS J. Y. KONO,\* KIRAN SETH,\* JESSE A. POLAND†‡ and PETER L. MORRELL\*

\*Department of Agronomy & Plant Genetics, University of Minnesota, 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA, †Hard Winter Wheat Genetics Research Unit, USDA ARS, 2021 Claflin Road, 4008 Throckmorton Hall, Kansas State University, Manhattan, KS 66506, USA, ‡Department of Agronomy, 2021 Claflin Road, 2004 Throckmorton Plant Sciences Center, Kansas State University, Manhattan, KS 66506, USA

# Resources

## Deleterious SNP Survey:

BAD\_Mutations program code

[https://github.com/MorrellLAB/BAD\\_Mutations](https://github.com/MorrellLAB/BAD_Mutations)

Sequence analysis pipeline

[https://github.com/MorrellLAB/Deleterious\\_Mutations](https://github.com/MorrellLAB/Deleterious_Mutations)

Sequence variants

<http://conservancy.umn.edu/handle/11299/181502>

## The Role of Deleterious Substitutions in Crop Genomes

Thomas J. Y. Kono,\*<sup>1</sup> Fengli Fu,<sup>1</sup> Mohsen Mohammadi,<sup>1,2</sup> Paul J. Hoffman,<sup>1</sup> Chaochih Liu,<sup>1</sup> Robert M. Stupar,<sup>1</sup> Kevin P. Smith,<sup>1</sup> Peter Tiffin,<sup>3</sup> Justin C. Fay,<sup>4</sup> and Peter L. Morrell\*,<sup>1</sup>

<sup>1</sup>Department of Agronomy and Plant Genetics, University of Minnesota

<sup>2</sup>Department of Agronomy, Purdue University

<sup>3</sup>Department of Plant Biology, University of Minnesota

<sup>4</sup>Department of Genetics, Washington University

# Resources

## **Deleterious SNPs and Phenotypes:**

Analysis scripts

[https://github.com/MorrellLAB/Deleterious\\_GP](https://github.com/MorrellLAB/Deleterious_GP) (Private until preprint/pub)

# Resources

**Slide Deck:**

<http://>