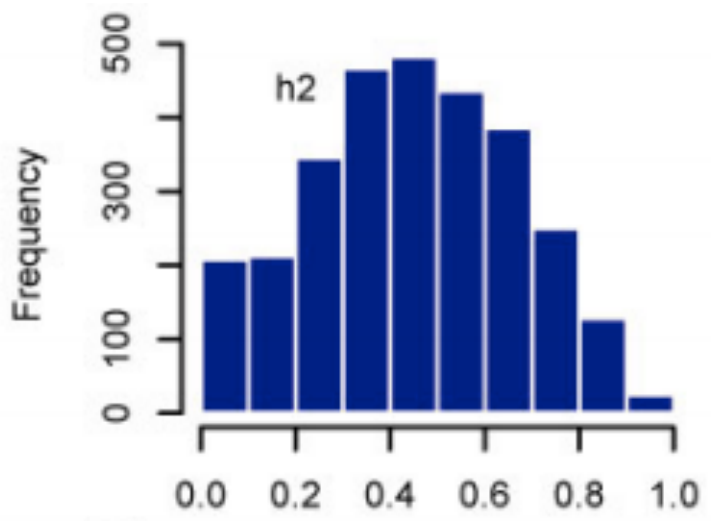
COMPLEX TRAITS AND DELETERIOUS ALLELES IN HUMAN POPULATIONS

Evolution, Winter 2017

Can we discover the genetic basis of heritable traits?

Twin studies revealed that many phenotypes have a genetic component

- Height: 0.68-0.93
- Fasting blood glucose: 0.38-0.66
- RNA expression levels: Most genes > 0.1 and < 0.5



Silventoinen *et al* 2003 (*Twin Res*), Wright *et al* 2014 (*Nature Genetics)*
Simonis-Bik *et al* 2008 *(Twin Res Hum Genet)*, Polderman *et al* 2015 *(Nature Genetics)*

Hypothesis: common-phenotype, common-variant

- Suppose a trait is relatively common (e.g., asthma)

Hypothesis: common-phenotype, common-variant

- Suppose a trait is relatively common (e.g., asthma)

- Suppose the trait has a genetic component

  ATGCTATCCTATCAAACTATCTA

  ATGCTATCCTATCAAAGTATCTA -> G *is a risk allele*

Hypothesis: common-phenotype, common-variant

- Suppose a trait is relatively common (e.g., asthma)

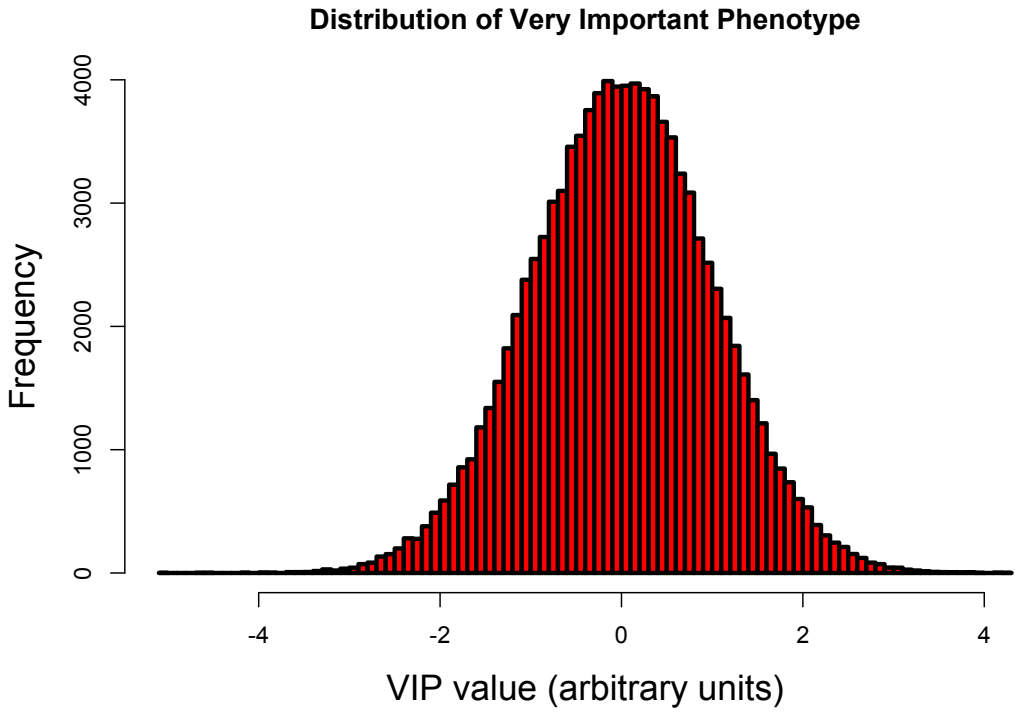
- Suppose the trait has a genetic component

  ATGCTATCCTATCAAACTATCTA

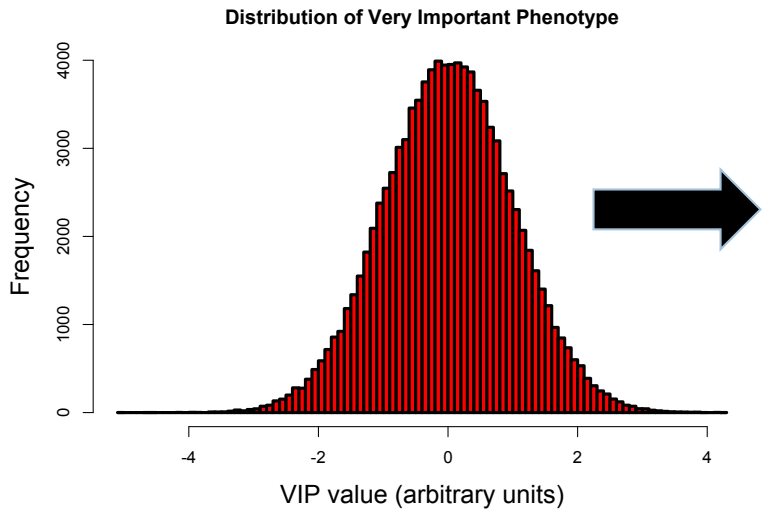
  ATGCTATCCTATCAAAGTATCTA -> G *is a risk allele*

- "Reasonable" hypothesis: **If a hereditary trait is common, its genetic basis must also be common**

“Explaining” phenotypic variation

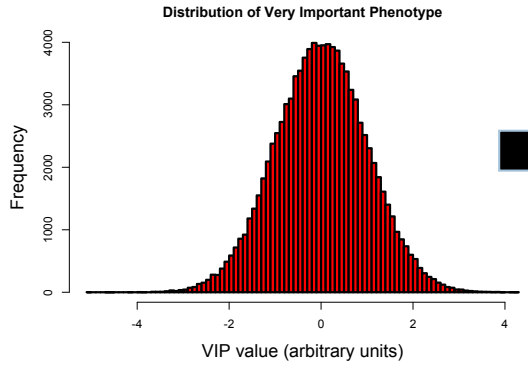


“Explaining” phenotypic variation

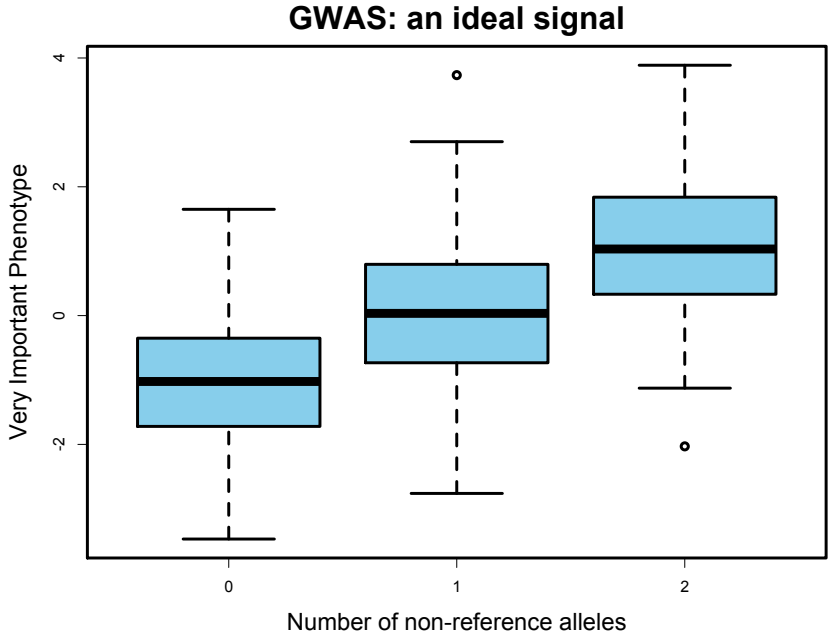


IND. →	i1	i2	...	i100
SNP	A/A	A/T	...	A/T
↓	G/T	T/T	...	G/G
	A/T	A/A	...	A/A
	A/C	C/C	...	A/C
	T/C	T/T	...	T/C
	G/G	A/A	...	G/A

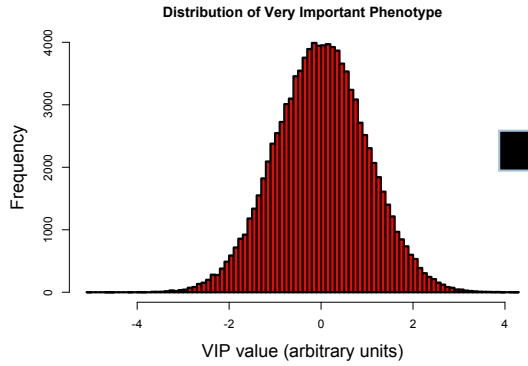
“Explaining” phenotypic variation



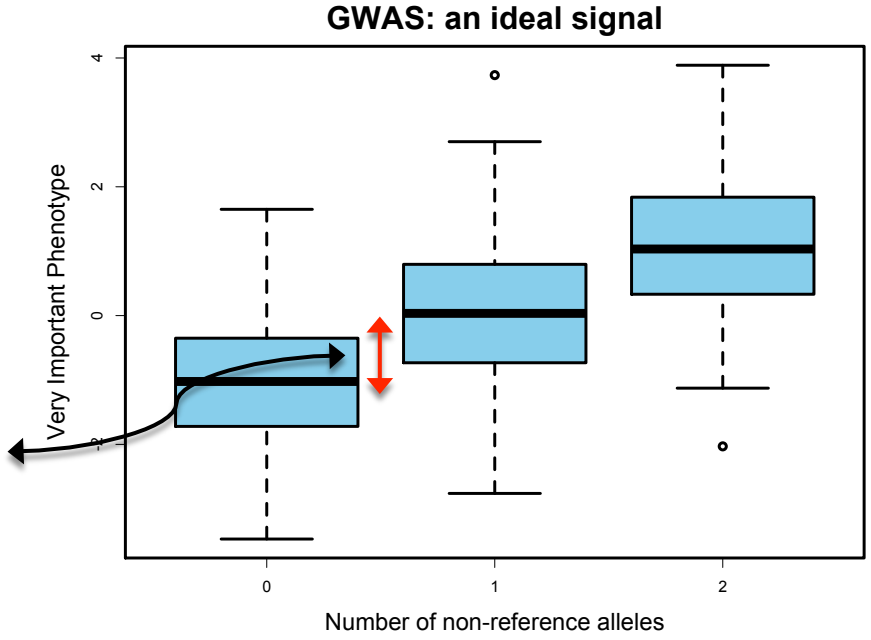
IND. →	i1	i2	...	i100
SNP	A/A	A/T	...	A/T
↓	G/T	T/T	...	G/G
	A/T	A/A	...	A/A
	A/C	C/C	...	A/C
	T/C	T/T	...	T/C
	G/G	A/A	...	G/A



“Explaining” phenotypic variation

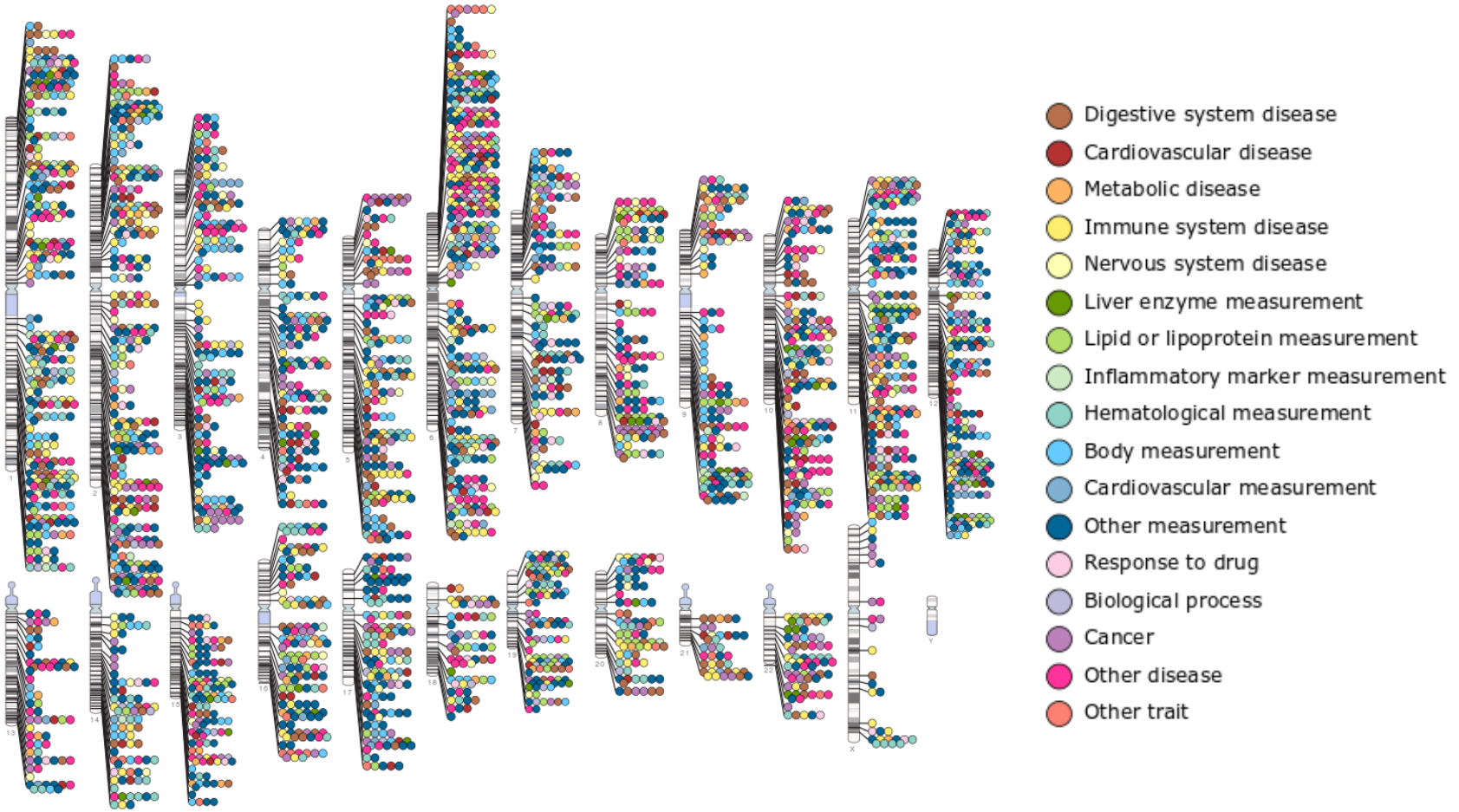


IND. →	i1	i2	...	i100
SNP	A/A	A/T	...	A/T
↓	G/T	T/T	...	G/G
	A/T	A/A	...	A/A
	A/C	C/C	...	A/C
	T/C	T/T	...	T/C
	G/G	A/A	...	G/A



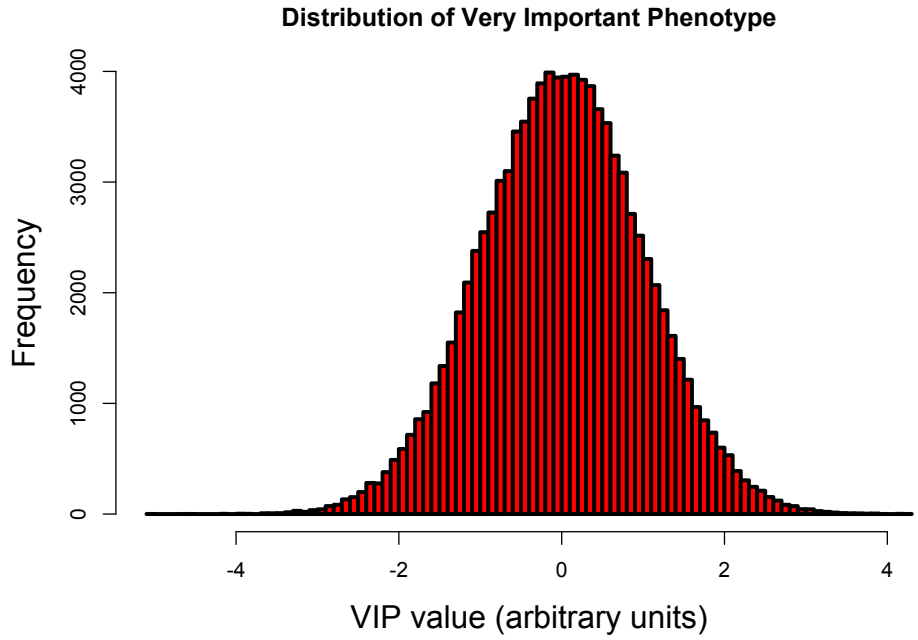
The difference in trait values between genotypes is called the effect size, **Z**

Many loci that contribute to trait variation have been discovered with GWAS

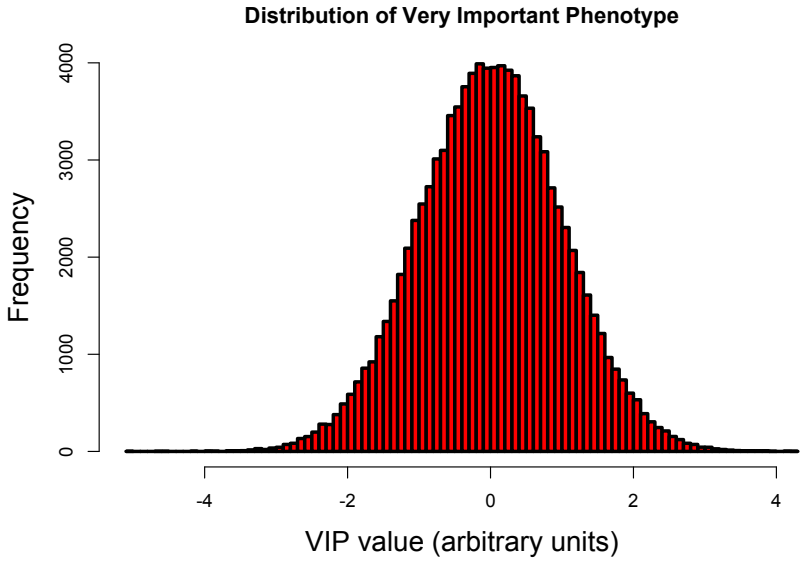


Obtained from http://www.ebi.ac.uk/fgpt/gwas/

But do known associations explain variance in traits?



“Missing” heritability - calculating variance accounted for by GWAS



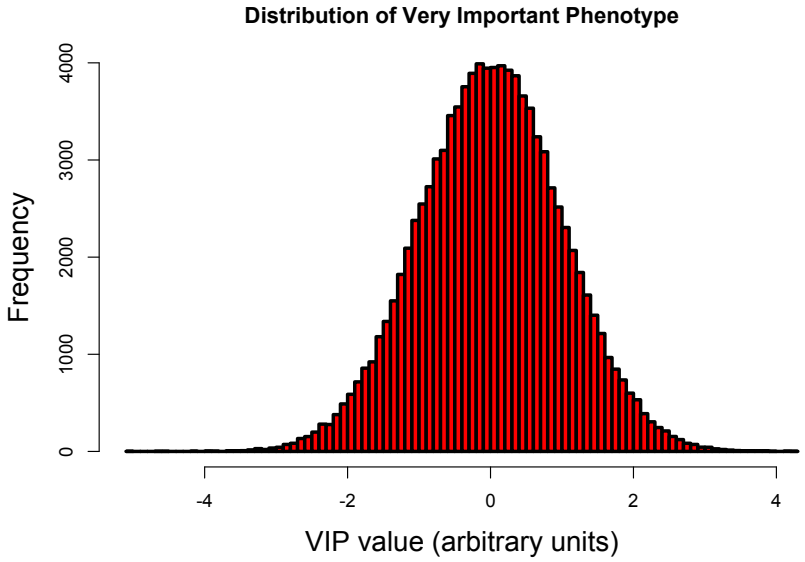
Suppose *k* variants are found to be associated with VIP…

Contribution from each SNP $v = \frac{1}{2} z^2 x(1-x)$

Total variance from GWAS $V_{\text{GWAS}}(P) = \sum_k v_k$

Compare to heritability $V_{\text{GWAS}}(P) \stackrel{?}{=} h^2 \times V(P)$

“Missing” heritability - calculating variance accounted for by GWAS



Suppose *k* variants are found to be associated with VIP…

Contribution from each SNP $v = \frac{1}{2} z^2 x(1-x)$

Total variance from GWAS $V_{\text{GWAS}}(P) = \sum_k v_k$

Compare to heritability $V_{\text{GWAS}}(P) < h^2 \times V(P)$

Potential sources of "missing" heritability

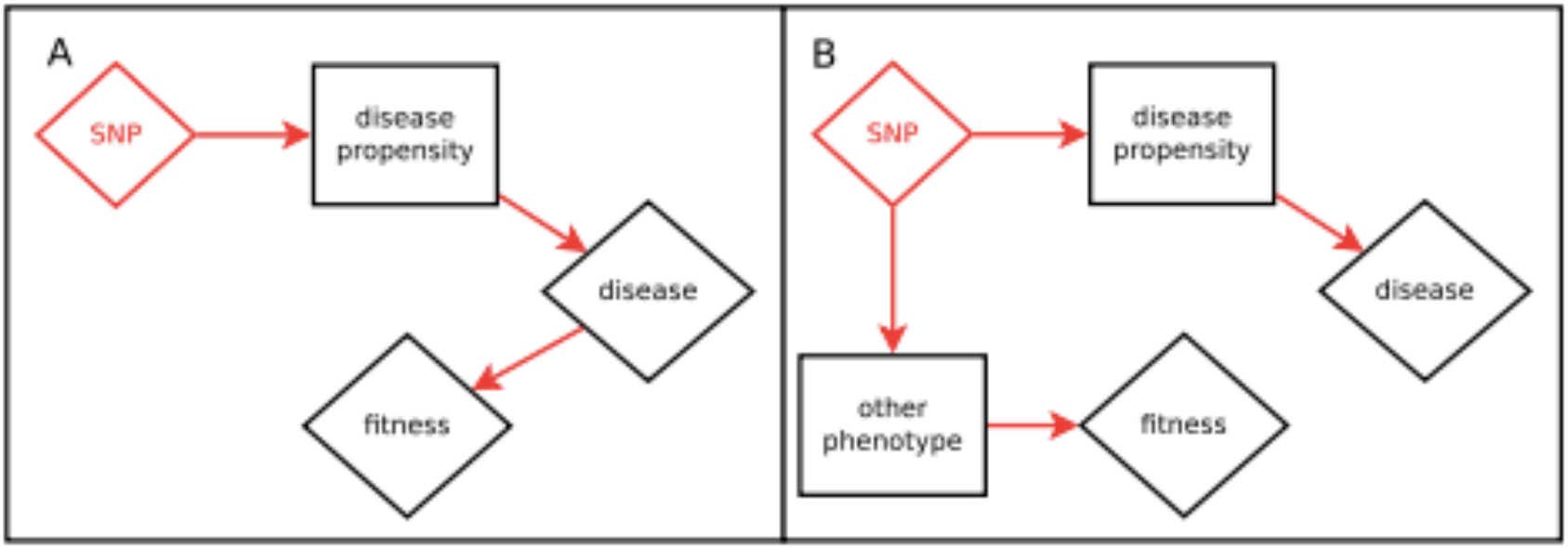
- Bias in heritability estimates
- Interactions (between groups of genes, or genes and the environment)
- Common variants of small effect
- Structural variants
- Rare variants of large effect

Potential sources of "missing" heritability

- Bias in heritability estimates
- Interactions (between groups of genes, or genes and the environment)
- Common variants of small effect
- Structural variants
- Rare variants of large effect

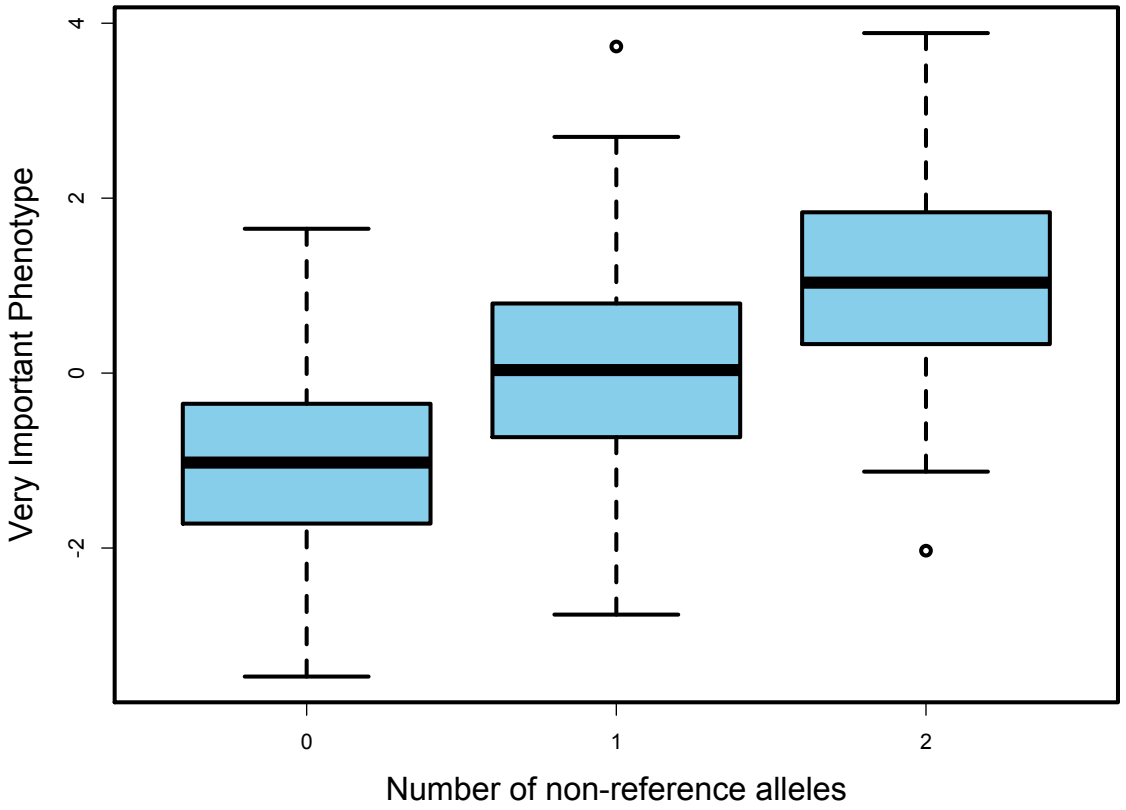
$$v = \frac{1}{2} z^2 x(1-x)$$

Why might we expect rare alleles to matter for some phenotypes?

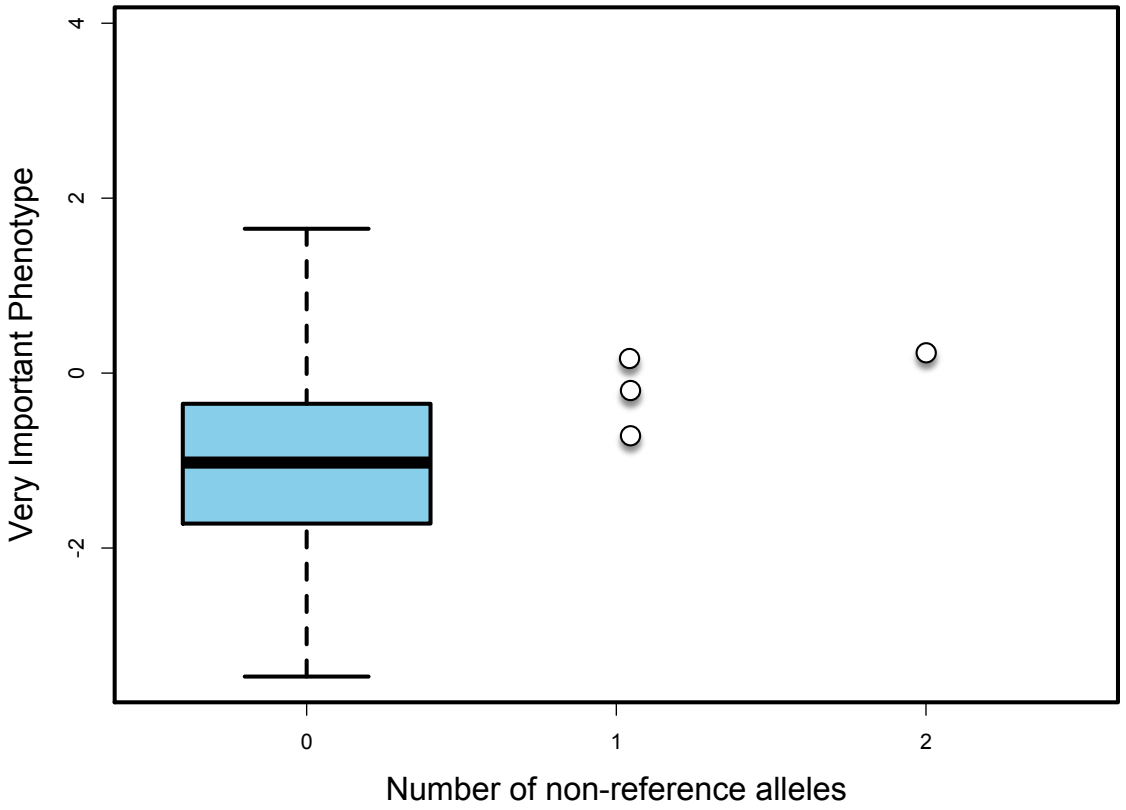


King *et al* 2010 (*Plos Genet*)

But it's hard to detect causal rare alleles…



But it's hard to detect causal rare alleles…



We will…

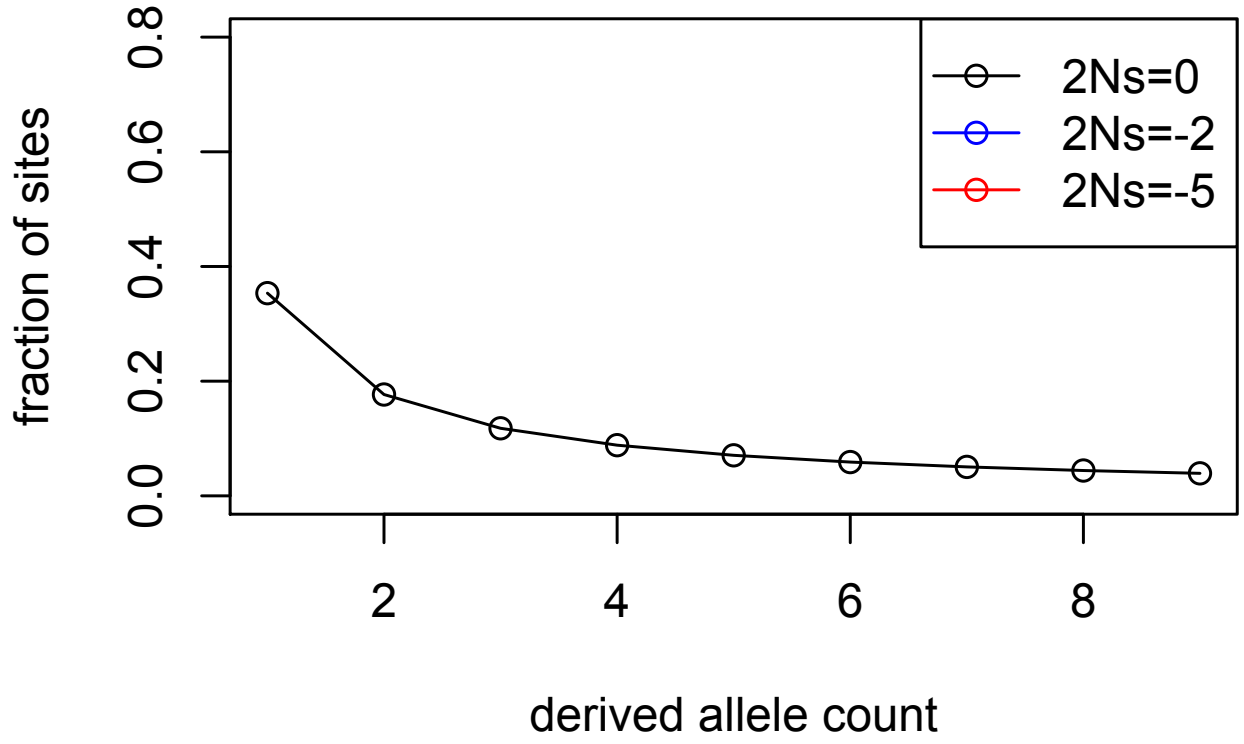
- Consider whether rare alleles (as a group) can ever contribute to complex trait variation
- Introduce recent literature on deleterious allele frequencies and “load” in humans
- Discuss whether there is empirical evidence for differences in “deleterious allele load” between human populations
- Summarize current thinking on contribution (or lack thereof) of rare alleles to complex traits

Quick refresher on evolutionary "forces"

- Mutation -> random introduction of new alleles
- Drift -> allele frequencies can randomly change over time
- Selection -> biases changes in allele frequency

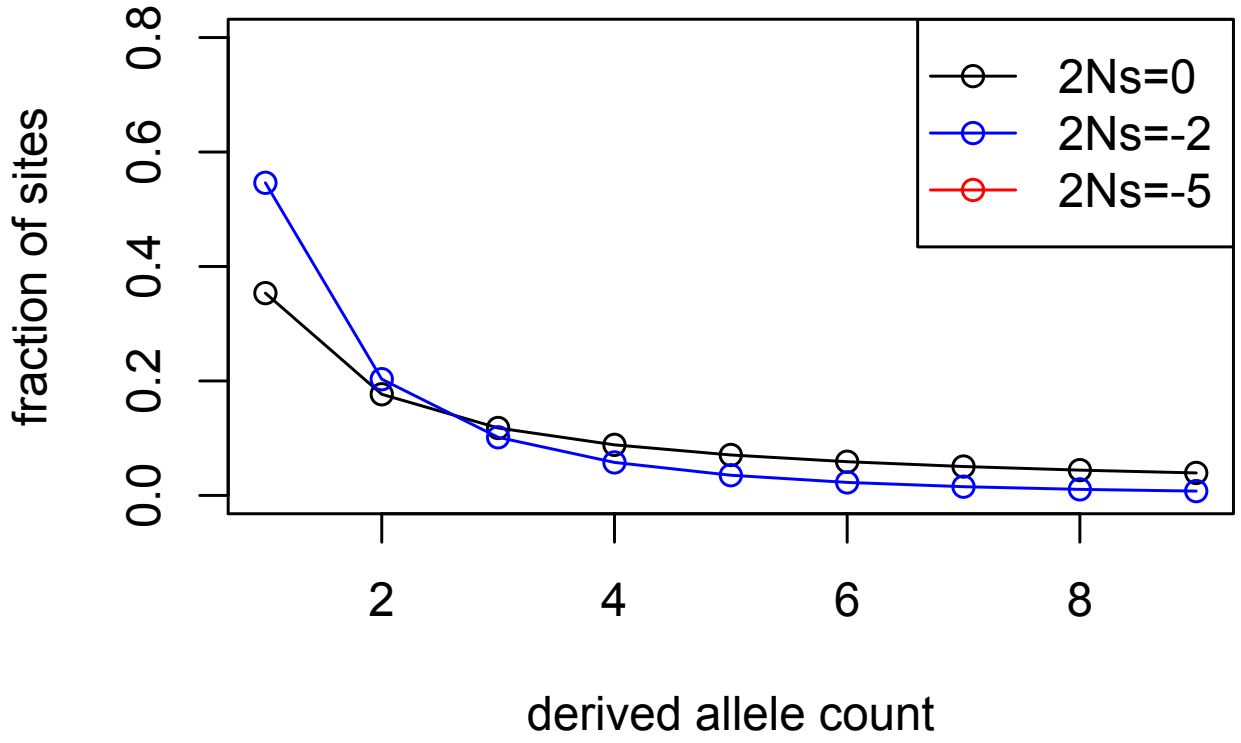
Mutation-selection-drift balance

- Suppose we sequence 5 diploid individuals
- For each variable site we observe, there can be from 1 to 9 copies of each derived allele in the population



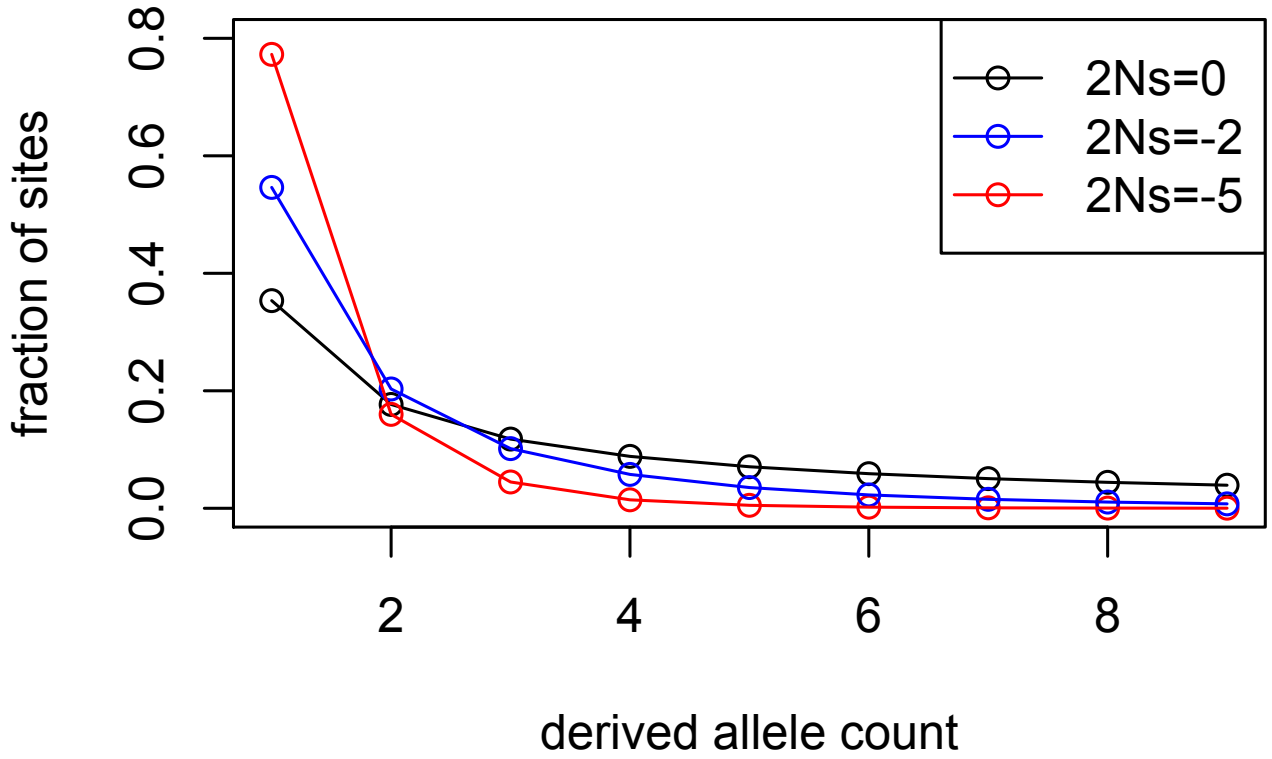
Mutation-selection-drift balance

- Suppose we sequence 5 diploid individuals
- For each variable site we observe, there can be from 1 to 9 copies of each derived allele in the population



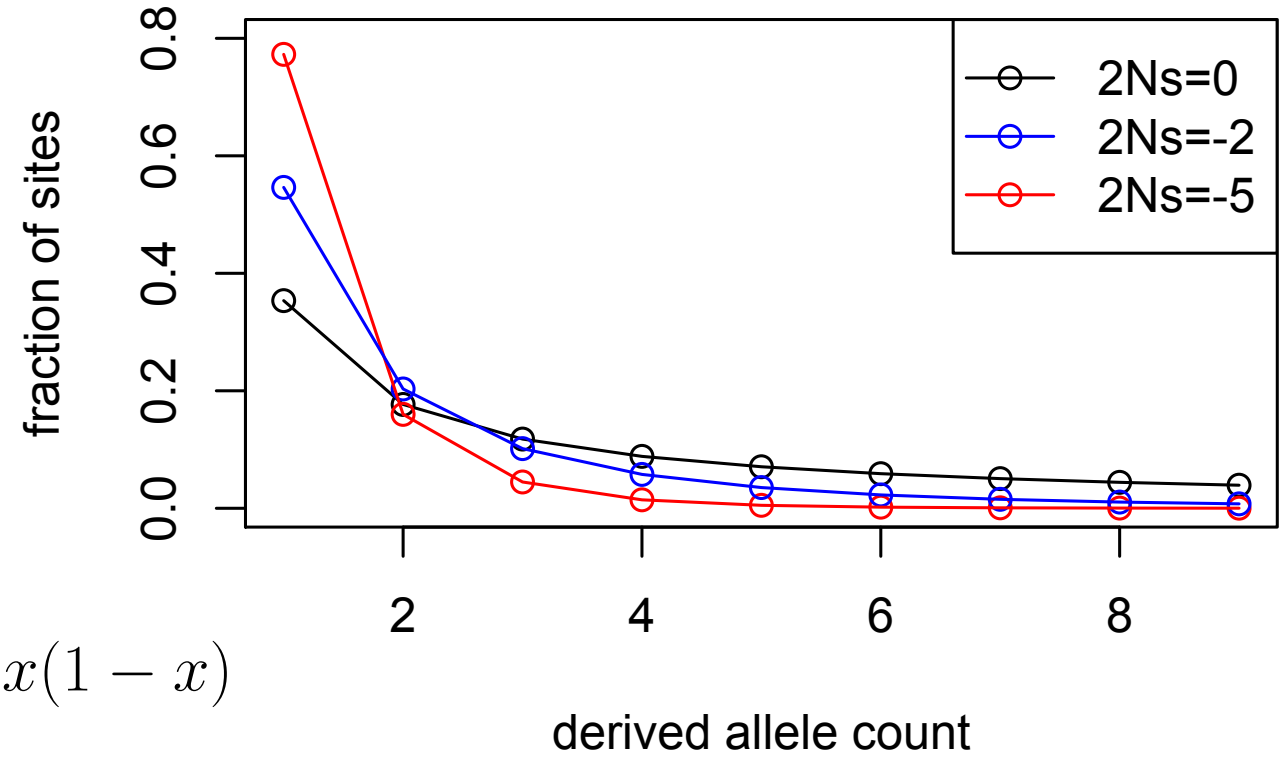
Mutation-selection-drift balance

- Suppose we sequence 5 diploid individuals
- For each variable site we observe, there can be from 1 to 9 copies of each derived allele in the population



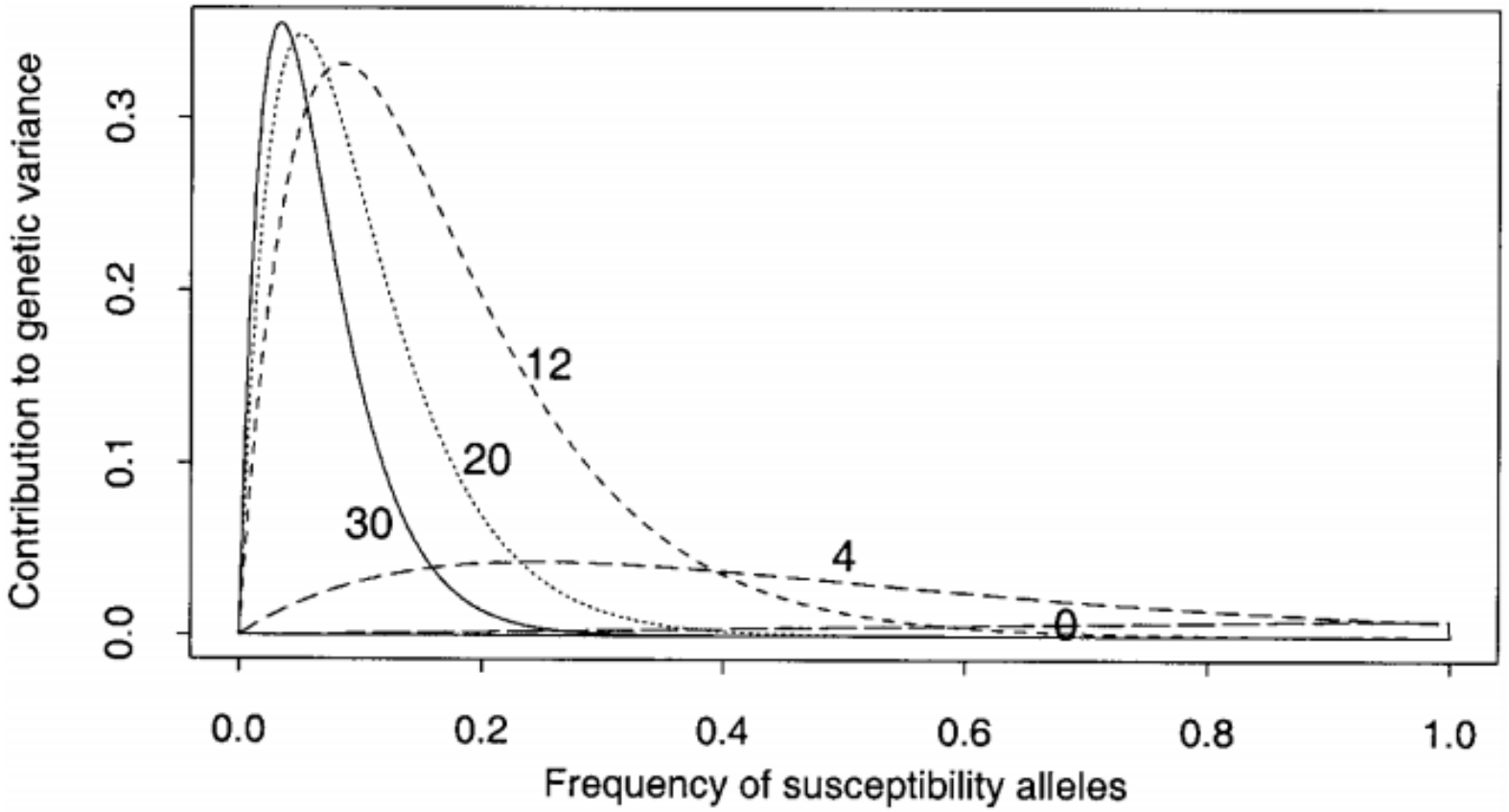
Mutation-selection-drift balance

- Suppose we sequence 5 diploid individuals
- For each variable site we observe, there can be from 1 to 9 copies of each derived allele in the population



$$v = \frac{1}{2} z^2 x(1-x)$$

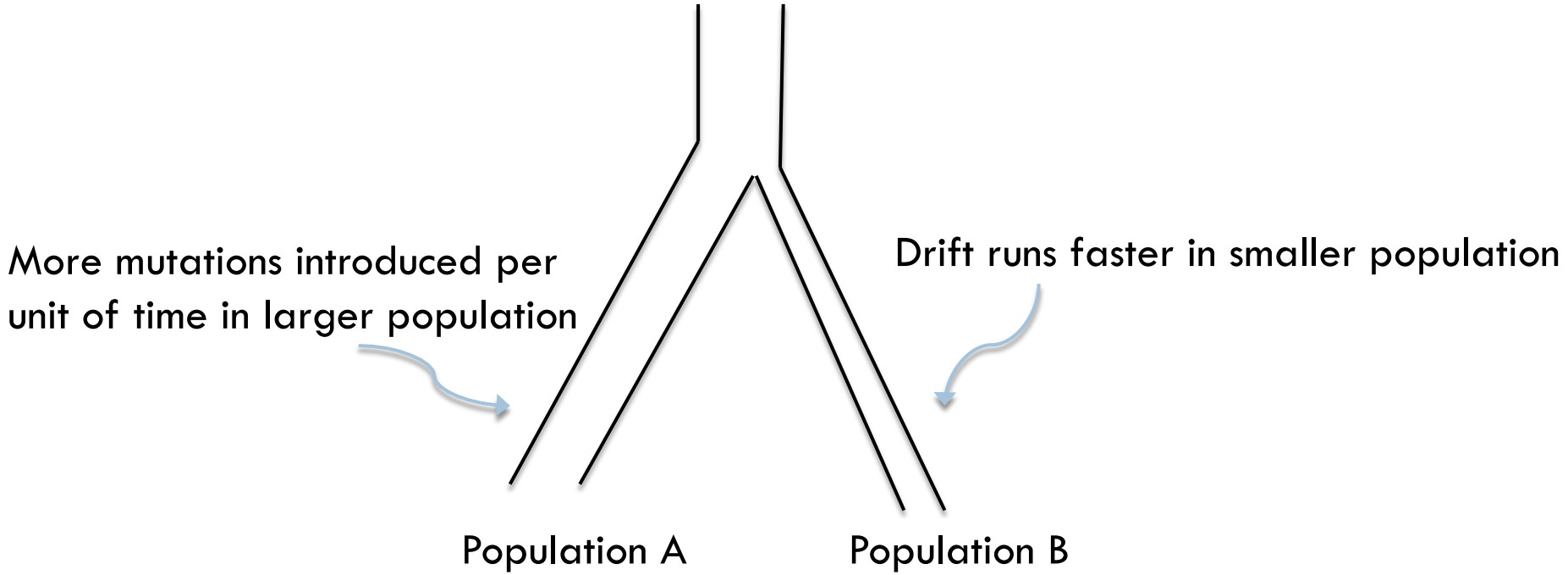
Common-disease common-variant… or not?



Pritchard & Cox 2002 (*Hum. Mol. Genet.*)
Pritchard 2001 (*AJHG*)

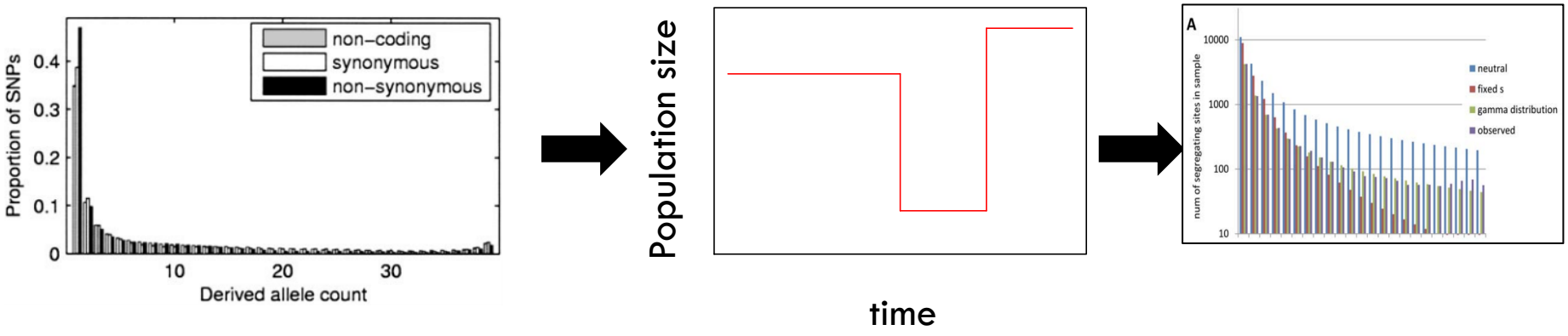
How strong is selection against human polymorphism?

- Approach: compare site frequency spectra at NS and SYN sites, under assumption that SYN sites are neutral
- Complication: population sizes also affects frequency spectrum



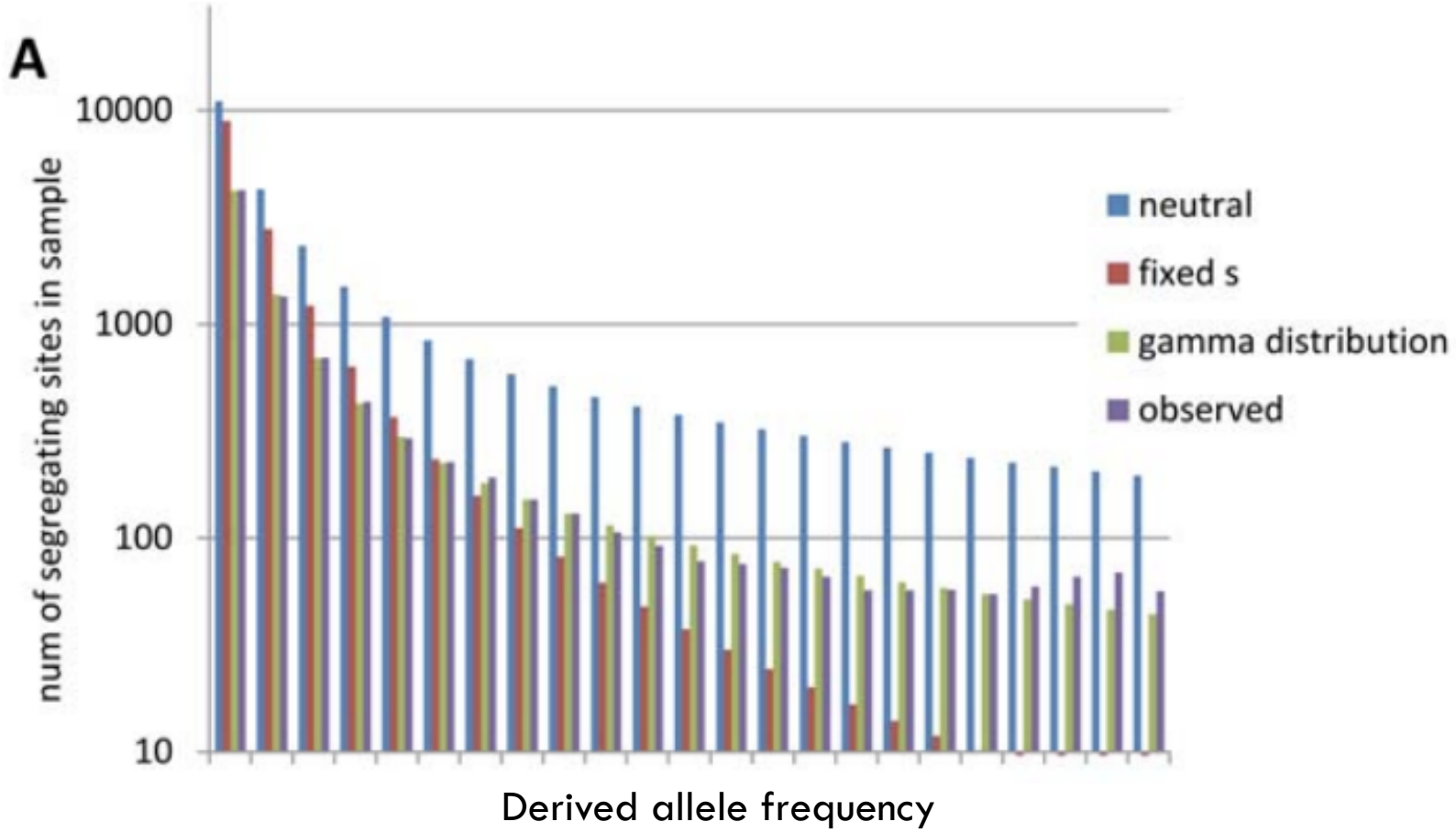
How strong is selection against human polymorphism?

- Time-dependence of demography/selection in Wright-Fisher model is not a solved problem
- Algorithmic approaches were developed to numerically compute the SFS as a function of time



Williamson *et al* 2005 (PNAS)
Boyko *et al* 2008 (Plos Genetics)
Keightley & Eyre-Walker 2007 (Genetics)

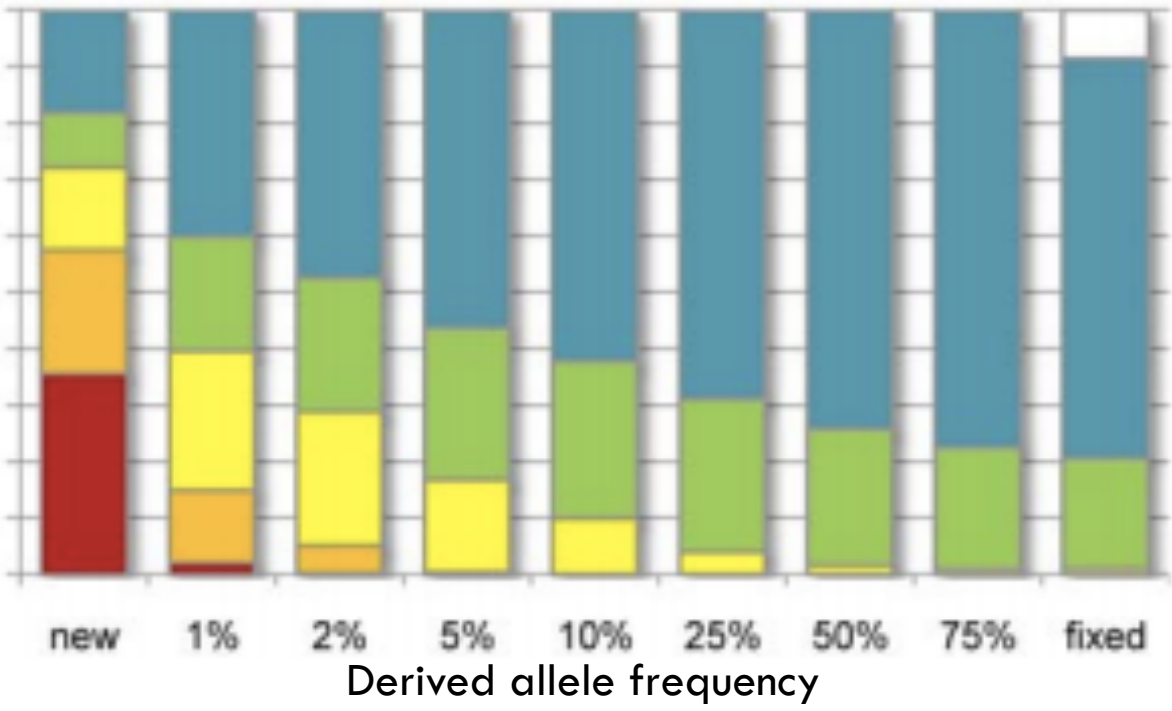
How strong is selection against human polymorphism?



Boyko *et al* 2008 (Plos Genetics)

How strong is selection against human polymorphism?

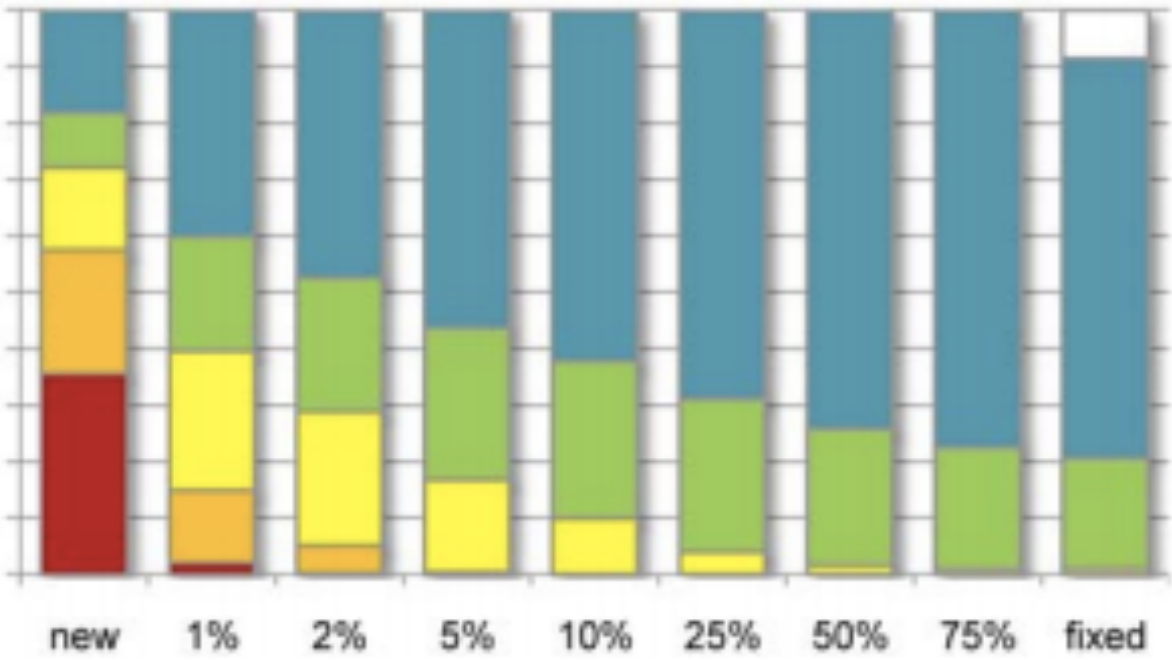
Red s<-0.01
Brown -0.01< s <-0.001
Yellow -0.001 < s< -0.0001
Etc…



Boyko *et al* 2008 (Plos Genetics)

How strong is selection against human polymorphism?

Red s<-0.01
Brown -0.01< s <-0.001
Yellow -0.001 < s< -0.0001
Etc…



Consequently, it is reasonable to suppose that correlations between effect size and allele frequency may exist and increase the role of rare alleles in complex traits

Boyko *et al* 2008 (Plos Genetics)

Requirements for a rare allele contribution

- Trait is heritable
- Large mutation rate for causal alleles (*i*.e., many loci genome wide that alter the trait)
- Coupling between allelic selection strength and effect on trait

Do deleterious allele frequencies differ across populations?

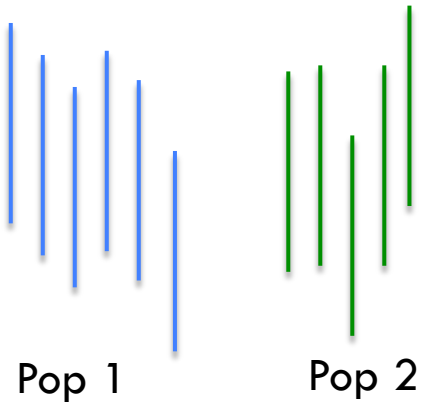
- Common alleles: almost entirely shared across human populations
- Rare alleles: less likely to be true
- Reasonable hypothesis: demographic events have altered deleterious allele frequencies in human populations relative to each other

Lohmueller *et al* 2008 (*Nature*)

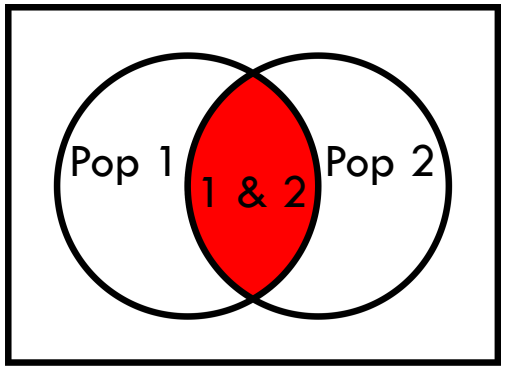
Do deleterious allele frequencies differ across populations?

- Common alleles: almost entirely shared across human populations
- Rare alleles: less likely to be true
- Reasonable hypothesis: demographic events have altered deleterious allele frequencies in human populations relative to each other

Sequence 2 human populations with different ancestral histories



Classify damaging (NS) alleles as private or shared



Compare proportion of private alleles in each population

$F_{priv}(1) > F_{priv}(2)$?

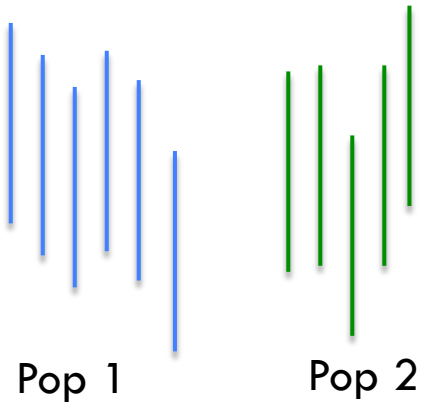
$F_{priv}(1) = F_{priv}(2)$?

$F_{priv}(1) < F_{priv}(2)$?

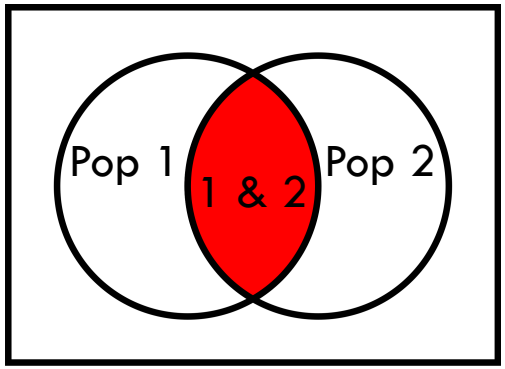
Do deleterious allele frequencies differ across populations?

- Common alleles: almost entirely shared across human populations
- Rare alleles: less likely to be true
- Reasonable hypothesis: demographic events have altered deleterious allele frequencies in human populations relative to each other

Sequence 2 human populations with different ancestral histories

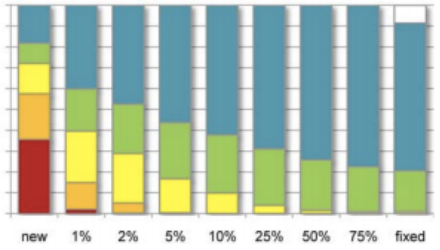


Classify damaging (NS) alleles as private or shared



Compare proportion of private alleles in each population

$F_{priv}(1) > F_{priv}(2)$



Do deleterious allele frequencies differ across populations?

- Common alleles: almost entirely shared across human populations
- Rare alleles: less likely to be true
- Reasonable hypothesis: demographic events have altered deleterious allele frequencies in human populations relative to each other

Table 1 | Distribution of Applera SNPs by population and functional class

Category	Shared	Private AA	Private EA	Mean derived frequency	
				AA*	EA†
Synonymous	8,056 (58.3%)	8,958 (53.0%)	3,879 (44.6%)	0.211	0.266
Non-synonymous	5,771 (41.7%)	7,950 (47.0%)	4,826 (55.4%)	0.174	0.202

Lohmueller *et al* 2008 (*Nature*)

Do deleterious allele frequencies differ across populations?

- Common alleles: almost entirely shared across human populations
- Rare alleles: less likely to be true
- Reasonable hypothesis: demographic events have altered deleterious allele frequencies in human populations relative to each other

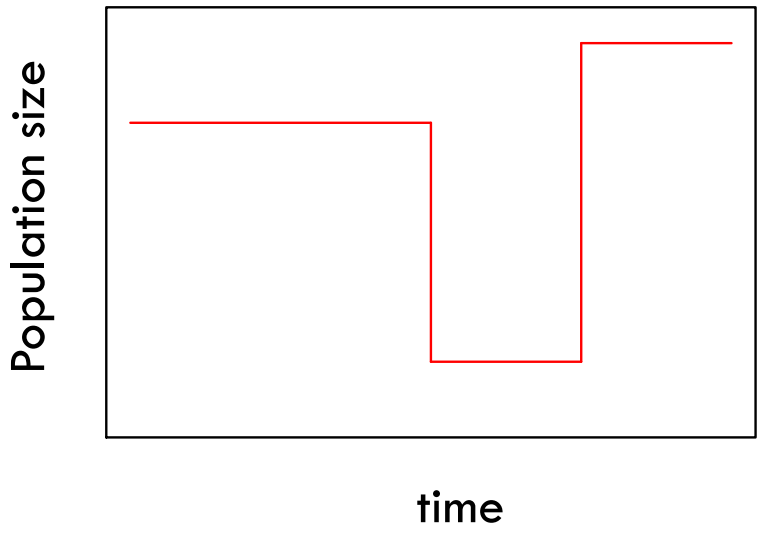
Table 1 | Distribution of Applera SNPs by population and functional class

Category	Shared	Private AA	Private EA	Mean derived frequency	
				AA*	EA†
Synonymous	8,056 (58.3%)	8,958 (53.0%)	3,879 (44.6%)	0.211	0.266
Non-synonymous	5,771 (41.7%)	7,950 (47.0%)	4,826 (55.4%)	0.174	0.202

"Our finding that each person carries several hundred potentially damaging SNPs indicates that large-scale medical resequencing will be useful to find common and rare SNPs of medical consequence"

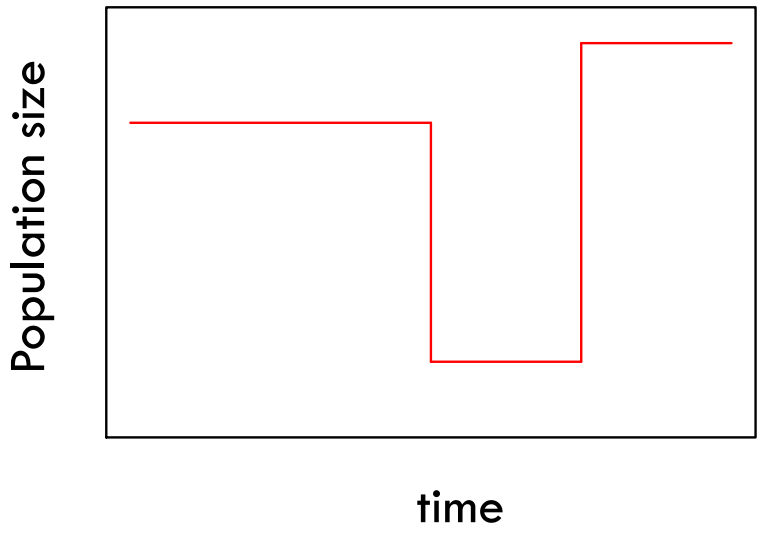
Lohmueller *et al* 2008 (*Nature*)

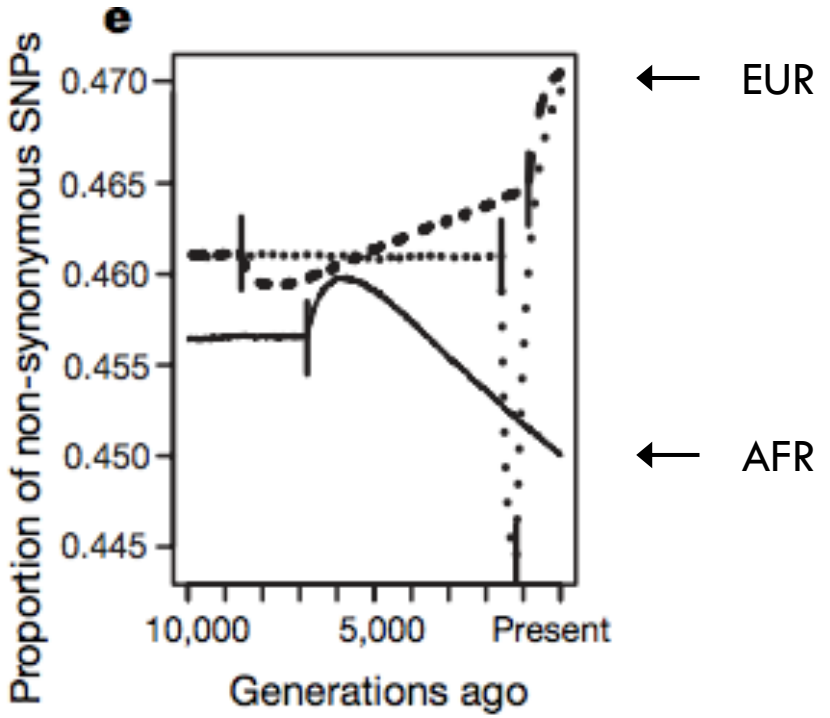
How can we explain this pattern?



Lohmueller *et al* 2008 (*Nature*)

How can we explain this pattern?





Lohmueller *et al* 2008 (*Nature*)

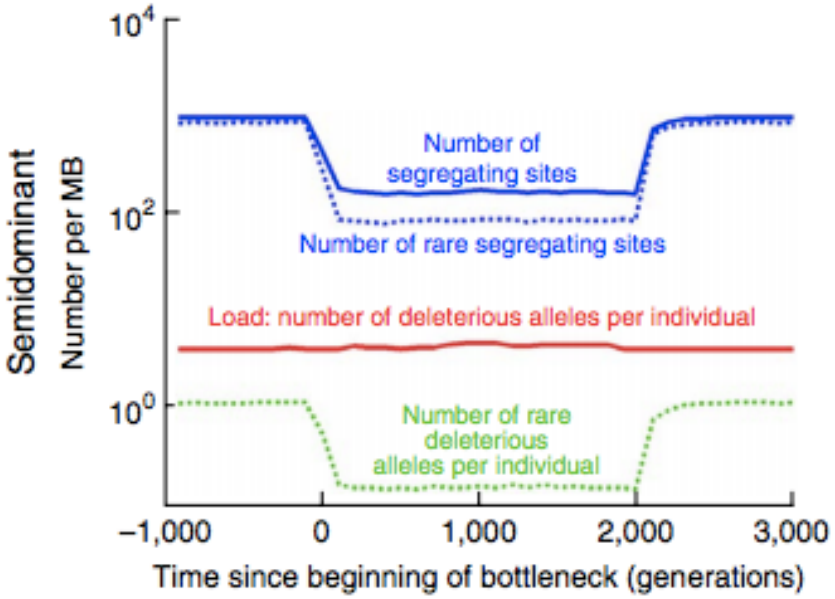
Does deleterious allele "load" differ between populations?

- Load: average number of deleterious alleles per person
- Do Lohmueller *et al's* results imply a difference in load?

Simons *et al* 2014 (*Nature Gen*)

Does deleterious allele "load" differ between populations?

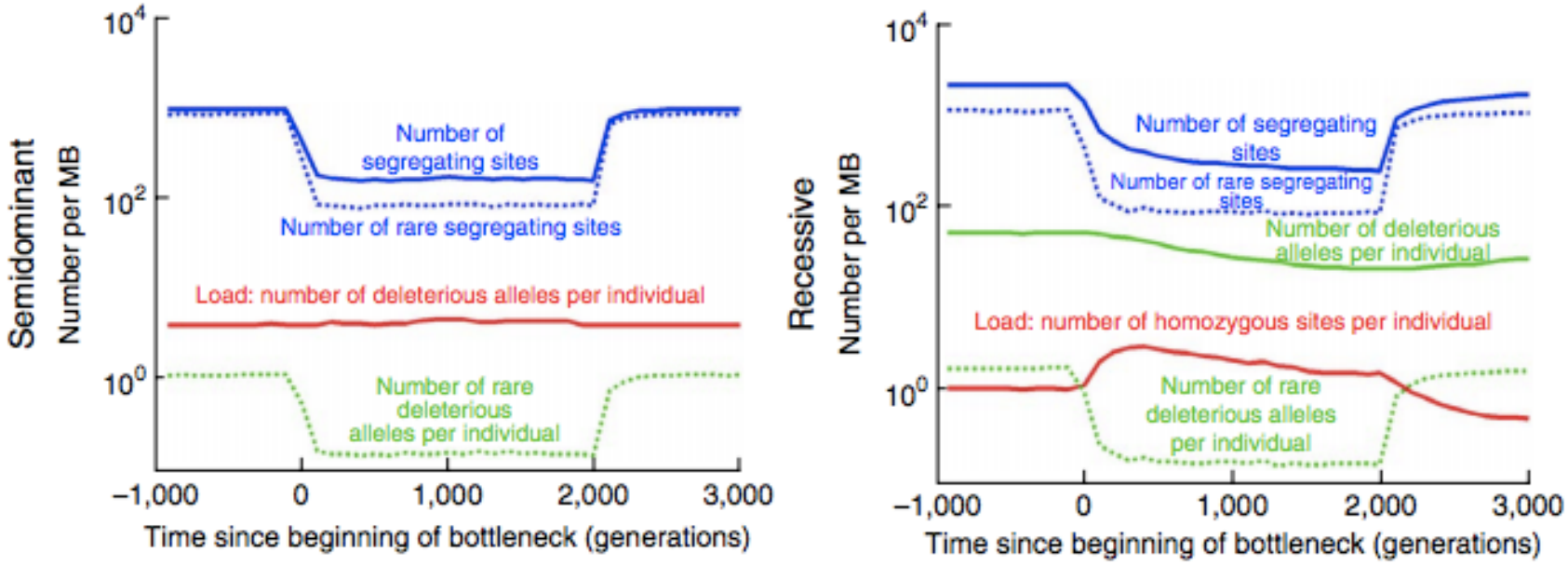
- Load: average number of deleterious alleles per person
- Do Lohmueller *et al's* results imply a difference in load?



Simons *et al* 2014 (*Nature Gen*)

Does deleterious allele "load" differ between populations?

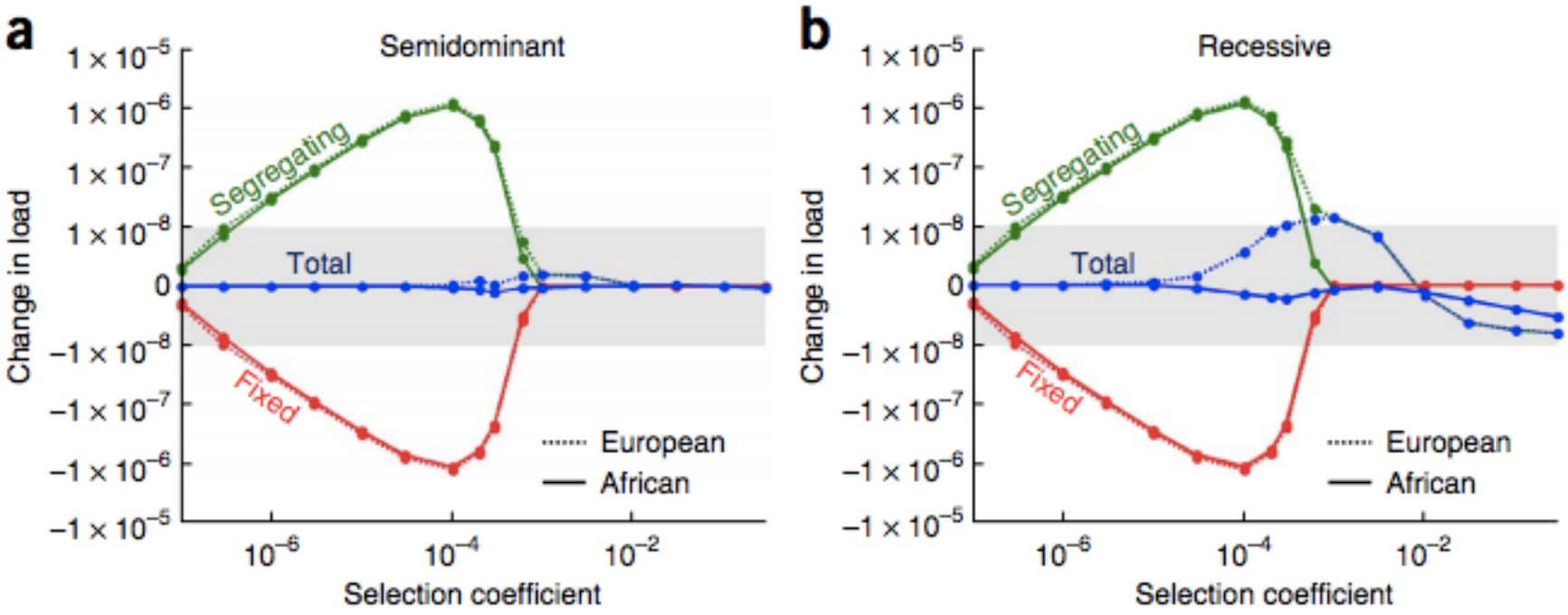
- Load: average number of deleterious alleles per person
- Do Lohmueller *et al's* results imply a difference in load?



Simons *et al* 2014 (*Nature Gen*)

Does deleterious allele "load" differ between populations?

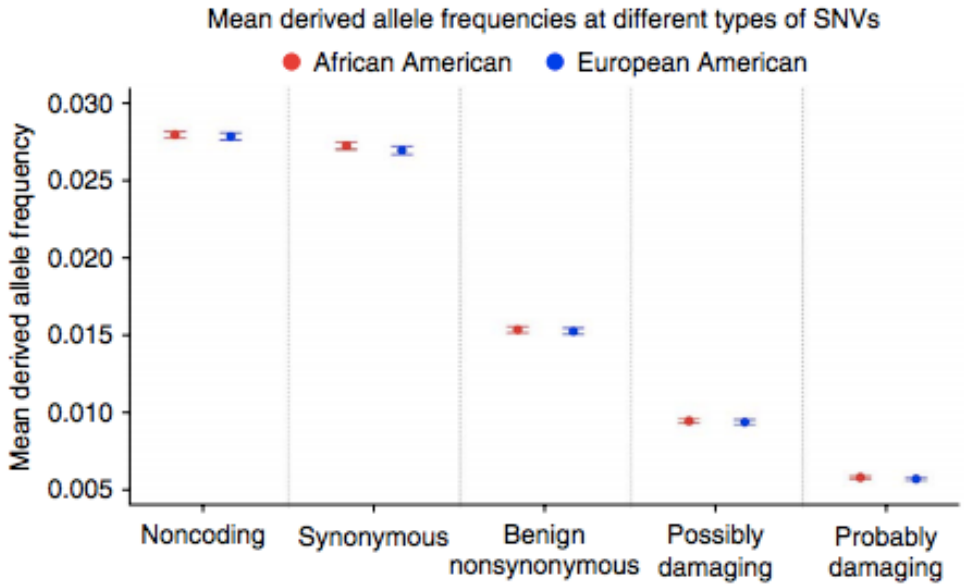
- Load: average number of deleterious alleles per person
- Do Lohmueller *et al's* results imply a difference in load?



Simons *et al* 2014 (*Nature Gen*)

Does deleterious allele "load" differ between populations?

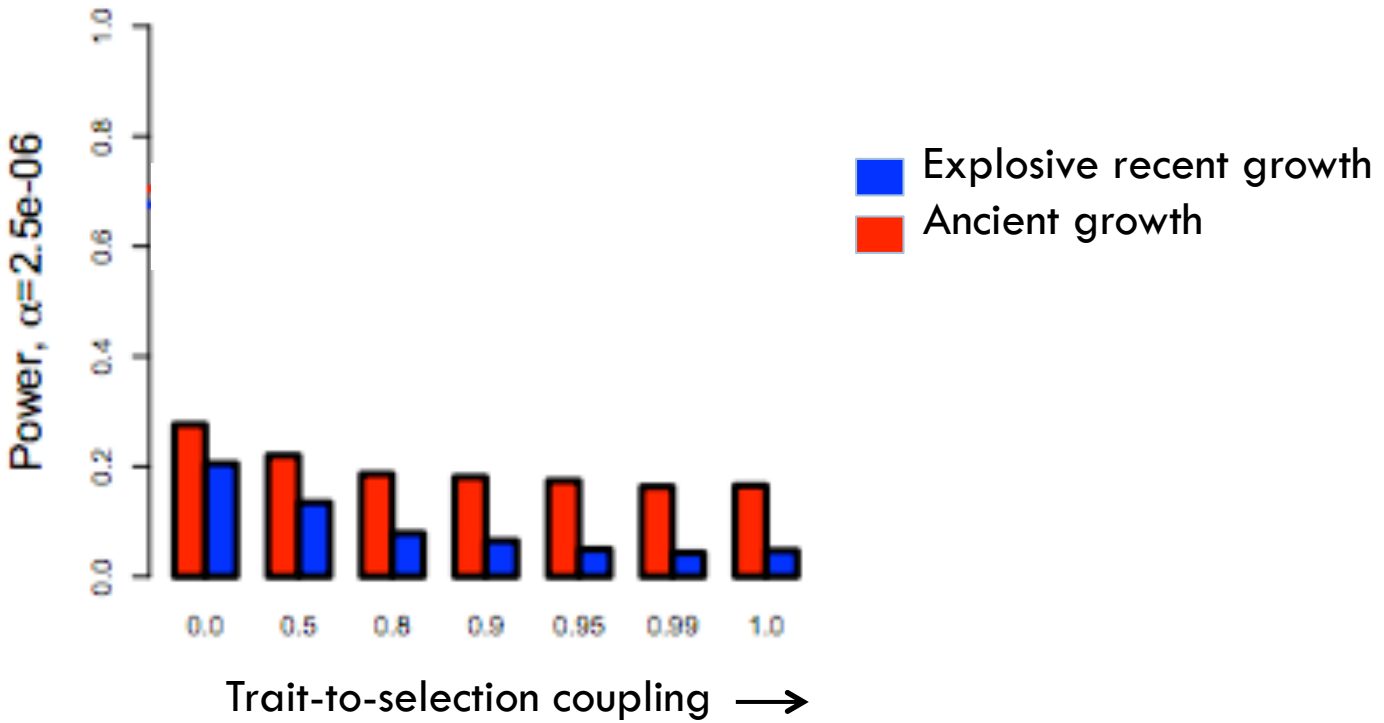
- Load: average number of deleterious alleles per person
- Do Lohmueller *et al's* results imply a difference in load?



	Noncoding	Synonymous	Benign nonsynonymous	Possibly damaging	Probably damaging
Number per individual, AA:	21,421	15,401	5,373	1,695	2,002
Number per individual, EA:	21,345	15,231	5,338	1,682	1,969

Simons *et al* 2014 (*Nature Gen*)

How do these ideas impact association tests?



Uricchio 2014 (UCSF dissertation); Uricchio *et al* 2016 (*Genome Research*)

So, do rare alleles make substantial contributions to complex trait genetic variance?

- Several papers suggest that many human traits driven by common alleles (including risk for diabetes and autism)
- Possible minor role in human height
- Ongoing work at UCSF on RNA expression suggests rare alleles are a major contributor

Gaugler *et al* 2014 (*Nature Genetics*), Lohmueller *et al* 2013 (*AJHG*); Yang *et al* 2015 (*Nature Genetics*); Fuchsberger *et al* 2016 (*Nature*)

My email: uricchio@stanford.edu

Bibliography

- Henn *et al* 2015 (*Nature Reviews Genetics)* -> a review on estimating deleterious allele load in human populations
- Pritchard 2001 (*AJHG)* -> shows that rare alleles can matter for complex traits in some parametric models
- Boyko *et al* 2008 (*Plos Genetics)* -> a paper that estimates the strength of selection on human nonsynonymous sites
- Lohmueller *et al* 2008 (*Nature*) -> argues that Europeans carry proportionally more deleterious variation than Africans
- Simons *et al* 2014 (*Nature Genetics*) -> a paper that argues that deleterious allele load is not sensitive to demographic effects
- Szpiech *et al* 2014 (AJHG) -> shows an enrichment in deleterious alleles in ROH in humans
- Henn *et al* 2016 (*PNAS*) -> a paper that argues that deleterious allele load increases with distance from Africa
- Gao *et al* 2015 *(Genetics)* -> argues that recessive lethal mutations in humans are rarer than previously appreciated
- Lohmueller 2014 (*Plos Genetics*) -> shows how evolutionary events impact diseases association studies
- Uricchio *et al* 2016 (*Genome Research)* -> Uses evolutionary models to argue that association tests may be severely underpowered to discover rare causal variation