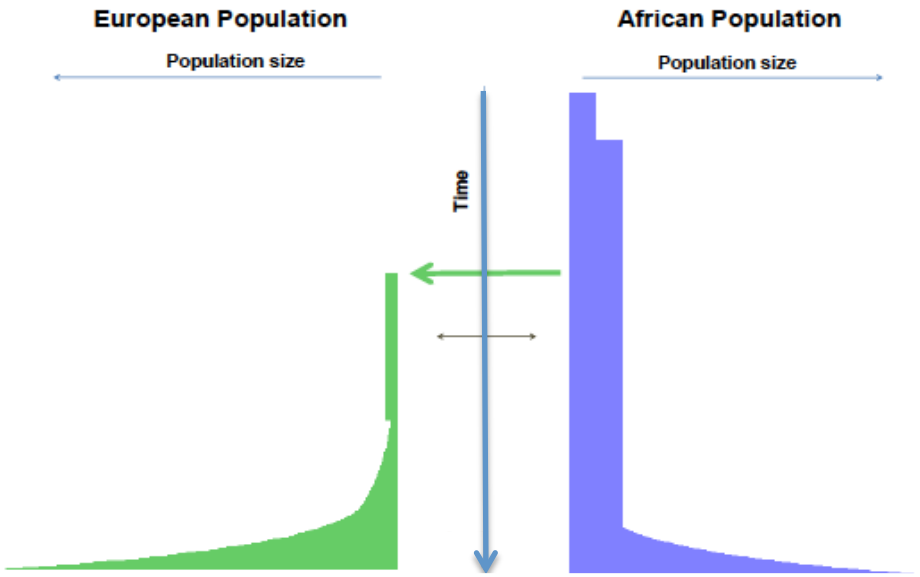
The impact of recent human demography on the deleterious mutation load and the genetic architecture of disease susceptibility

Guy Sella

Columbia University

Recent human demographic history had a substantial impact on genetic variation

- An abundance of very rare variants due to explosive growth over the past ~5Kya.
- A greater fraction of high frequency alleles in non-African population due to an Out-of-Africa bottleneck ~50Kya.



Following Tennessen et al (2012)

Wall & Przeworski (2000); Voight et al. (2005); Keinan et al. (2007); Gutenkunst et al. (2009); Coventry et al. (2010); Marth et al. (2011); Keinan & Clark (2012); Nelson et al. (2012); Tennessen et al. (2012);

How did demographic history affect genetic load and architecture?

"... negative selection is less effective at removing slightly deleterious alleles from European populations." Lohmueller et al. Nature 2008

"However, when the more conservative definition was used, this pattern was reversed, and AA individuals have a significantly higher proportion of predicted functional CNVs compared with EA individuals." Tennessen et al. Science 2012

"Some degree of genetic risk may be due to this recent rapid expansion of rare and common variants in the human population" Keinan & Clark, Science 2012

"…the increased mutational capacity of recent human populations has led to a larger burden of Mendelian disorders, increased the allelic and genetic heterogeneity of traits…" Fu et al. Nature 2012

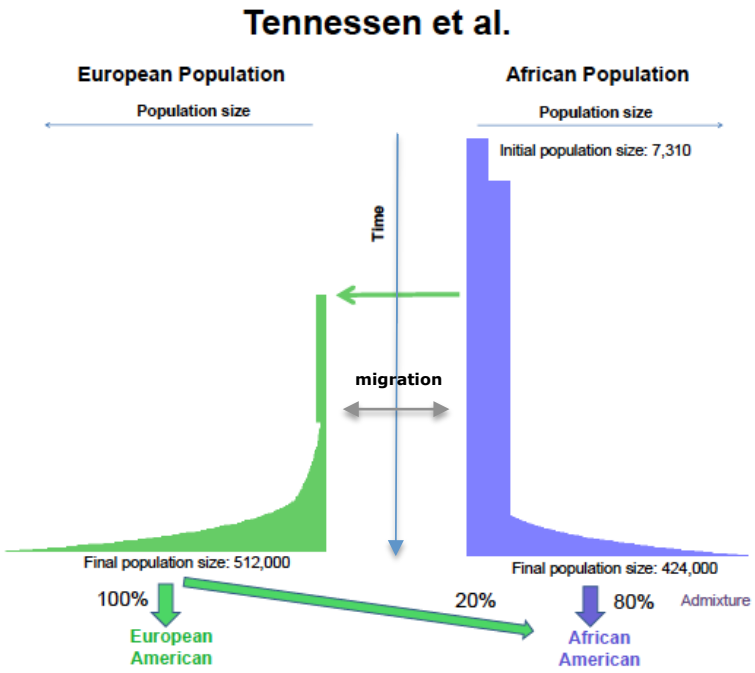
How did demographic history affect genetic load and architecture?

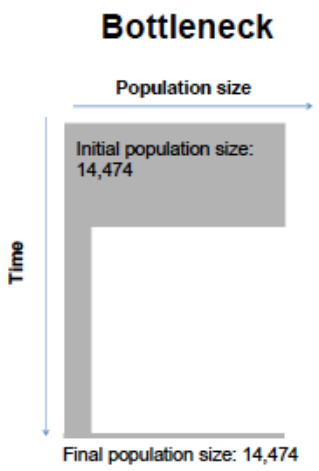
Little if at all

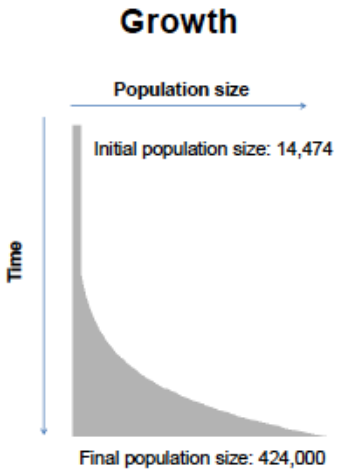
It depends how strongly the trait is coupled with fitness

The model

- Standard bi-allelic viability selection model (1, 1-hs, 1-s) with a finite diploid population and two-way mutation.
- We focus on the semi-dominant and recessive cases.
- LE and multiplicative fitness.
- Three demographic scenarios:

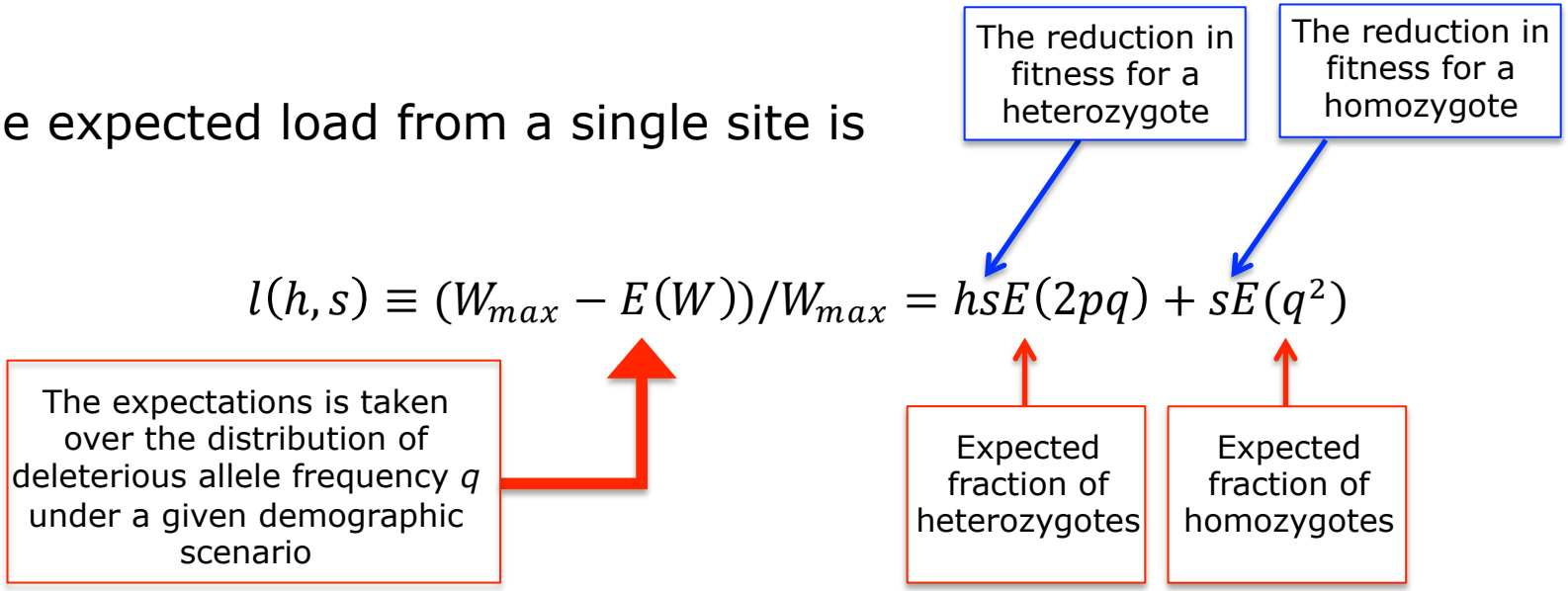






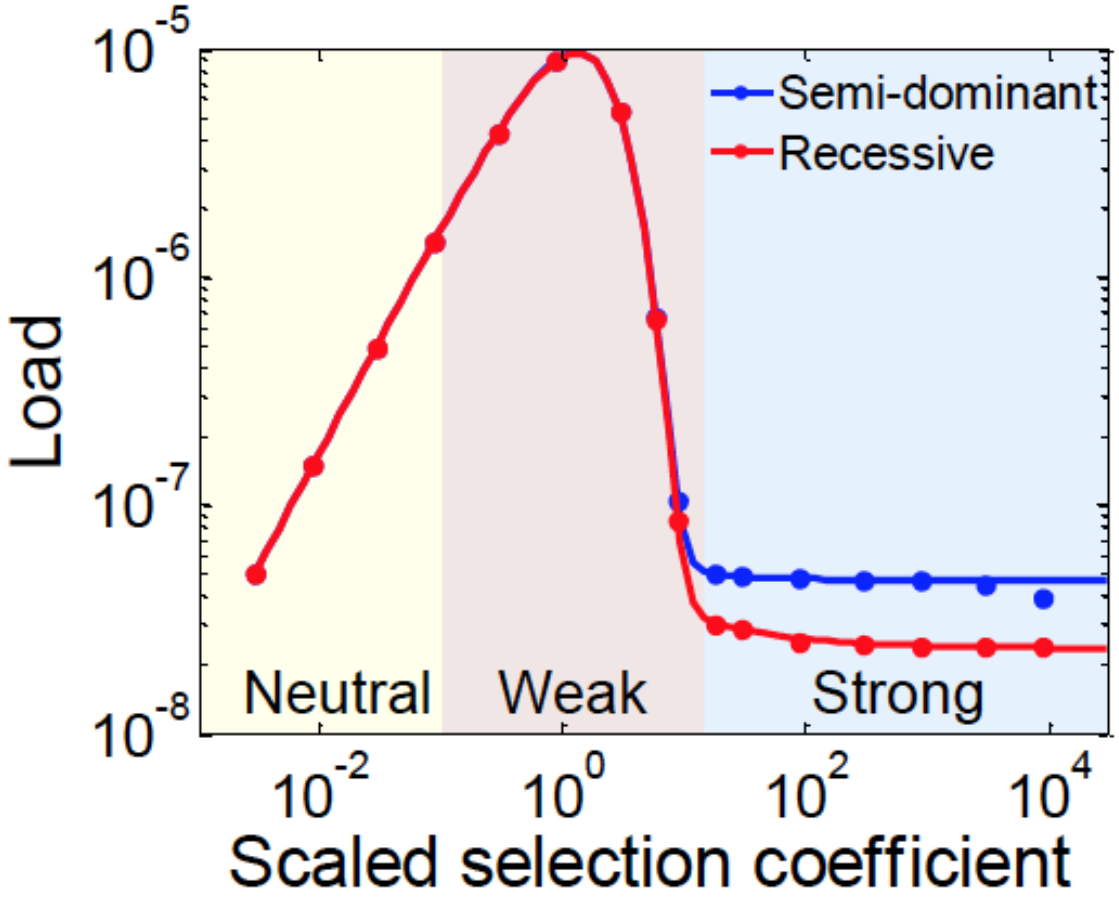
Genetic load is defined as the relative reduction in fitness caused by deleterious alleles

The expected load from a single site is

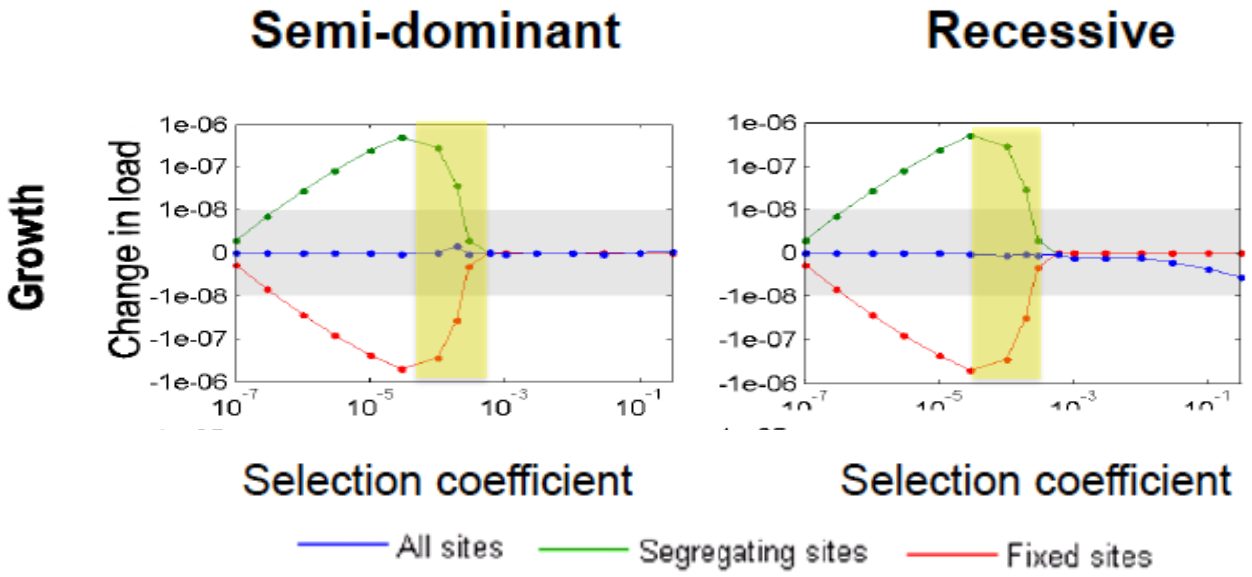


Under our model the load from multiple sites is a simple function of the sum of load over sites.

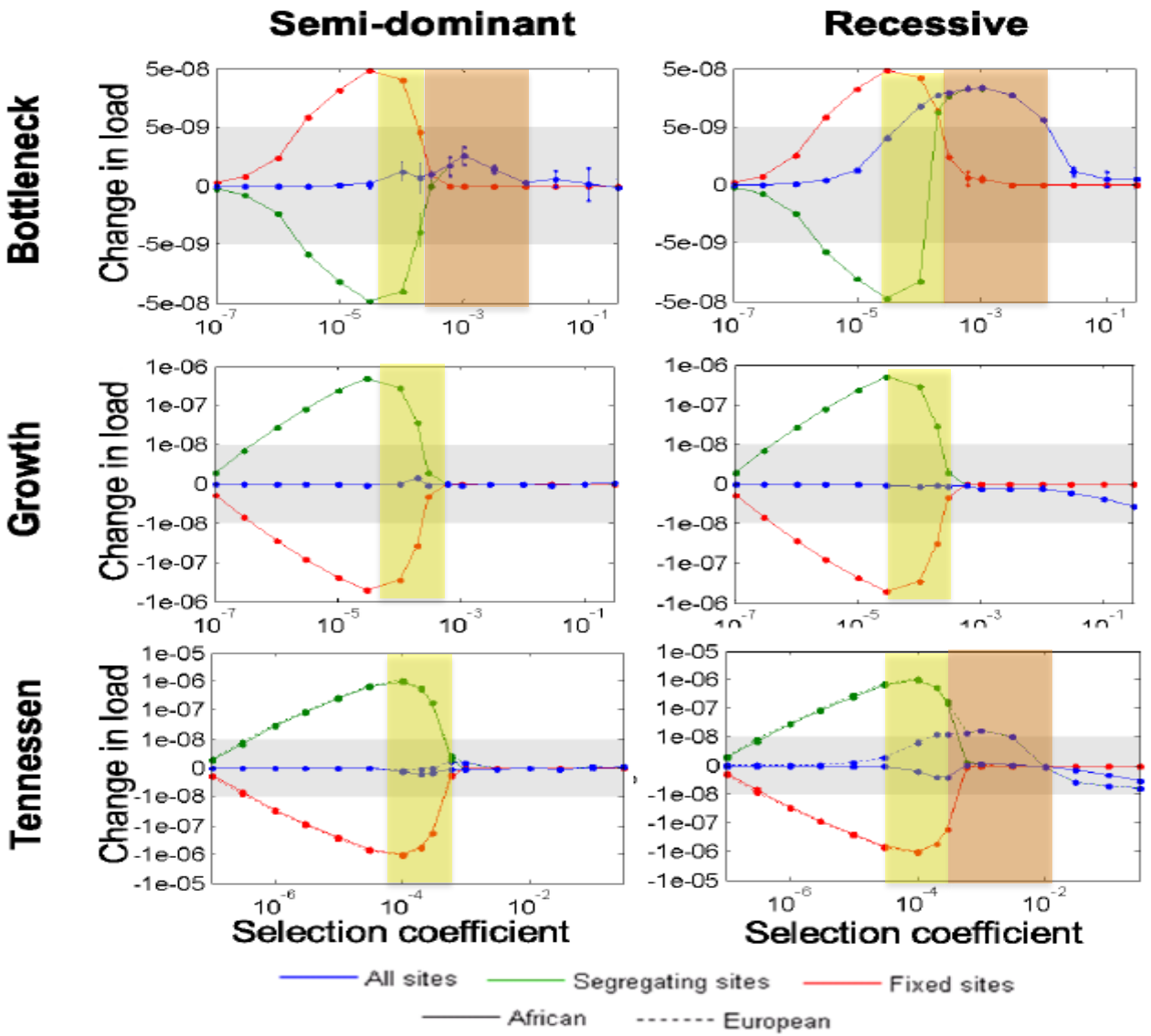
With constant population size genetic load exhibits three dynamic regimes



Recent demography has had a negligible effect on load



Recent demography has had a negligible effect on load



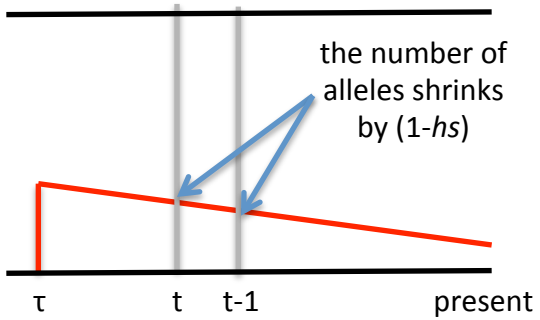
The detailed reasons depend on dominance and strength of selection

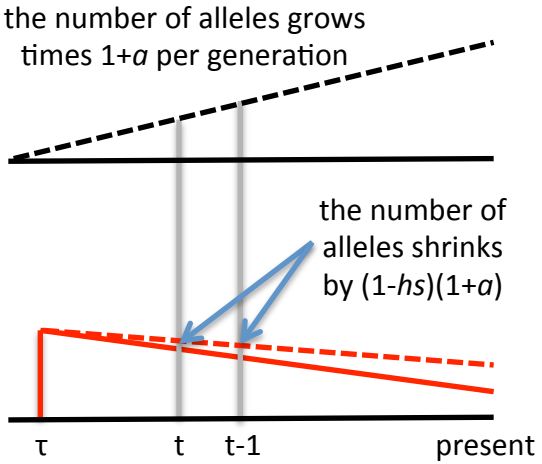
			Effectively neutral	Weak		Strong
				closer to neutral	closer to strong	
Bottleneck	Semi-dominant	fixed	increase	increase	increase	—
		segregating	decrease	decrease	increase	unchanged
		total	unchanged	increase	increase	unchanged
	Recessive	fixed	increase	increase	increase	—
		segregating	decrease	decrease	increase	transient increase
		total	unchanged	increase	increase	transient increase
Growth	Semi-dominant	fixed	decrease	decrease		—
		segregating	increase	increase		unchanged
		total	unchanged	unchanged		unchanged
	Recessive	fixed	decrease	decrease		—
		segregating	increase	increase		transient decrease
		total	unchanged	unchanged		transient decrease

The strong selection dominant case: load is not affected by changes in population size

Consider the case in which selection against heterozygotes is strong.

In this case, the change in the number of deleterious alleles is precisely offset by the change in population size.





The strong selection dominant case: load is not affected by changes in population size

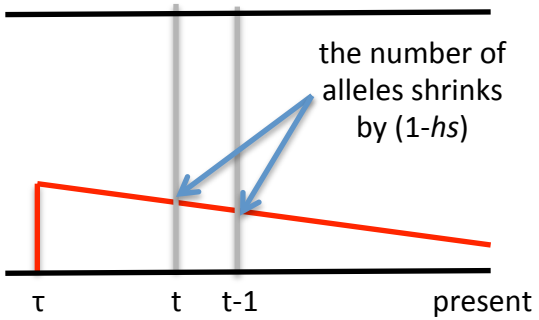
Consider the case in which selection against heterozygotes is strong.

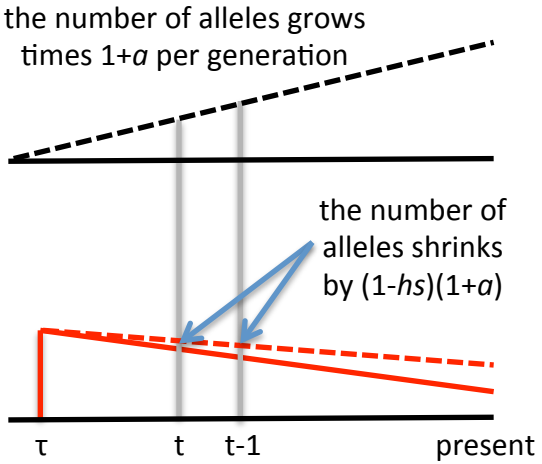
In this case, the change in the number of deleterious alleles is precisely offset by the change in population size.

It follows that the classic mutation-selection balance results

$$E(q) = \frac{u}{hs} \quad \text{and} \quad E(l) = 2u$$

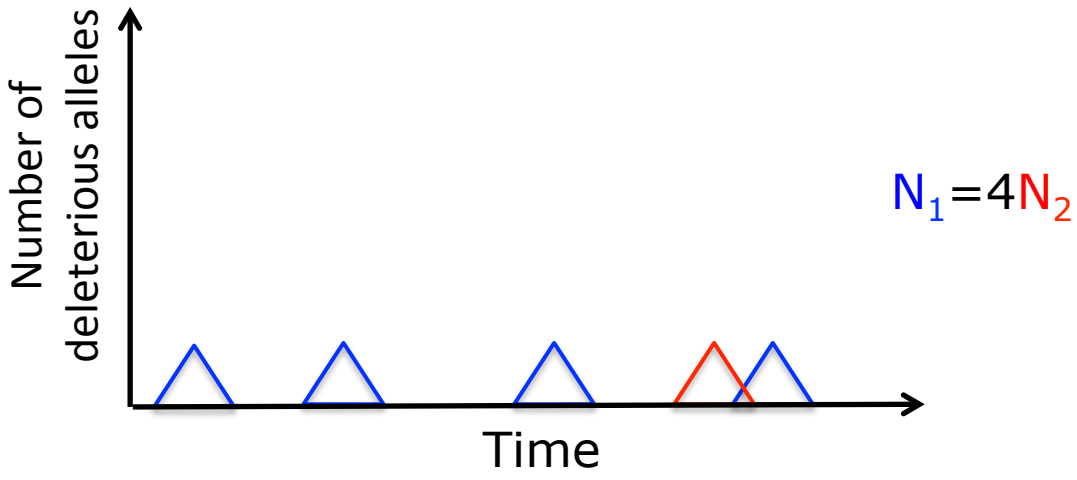
hold, regardless of changes in population size.





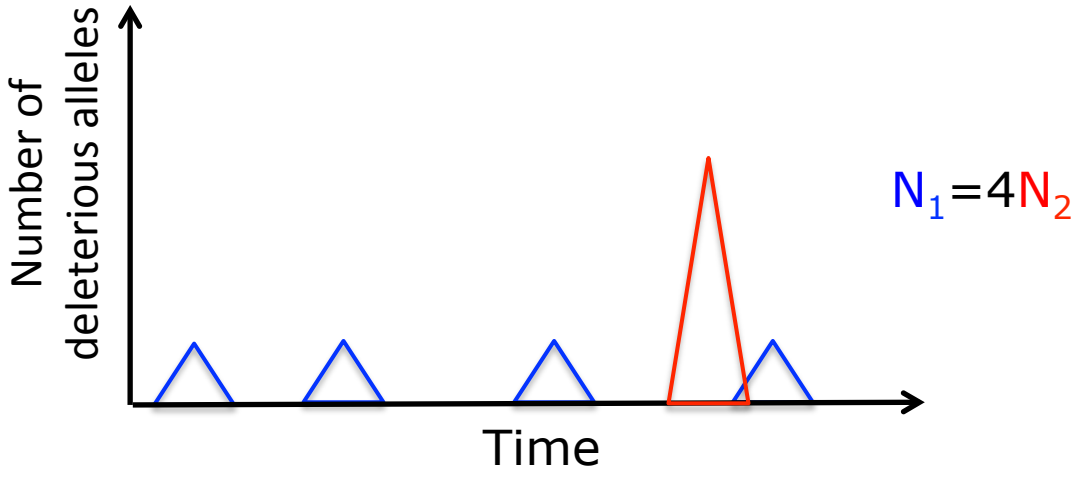
Intuition based on the trajectory of deleterious alleles

As long as the selection on heterozygotes is sufficiently strong, we can describe the trajectories using a branching process.



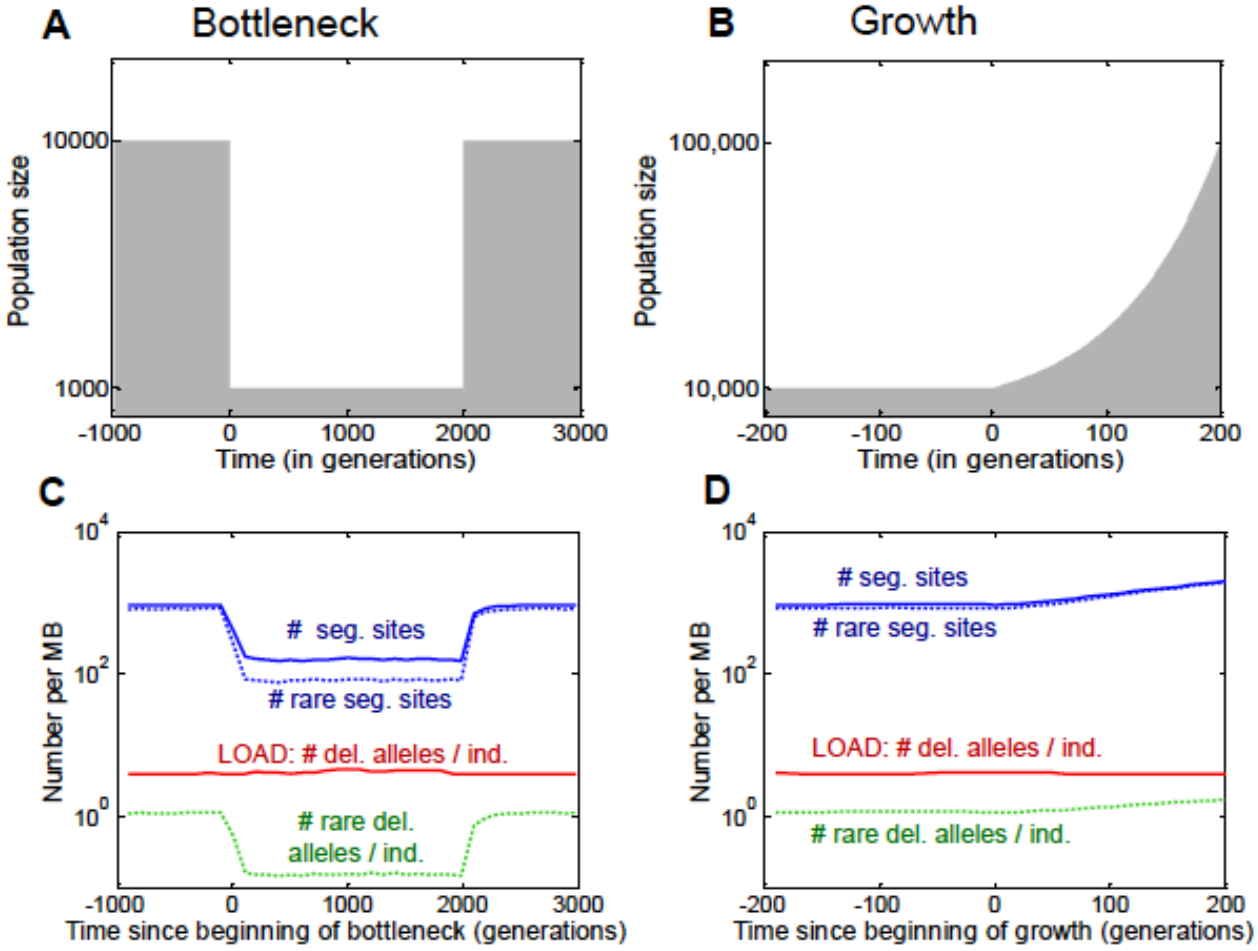
Intuition based on the trajectory of deleterious alleles

As long as the selection on heterozygotes is sufficiently strong, we can describe the trajectories using a branching process.



In larger populations or under growth an increase in the number of segregating deleterious alleles is offset by their decreased frequency, with a bottleneck it's the opposite.

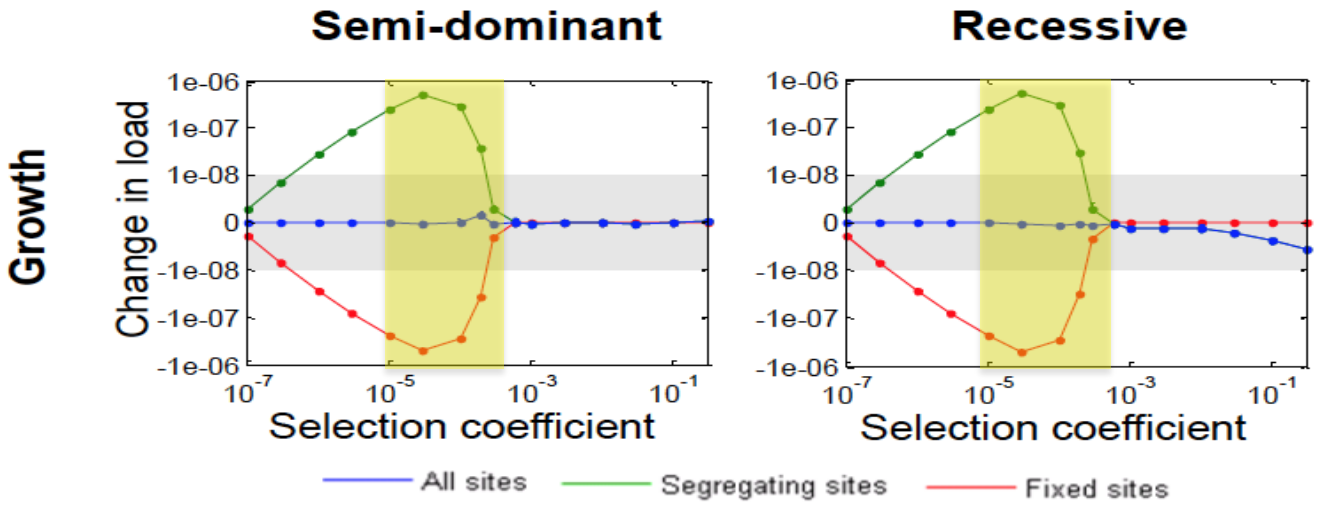
For strong selection the changes in genetic variation are quick



The detailed reasons depend on dominance and strength of selection

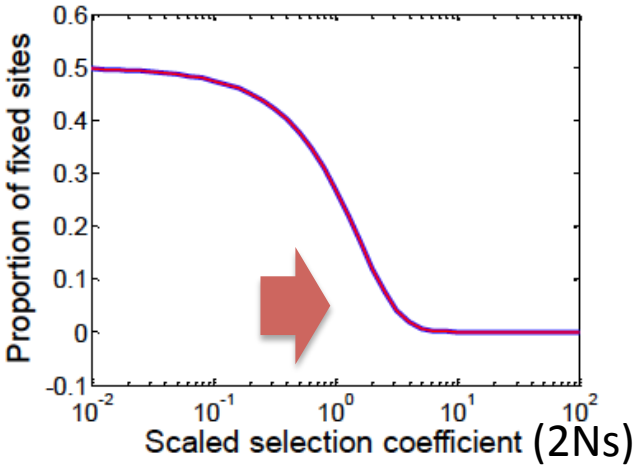
			Effectively neutral	Weak		Strong
				closer to neutral	closer to strong	
Bottleneck	Semi-dominant	fixed	increase	increase	increase	—
		segregating	decrease	decrease	increase	unchanged
		total	unchanged	increase	increase	unchanged
	Recessive	fixed	increase	increase	increase	—
		segregating	decrease	decrease	increase	transient increase
		total	unchanged	increase	increase	transient increase
Growth	Semi-dominant	fixed	decrease	decrease		—
		segregating	increase	increase		unchanged
		total	unchanged	unchanged		unchanged
	Recessive	fixed	decrease	decrease		—
		segregating	increase	increase		transient decrease
		total	unchanged	unchanged		transient decrease

Why does recent growth have a negligible effect on load from weakly selected sites?



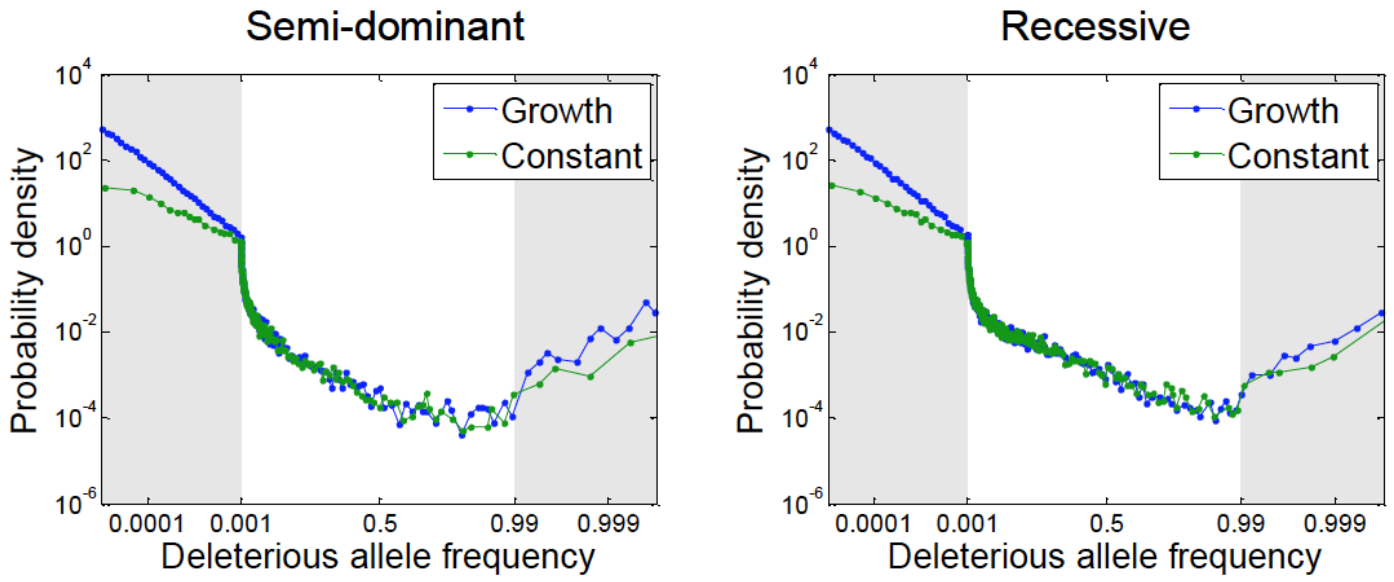
- Practically no effect on total load
- Load moves from fixed to segregating

Change in load due to fixations take too long

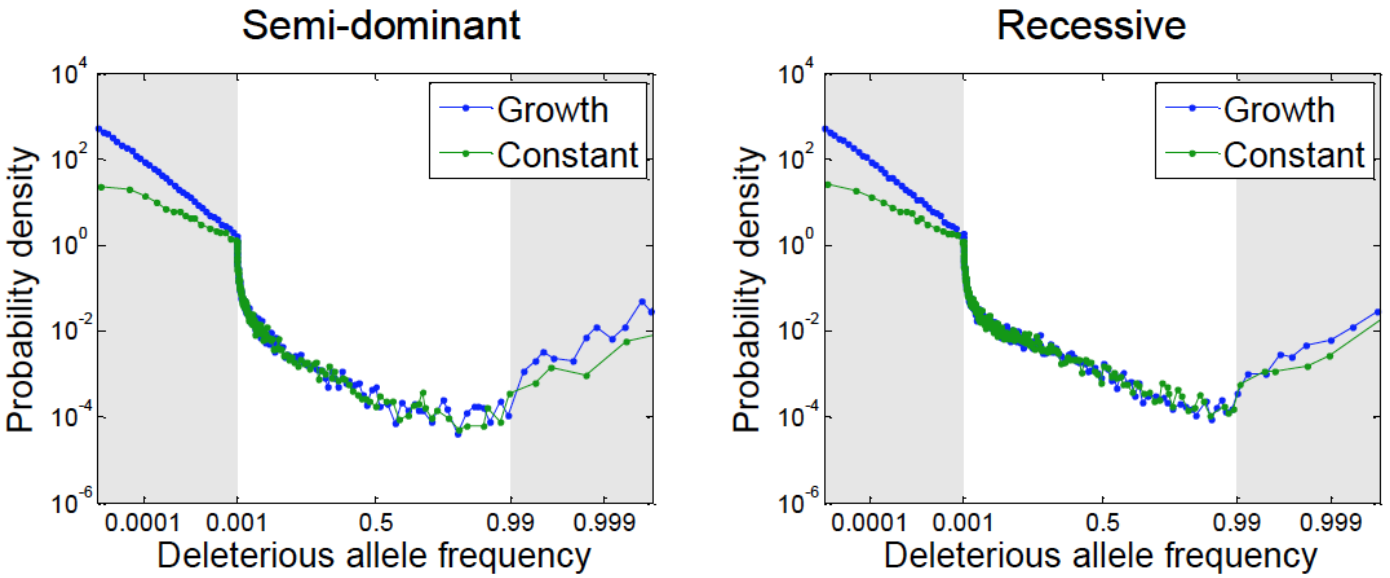


Change due to drift at segregating sites also take too long

Weakly selected sites respond to demographic changes much slower than strongly selected ones



Change due to drift at segregating sites also take too long

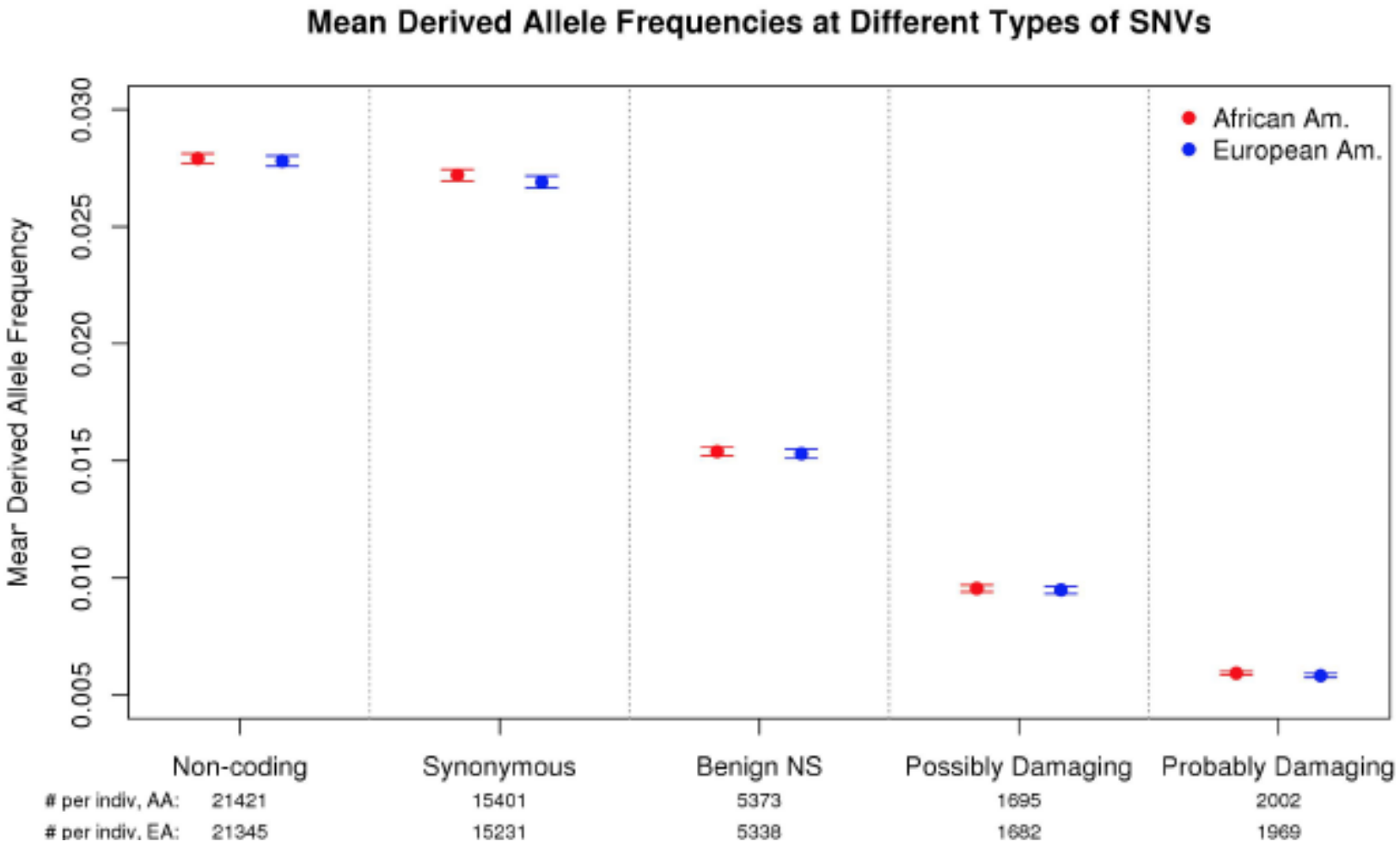


A testable prediction: recent demography had little effect on load

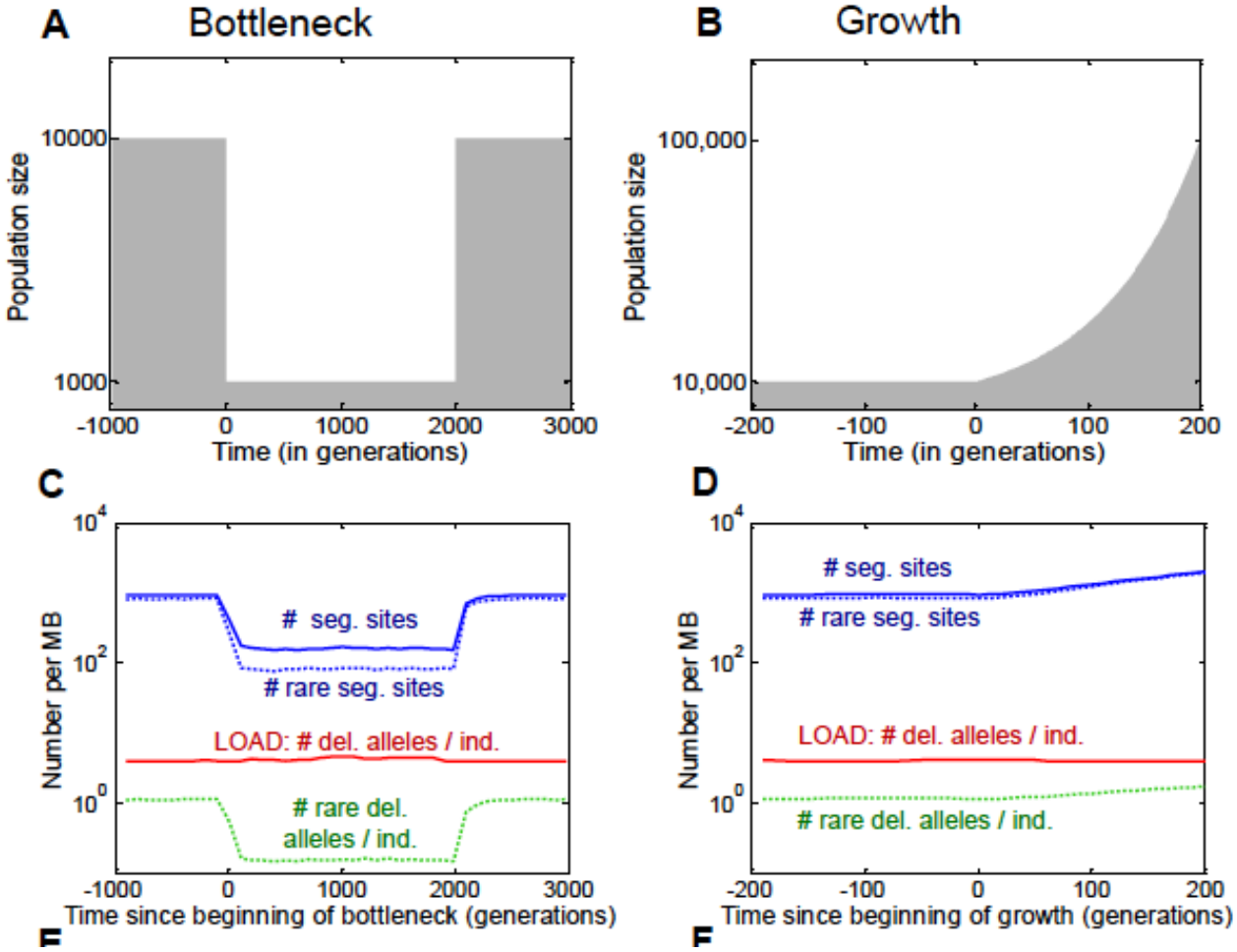
African and European populations, which differ in their recent histories, should carry similar numbers of deleterious alleles. Should be true across classes of variants with different fitness effects.

- Data: Deep sequencing of 15,336 exons from 2,217 AA and 4,298 EA (Fu et al. 2012 Nature). Also, 1000 Genomes Phase I CEU and YRI.
- PolyPhen 2 (and other) classifications of NS changes.
- Statistic: Mean derived allele frequency of variants segregating in total sample.

Derived allele frequencies in AA and EA are similar across functional categories



While the load is similar the architecture of complex traits doesn't have to be…



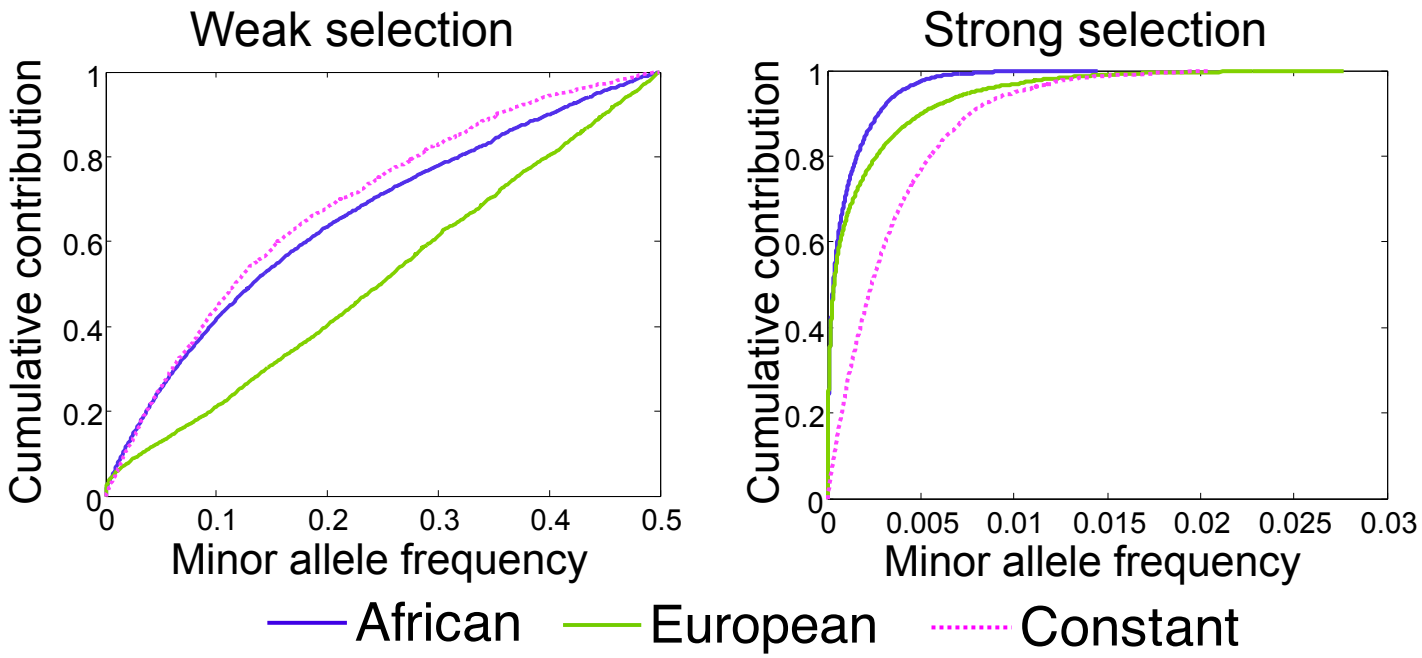
Under growth an increase in the number of segregating deleterious alleles is offset by their decreased frequency, with a bottleneck it's the opposite.

How did demographic history affect the contribution of rare and common alleles to disease risk?

Did recent population growth lead to a greater role of rare alleles in disease risk?

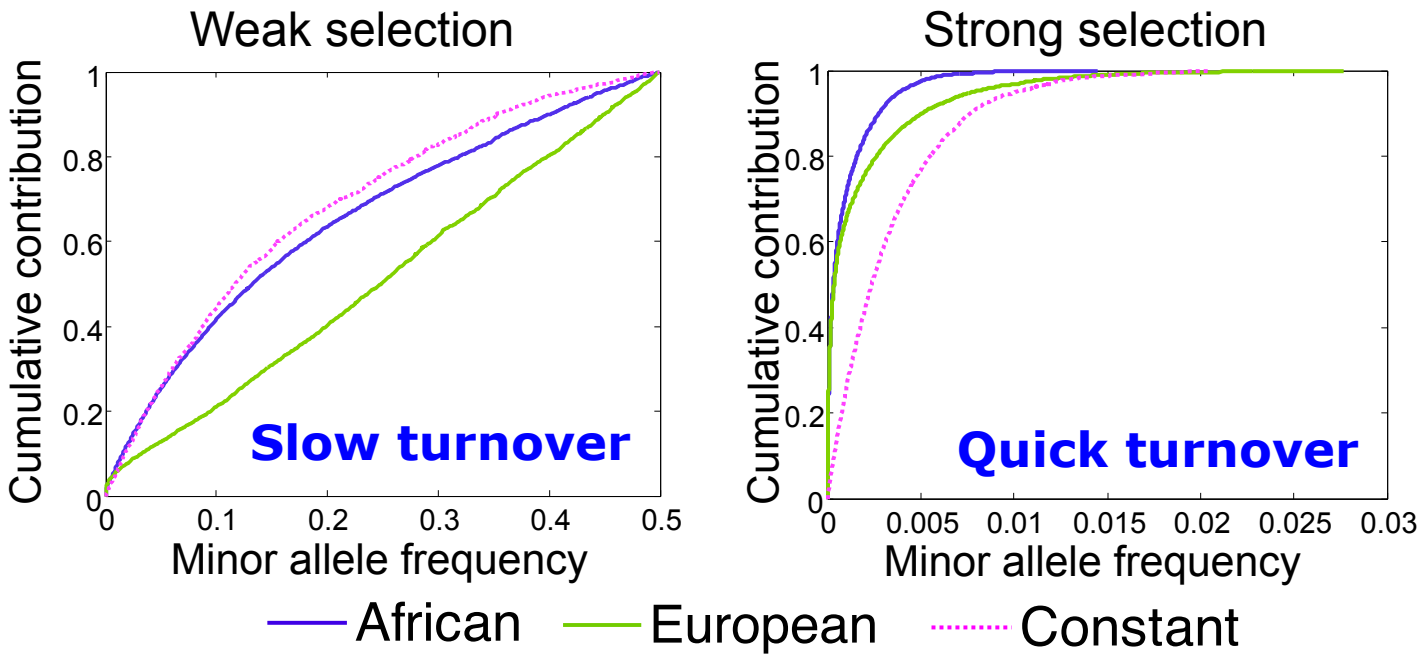
Did the Out-of-Africa bottleneck lead common alleles to play a greater role in non-Africans compared to Africans?

The contribution of rare/common alleles strongly depends on their fitness effects



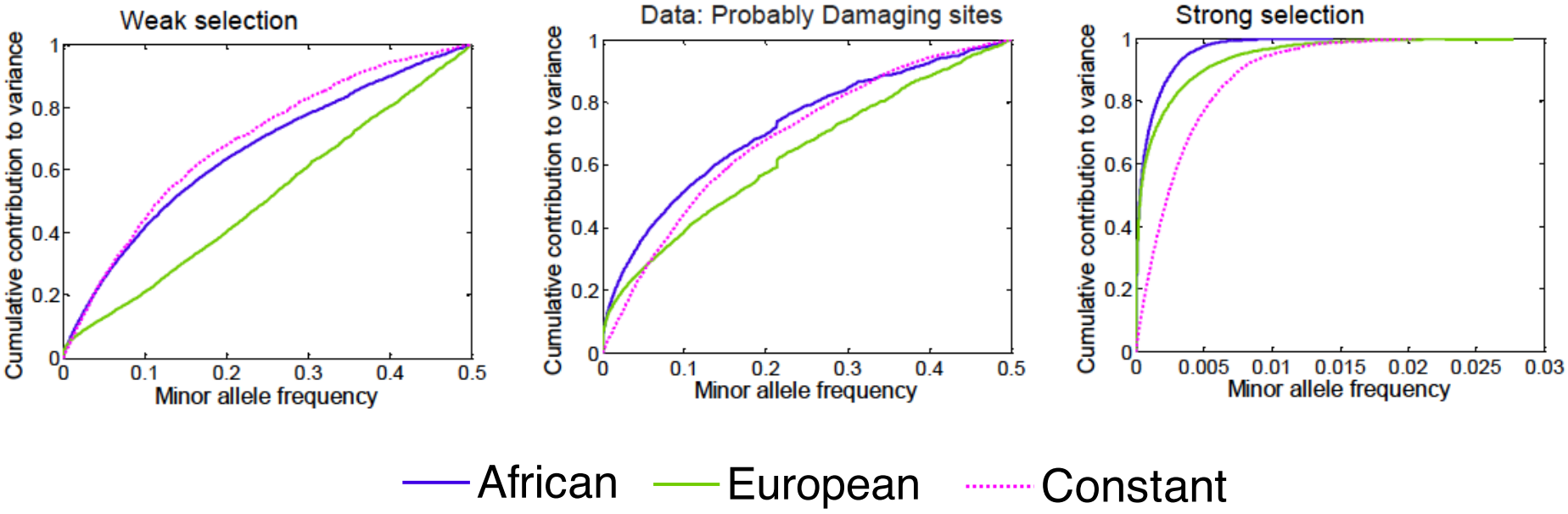
- Assume that selection and effect size are semi-dominant.
- The contribution to overall variance from variants with frequency x is then proportional to $x(1-x)f(x)$, where $f(x)$ is the density of alleles with frequency x.

The contribution of rare/common alleles strongly depends on their fitness effects



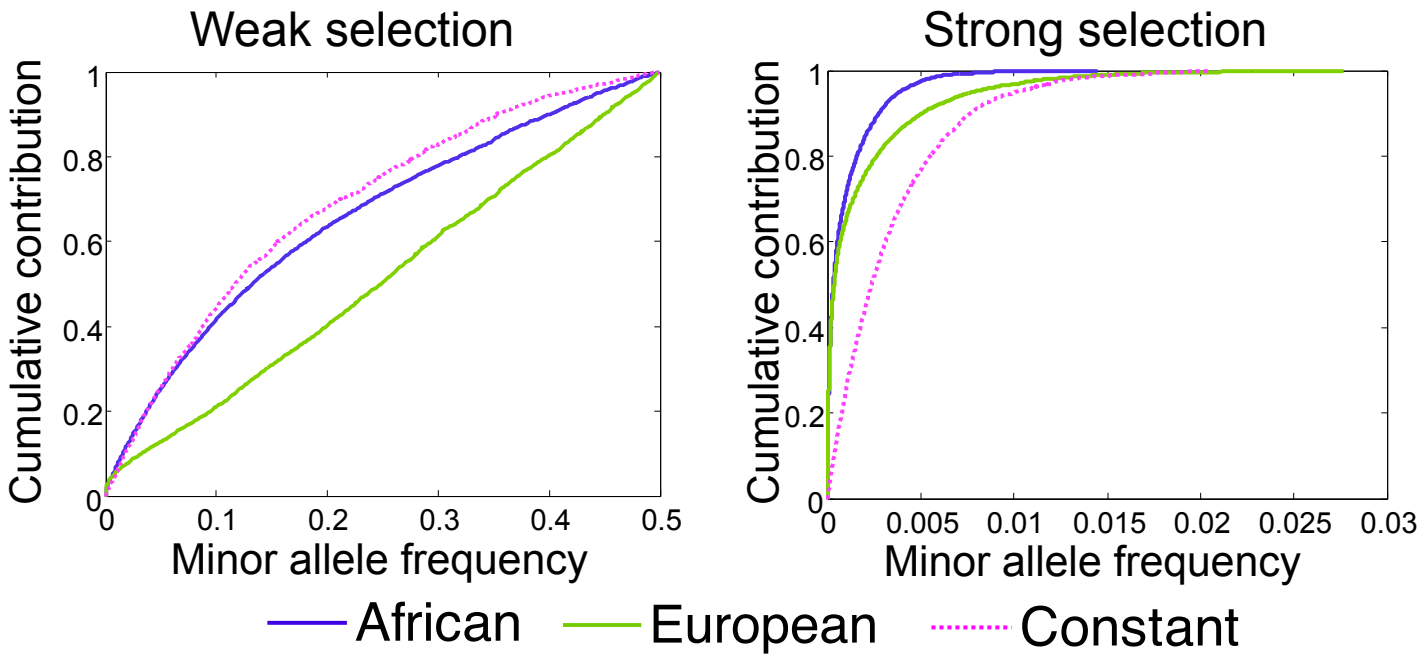
- For strongly selected variants, growth greatly increases the contribution of rares.
- For neutral, weakly selected and all recessives, growth has little effect.

The contribution of rare/common alleles strongly depends on their fitness effects



- For strongly selected variants, growth greatly increases the contribution of rares.
- For neutral, weakly selected and all recessives, growth has little effect.

The contribution of rare/common alleles strongly depends on their fitness effects



So, demography can substantially affects the contribution of rare/common variants but the effect depend on the selection acting on variants.

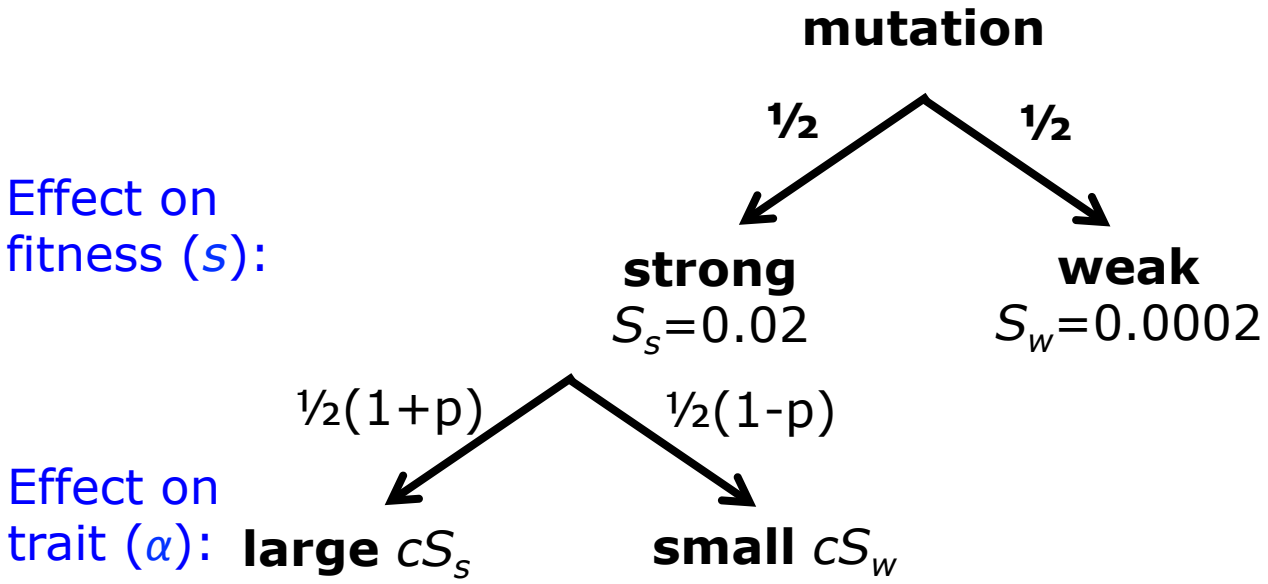
In reality we would expect the variants contributing to phenotypic variation to have a mixture of selection coefficients and effect sizes.

The relationship between a variant's effect on a trait and on fitness: two extremes

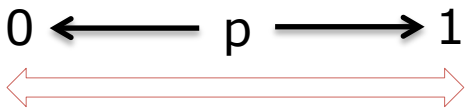
Effect on fitness and trait are independent (e.g., anthropomorphic traits and late onset diseases)

Effect on fitness and trait are directly related (e.g., diseases that are early onset and affect fertility)

The relationship between a variant's effect on a trait and on fitness: a toy model

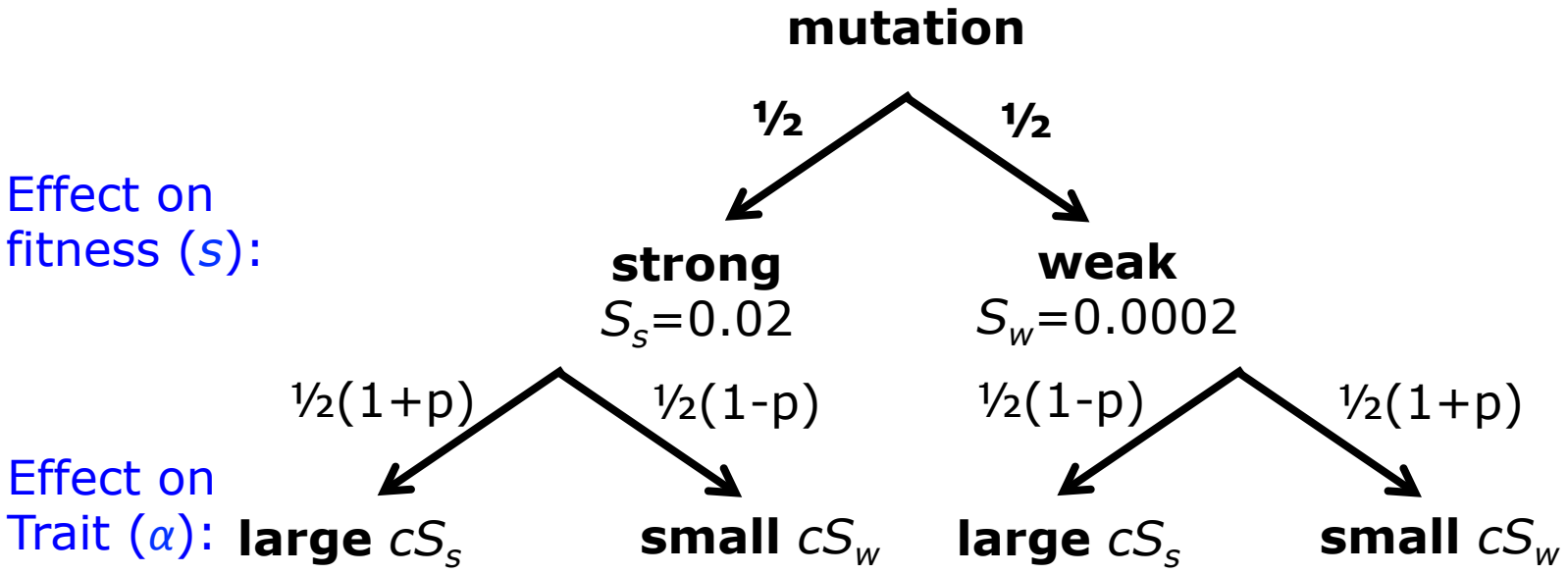


Effect on fitness and trait are independent (e.g., anthropomorphic traits and late onset diseases)



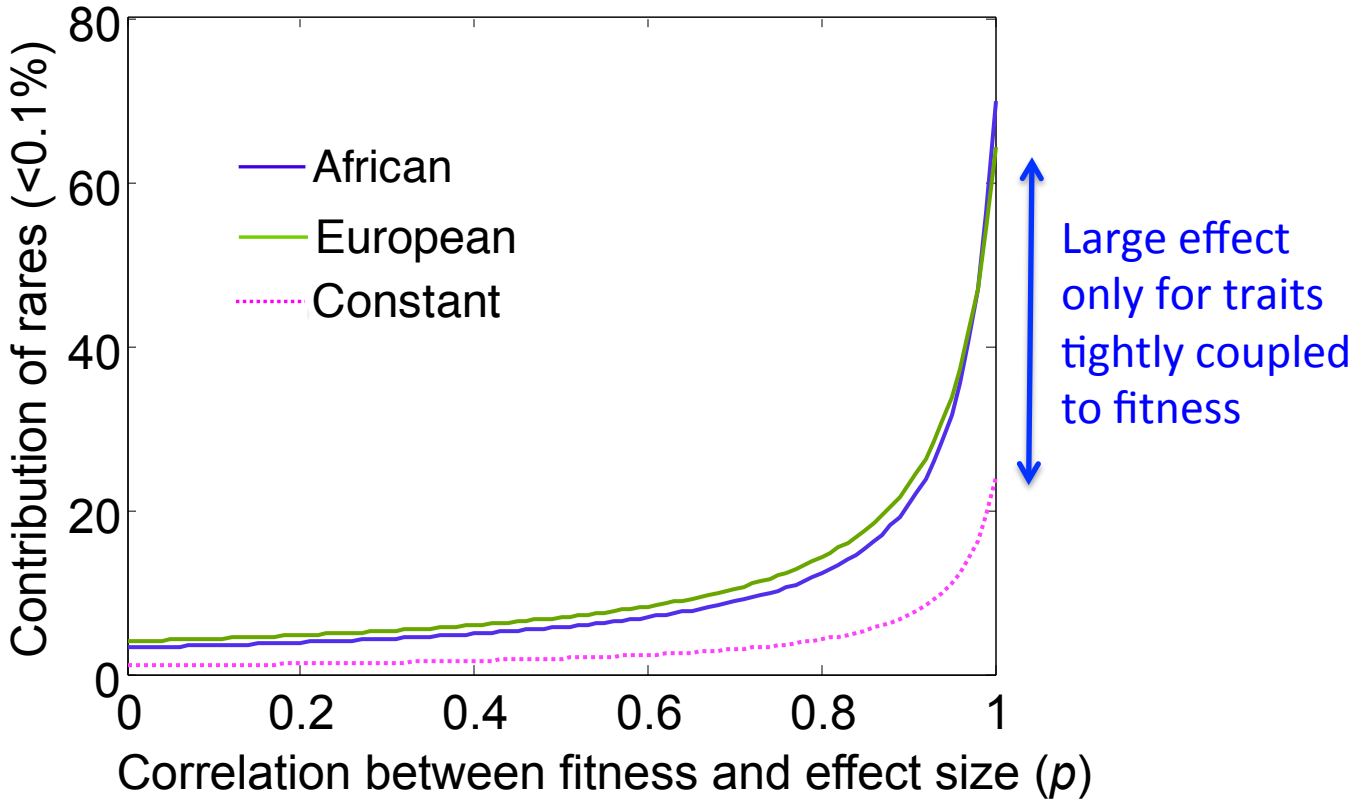
Effect on fitness and trait are directly related (e.g., diseases that are early onset and affect fertility)

The relationship between a variant's effect on a trait and on fitness: a toy model



- The marginal distributions of s and α remain constant for any *p.*
- The correlation between s and α is *p*.

Growth should only affect traits tightly coupled with fitness



Effect on fitness and trait are independent (e.g., anthropomorphic traits and late onset diseases)

Effect on fitness and trait are directly related (e.g., diseases that are early onset and affect fertility)

Summary

- The Out-of-Africa bottleneck or recent population growth should have had negligible effect on genetic load.
- Data analysis confirms this prediction.
- For most traits, growth should have had little effect on the genetic architecture.
- A possible exception are traits that are strongly coupled with fitness, for which the large contribution of rare alleles would be amplified by recent growth.
- For most traits, the Out-of-Africa bottleneck should have increased the contribution of common alleles.

more at http://arxiv.org/abs/1305.2061 or *Nature Genetics* (2014) doi:10.1038/ng.2896

Acknowledgements

Yuval Simons
(Columbia U.)

Michael Turchin
(U. Chicago)

Jonathan Pritchard
(Stanford U.)

Also thanks to David Reich (Harvard U.),
Shamil Sunyaev (Harvard U.),
Josh Akey (Wash U.)
and Molly Przeworski (Columbia U.)

Interested in joining the lab?

Biological Sciences, Columbia

more at http://arxiv.org/abs/1305.2061
or *Nature Genetics* (2014) doi:10.1038/ng.2896