

The genomic footprint of polygenic adaptation: from sweeps to small frequency shifts

Joachim Hermisson

Mathematics & Biology, University of Vienna

Three adaptive scenarios

Current Biology 20, R208–R215, February 23, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.11.055

The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation Review

Jonathan K. Pritchard^{1,2,*}, Joseph K. Pickrell¹,
and Graham Coop³

There has long been interest in understanding the genetic basis of human adaptation. To what extent are phenotypic differences among human populations driven by natural selection? With the recent arrival of large genome-wide data sets on human variation, there is now unprecedented opportunity for progress on this type of question. Several lines of evidence argue for an important role of positive

the key findings thus far, and then focus on what we see as some of the major open questions. A number of other recent reviews discuss either the general principles for detecting selection or summarize the overall results in more detail than we attempt here [1–6].

Recent Human Adaptations

While human populations differ in various phenotypes, there is a considerable burden of proof to show that phenotypic differences have a genetic basis and are adaptive. However, we do now have reasonable evidence of differential adapta-

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Three adaptive scenarios

*molecular
popgen*



*quantitative
genetics*

Hard sweeps

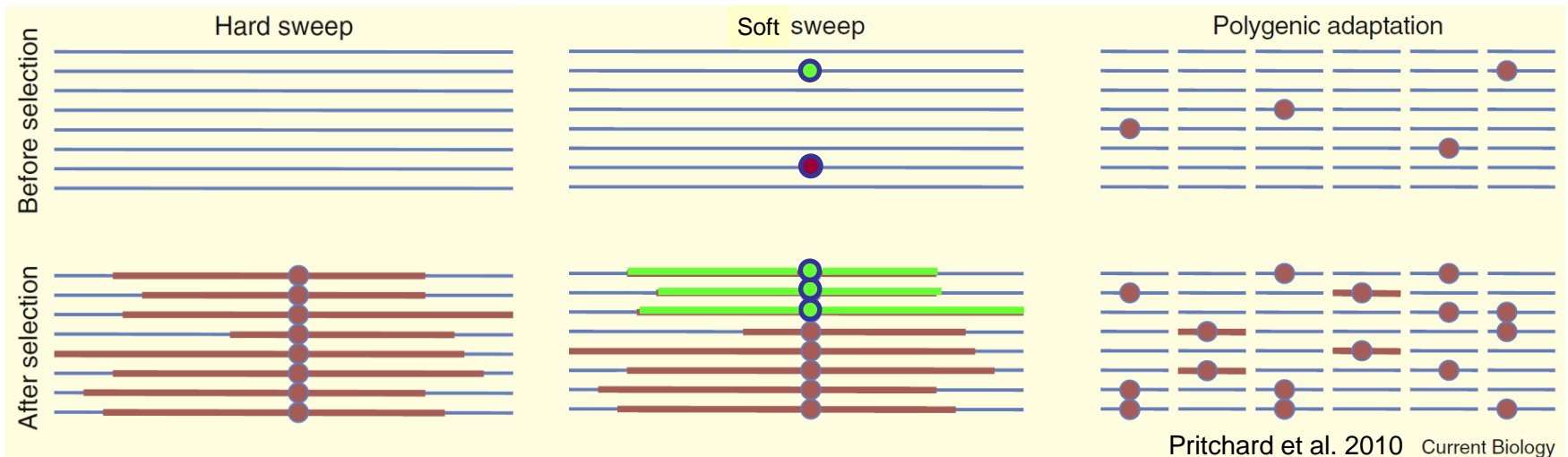
- mutation limited:
from single new mutation
- classical sweep pattern

Soft sweeps

- not mutation limited:
multiple copies from SGV
or recurrent new mutation
- modified sweep pattern

Polygenic adaptation

- adaptation from SGV due
to **subtle frequency shifts**
across many loci
- no sweep patterns



Three adaptive scenarios

... more than polygenic: *omnigenic* !

(June 2017 – more than 320 citations)

Leading Edge

Perspective

Cell

An Expanded View of Complex Traits: From Polygenic to Omnigenic

Evan A. Boyle,^{1,*} Yang I. Li,^{1,*} and Jonathan K. Pritchard^{1,2,3,*}

¹Department of Genetics

²Department of Biology

³Howard Hughes Medical Institute

Stanford University, Stanford, CA 94305, USA

*Correspondence: eaboyle@stanford.edu (E.A.B.), yangili@stanford.edu (Y.I.L.), pritch@stanford.edu (J.K.P.)

<http://dx.doi.org/10.1016/j.cell.2017.05.038>

A central goal of genetics is to understand the links between genetic variation and disease. Intuitively, one might expect disease-causing variants to cluster into key pathways that drive disease etiology. But for complex traits, association signals tend to be spread across most of the genome—including near many genes without an obvious connection to disease. We propose that gene regulatory networks are sufficiently interconnected such that all genes expressed in disease-relevant cells are liable to affect the functions of core disease-related genes and that most heritability can be explained by effects on genes outside core pathways. We refer to this hypothesis as an “omnigenic” model.

Three adaptive scenarios

... *but is it* ?

Reduced signal for polygenic adaptation of height in UK Biobank

Jeremy J. Berg^{*,†,1}, Arbel Harpak^{*,1,2}, Nasa Sinnott-Armstrong^{*,3},
Anja Moltke Jørgensen⁴, Hakhamanesh Mostafavi¹, Yair Field³, Evan A Boyle³, Xinjun Zhang⁵,
Fernando Racimo⁴, Jonathan K Pritchard^{†,2,3,6}, Graham Coop^{†,7}

Signals of polygenic adaptation on height have been overestimated due to uncorrected population structure in genome-wide association studies

Mashaal Sohail^{1,2,3+}, Robert M. Maier^{3,4,5+}, Andrea Ganna^{3,4,5,6,7}, Alex Bloemendal^{3,4,5},
Alicia R. Martin^{3,4,5}, Michael C. Turchin^{8,9}, Charleston W. K. Chiang¹⁰, Joel N.
Hirschhorn^{3,11,12}, Mark J. Daly^{3,4,5,7}, Nick Patterson^{3,13}, Benjamin M. Neale^{3,4,5*}, Iain
Mathieson^{14*}, David Reich^{3,13,15*}, Shamil R. Sunyaev^{1,2,3*}

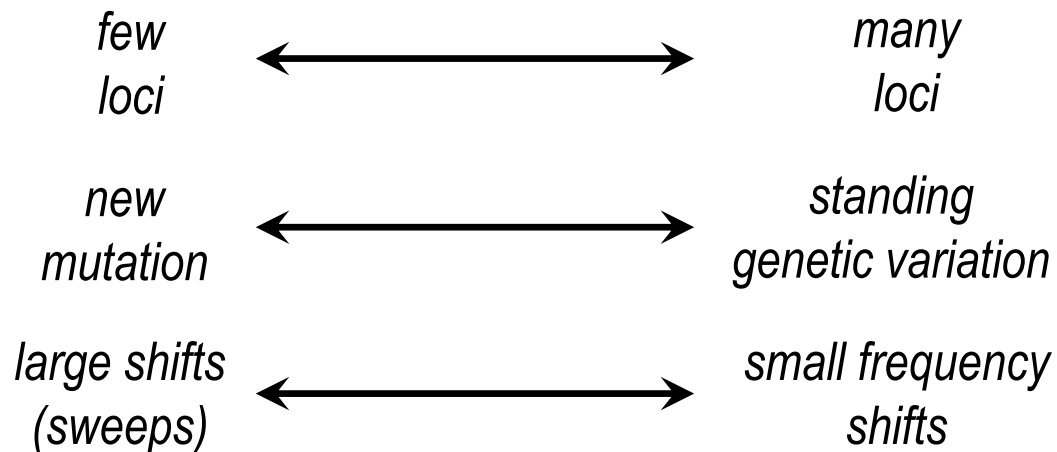
POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Three adaptive scenarios

Which scenario is favored under which conditions?

Consider:

- panmictic population
- adaptation from mutation-selection-drift balance



Depending on:

- mutation rate,
- selection strength,
- number of loci (target size)

Concepts & Models

Consider :

Adaptation at a polygenic trait
(panmictic, constant selection)

What do we need for adaptation
by small shifts at many loci ?

Concepts & Models

Consider :

Additive trait – **no fitness epistasis**, no linkage

- independent single-locus dynamics
- strong selection: *hard or soft sweeps* (overdominance: *partial sweeps*)
- weak selection: “*slow sweeps*” (no clear footprint)
- ***never*** : polygenic small shifts

Adaptation by small shifts is a
collective mode of adaptation
at many ***interacting*** loci

Concepts & Models

Consider :

Trait with **linked** loci, no fitness epistasis

- temporary **interference**, but in the long term still sweeps

Epistatic trait with **redundancy** among loci

- alternative mutations at different loci solve the same task
- diminishing returns (negative epistasis) for fitness
 - e.g. stabilizing selection on quantitative trait,
 - redundancy on the level of genes or networks,
 - saturation effects in metabolic pathways,
 - negative feedbacks, top-down control ...

Patterns of polygenic adaptation

Consider polygenic adaptation in an ideal model for small shifts



Ilse Hölliger
University of
Vienna



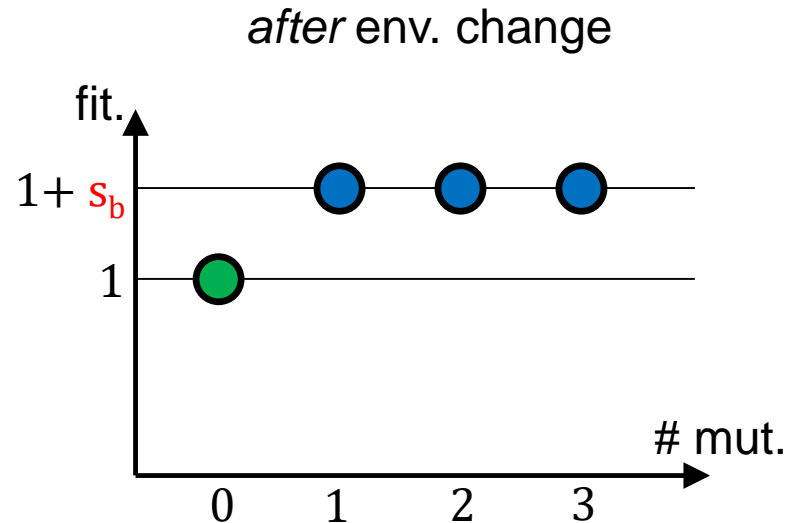
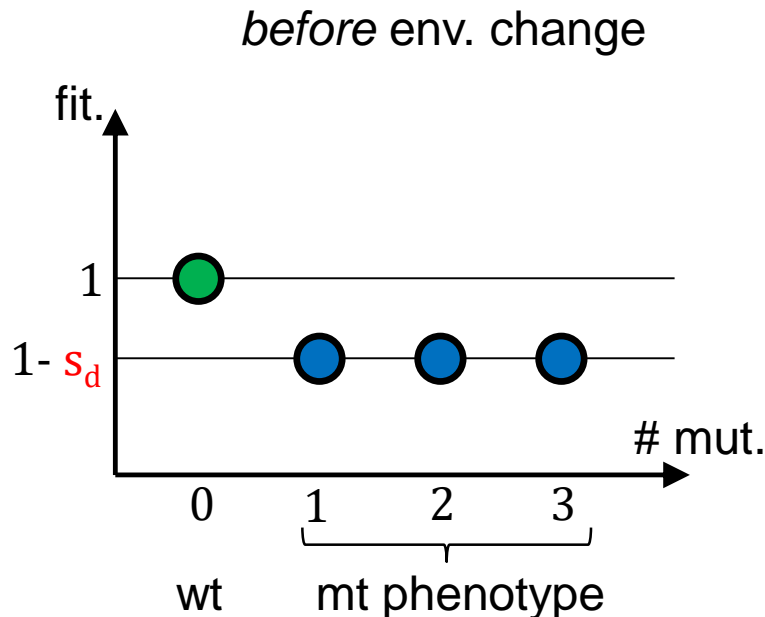
Pleuni Pennings
San Francisco State
University

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Patterns of polygenic adaptation

Basic model: *Binary trait with polygenic basis and complete redundancy*

- N haploids, L biallelic loci, all equal
- recurrent mutation at all loci $\theta = 2Nu$ (no back mut.)
- selection (e.g. resistance):



Patterns of polygenic adaptation

Basic model: *Binary trait with polygenic basis and complete redundancy*

- let trait evolve until mutant phenotype $> 95\%$
- consider:
 - **major locus**: locus with largest mutant frequency (after adaptation)
 - **minor loci**: further loci with mutant frequency > 0

few or many ?
independent or collective ?
sweeps or shifts ?

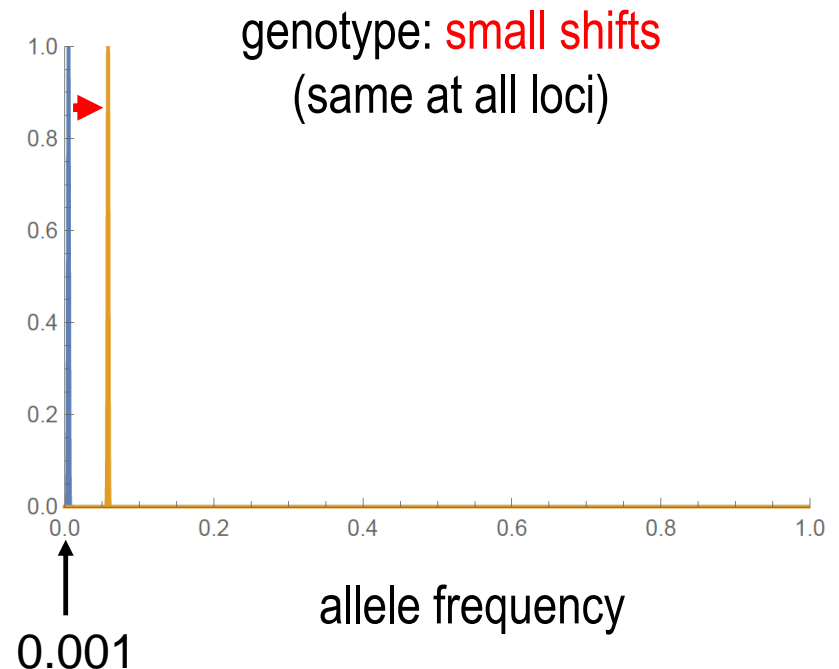
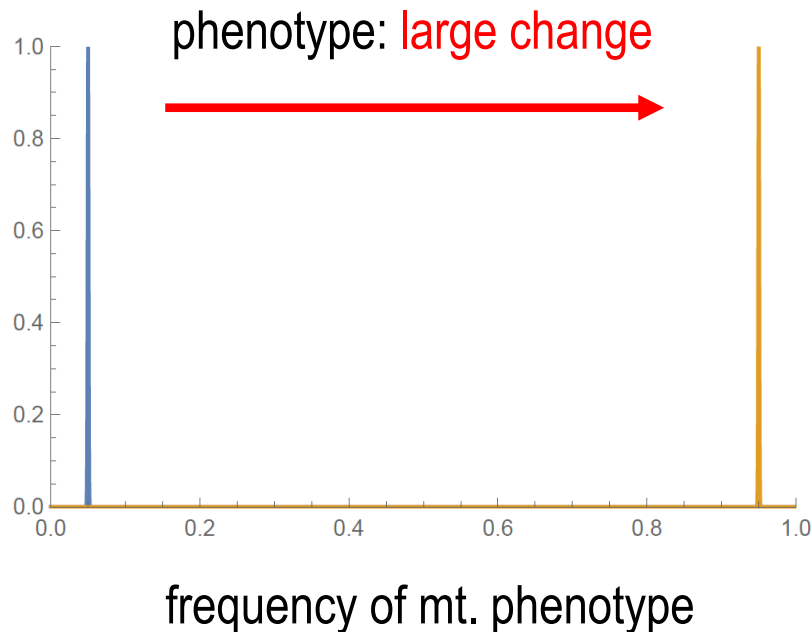
POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Patterns of polygenic adaptation

Adaptation from SGV to 95% mutant phenotype: **deterministic theory**

$L = 50$ loci, $\theta = 0.01$, $2Ns_d = 10$, $2Ns_b = 100$

➤ **archetypical scenario of “small shifts at many loci”**



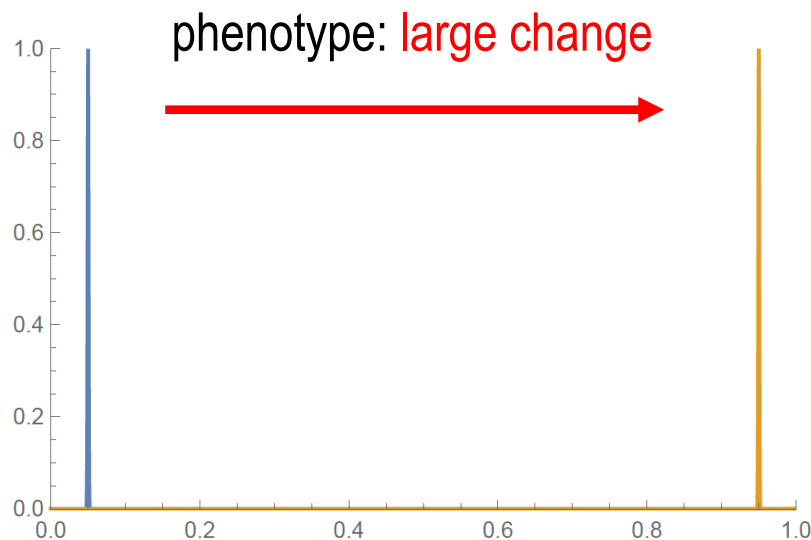
POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Patterns of polygenic adaptation

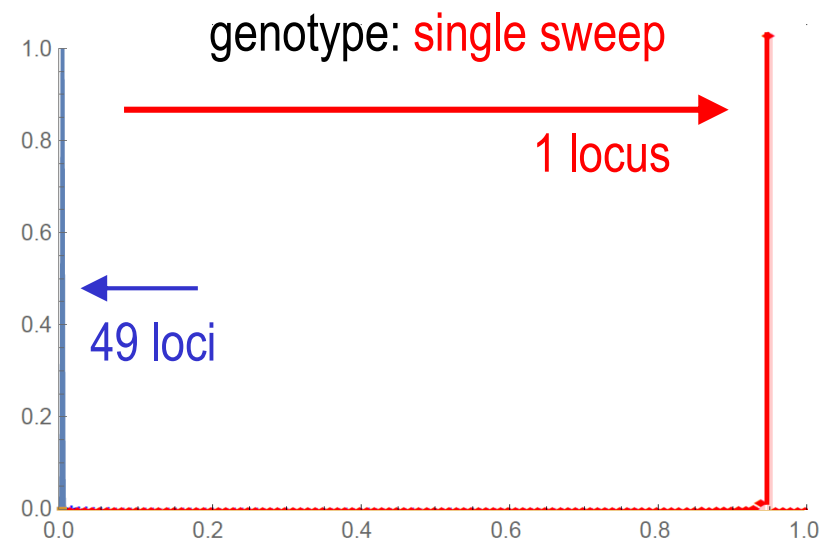
Adaptation from SGV to 95% mutant phenotype: **stochastic simulations**

$L = 50$ loci, $\theta = 0.01$, $2Ns_d = 10$, $2Ns_b = 100$

- **single major locus dominates and sweeps, no “small shifts”**



frequency of mt. phenotype



allele frequency

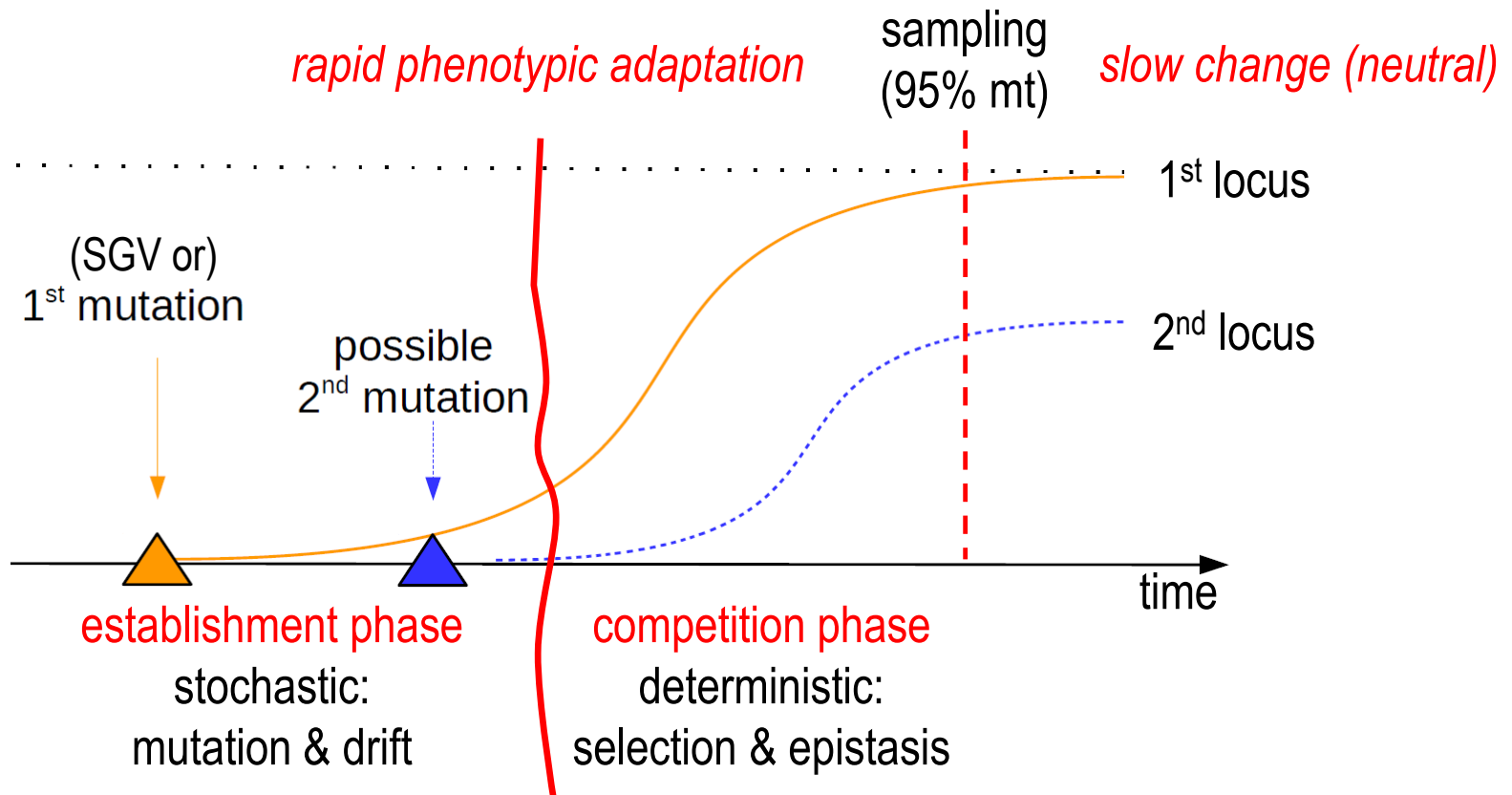
Patterns of polygenic adaptation

- maximal **symmetry**:
 - all loci equal, same starting conditions, no a-priory advantage
 - maximal **redundancy**:
 - loci restrict each other due to negative epistasis
- *Why no “adaptation by small frequency shifts at many loci” ?*

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Patterns of polygenic adaptation

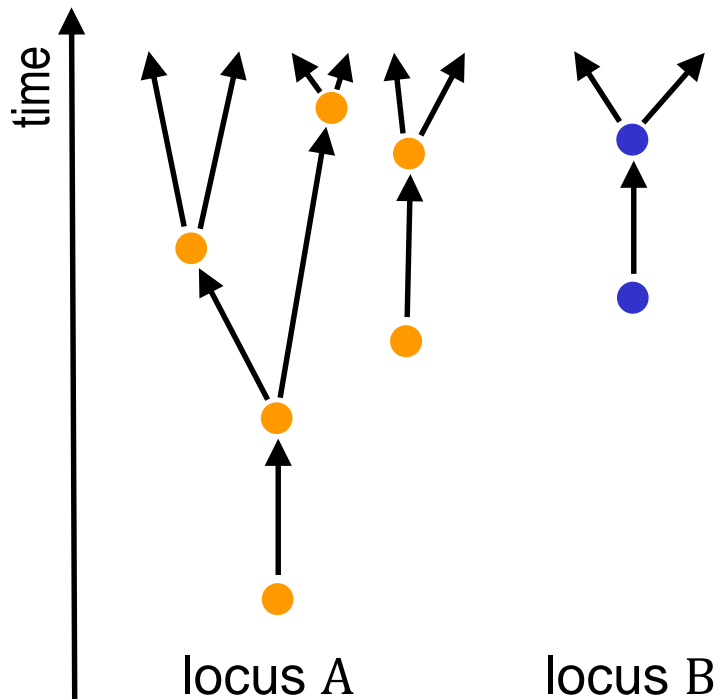
Evolutionary trajectories (2 loci, schematic):



Patterns of polygenic adaptation

Establishment phase: **stochastic origin of mutant clones**

- **new mutant copies** at all loci by **mutation** or **reproduction**
- establishment probability $2s_b$ for each new copy
- track only copies destined for establishment: **"Yule process with immigration"**



split rate s_b per line

immigration rate θs_b per locus

n_A, n_B, \dots copies at loci A, B, ...

➤ joint distribution $\Pr[n_A, n_B, \dots]$
"inverted Dirichlet distribution"

➤ ratios n_B/n_A **independent of s_b**

*mutation and drift during establishment
 create **stochastic differences** among loci*

Patterns of polygenic adaptation

2. Competition phase

- deterministic allele frequency changes due to **epistatic selection** (ignore mutation)

$$\dot{p}_A = s_b p_A (1 - p_A)(1 - p_B) + s_b p_A D$$

$$\dot{p}_B = s_b p_B (1 - p_A)(1 - p_B) + s_b p_B D$$

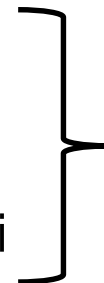
(D linkage disequilibrium)

$$\Rightarrow \frac{d}{dt} \left(\frac{p_B}{p_A} \right) = 0$$

➤ selection among redundant alleles

- maintains ratios n_B/n_A
- zooms up differences $n_A - n_B$ created during the establishment phase

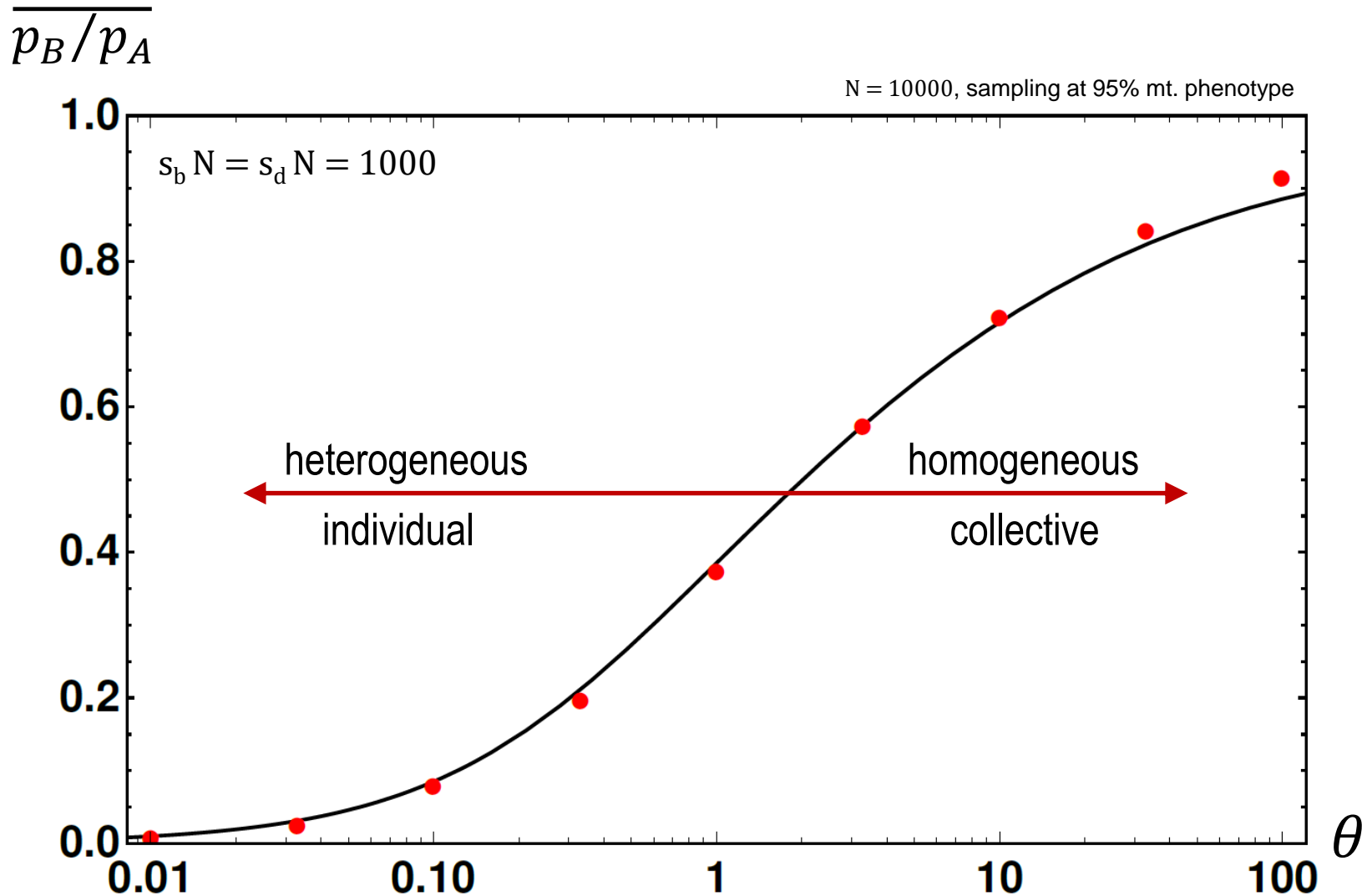
- **independent of s_b and D**
- independent of the stopping condition
- **L loci**: all allele frequency ratios maintained, one major locus & L-1 independent minor loci



➤ strong major/minor locus structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)



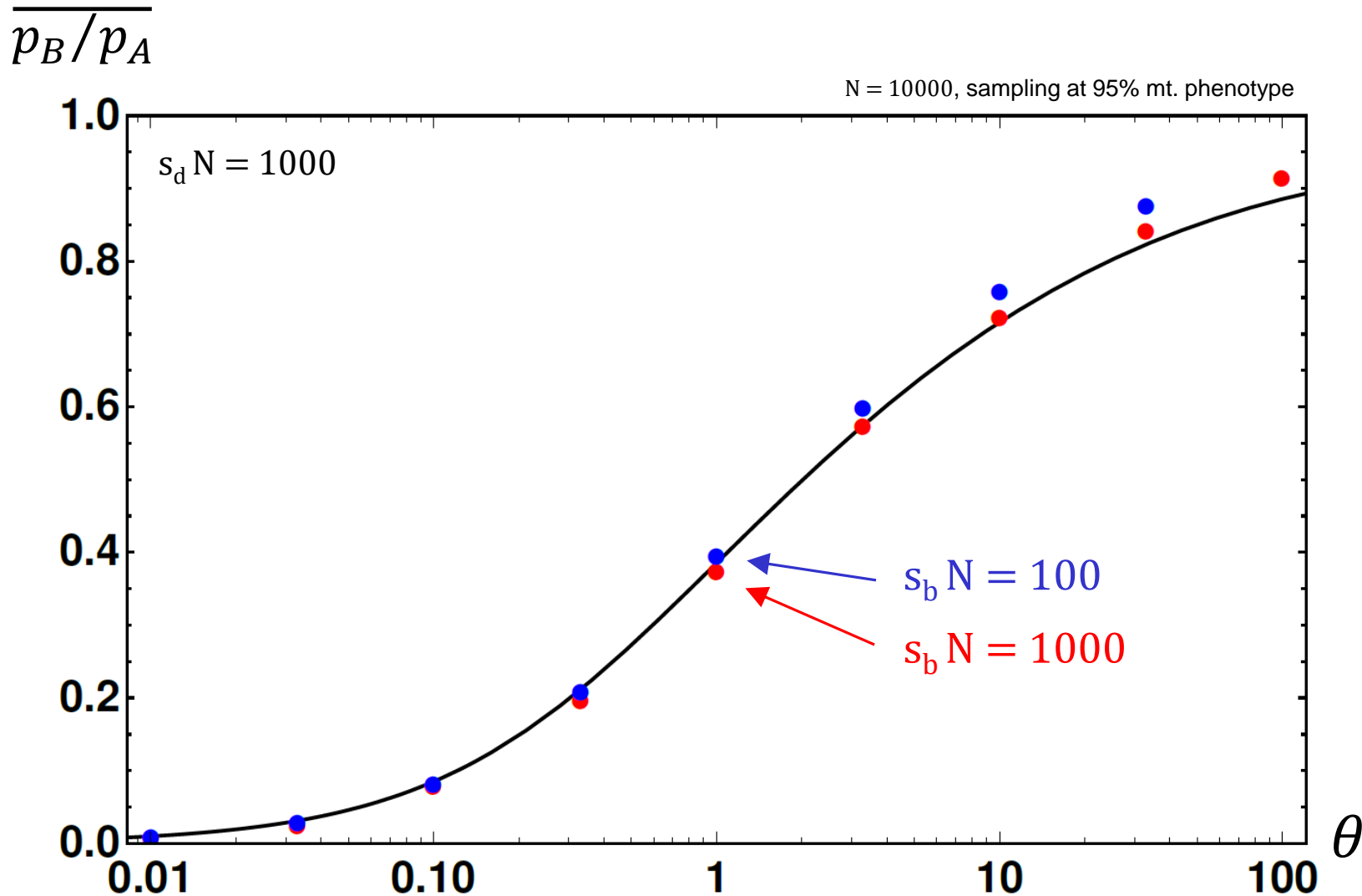
POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)

Influence of selection?

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)



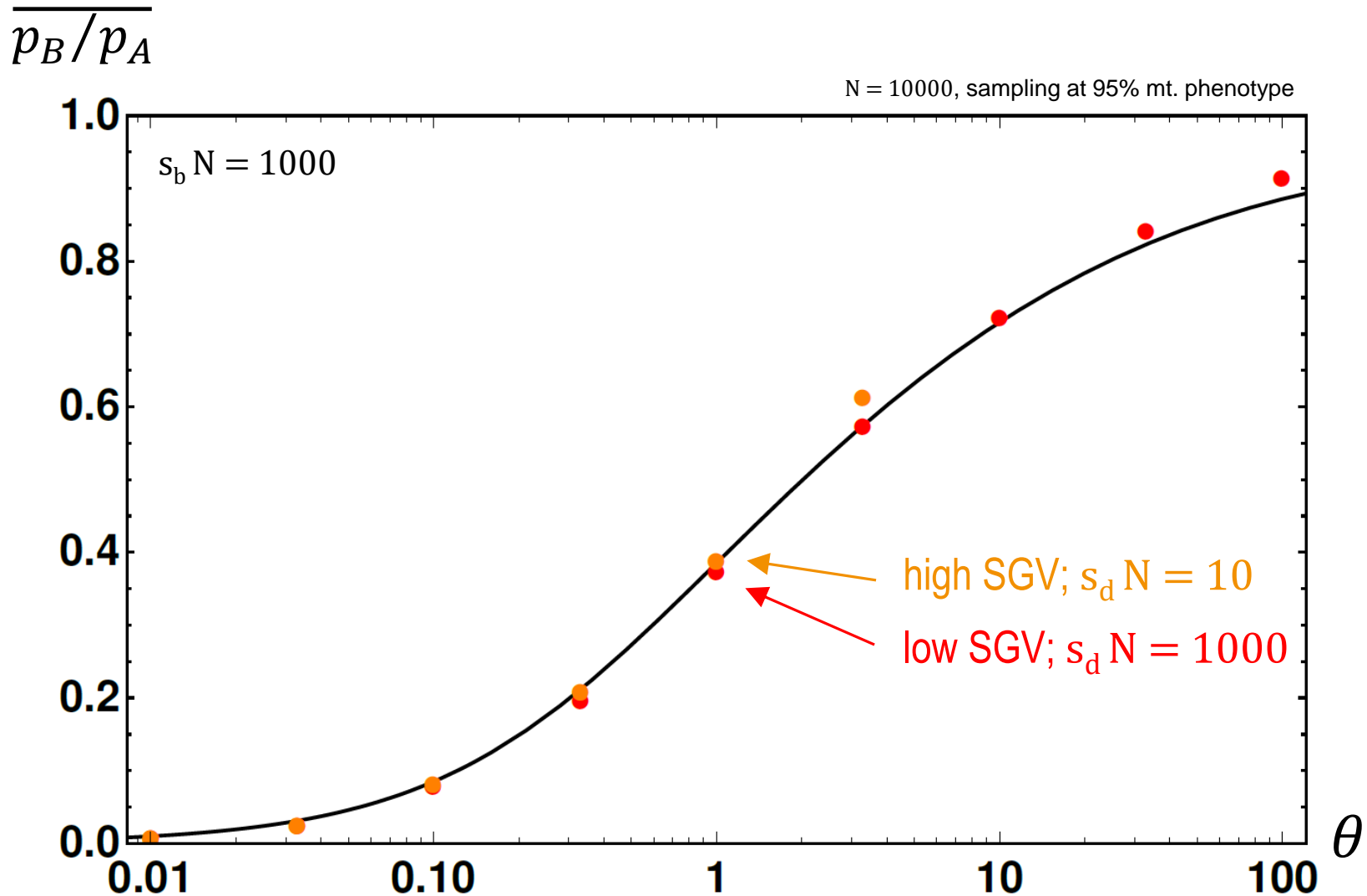
POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)

How about adaptation
from standing genetic variation?

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)

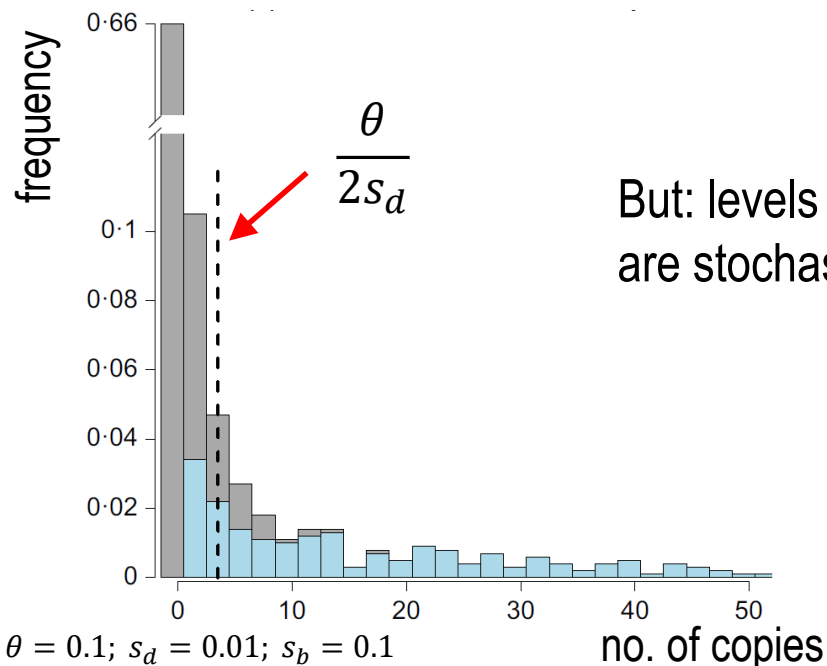
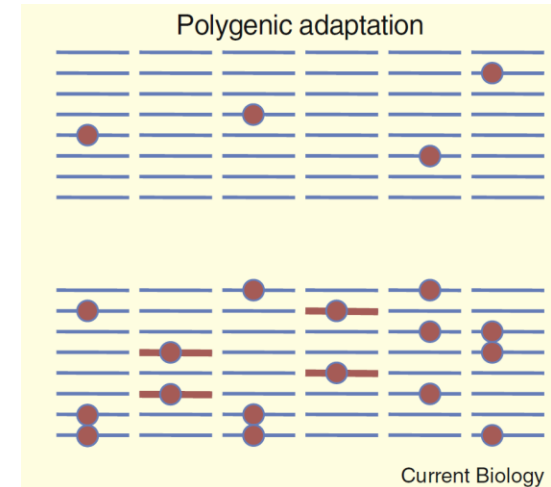


POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Relative adaptive response (2 loci)

Equal levels of SGV $N \frac{u}{s_d} = \frac{\theta}{2s_d}$ at many loci ...

... should lead to equal proportions n_i/n_j after adaptation



But: levels of SGV in mutation-selection-drift balance are stochastic and have a large variance

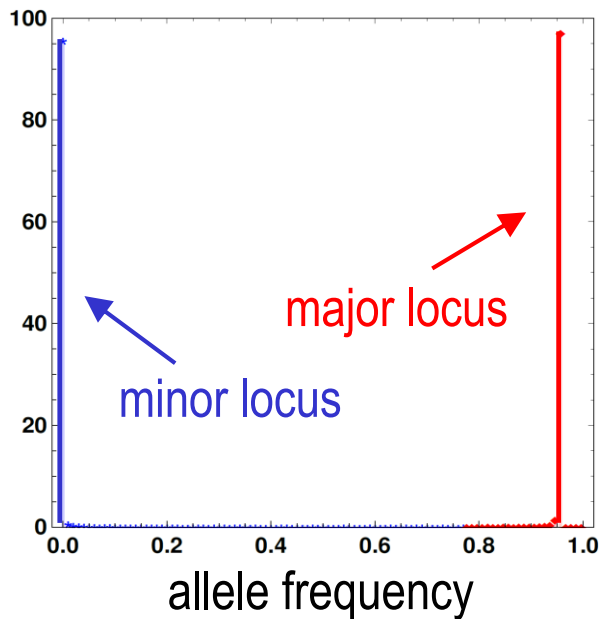
- Same major-minor locus structure as for new mutation

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

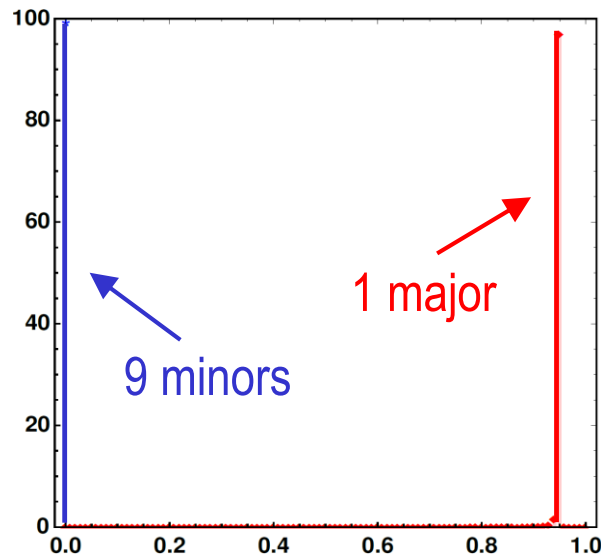
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta = 0.01 \longrightarrow \theta_{Bg} = (L - 1)\theta = 0.01 \quad (\approx \text{“genome-wide } \theta \text{”})$$

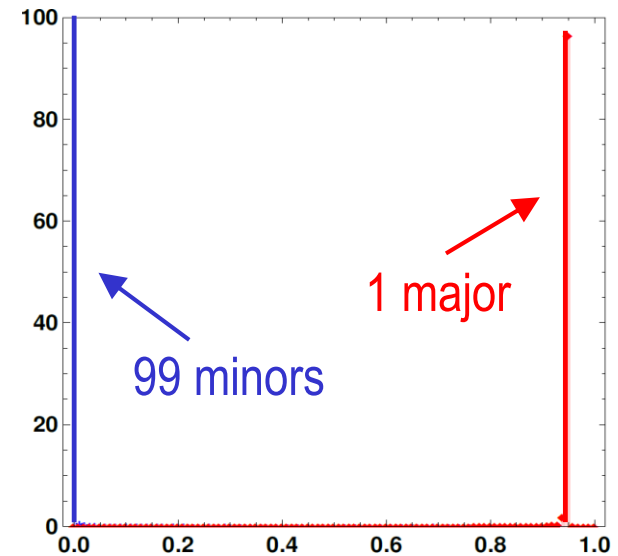
2 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

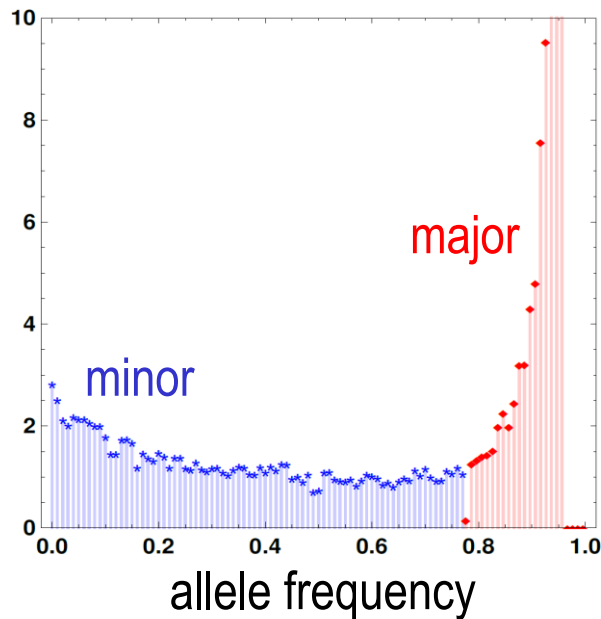
sweep @ single major locus

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

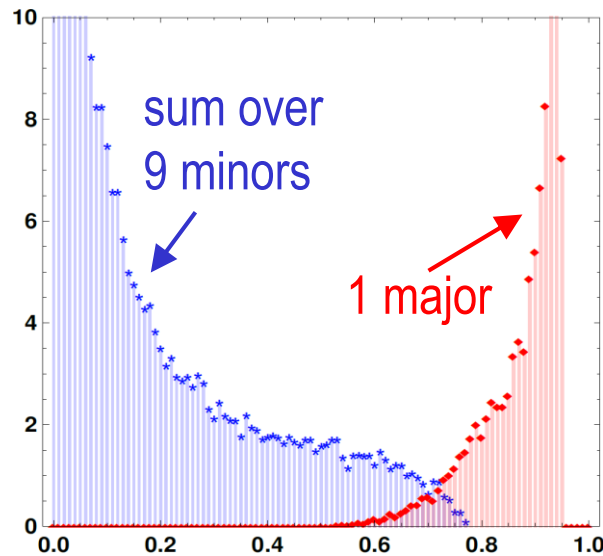
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta_{Bg} = (L - 1)\theta = 1$$

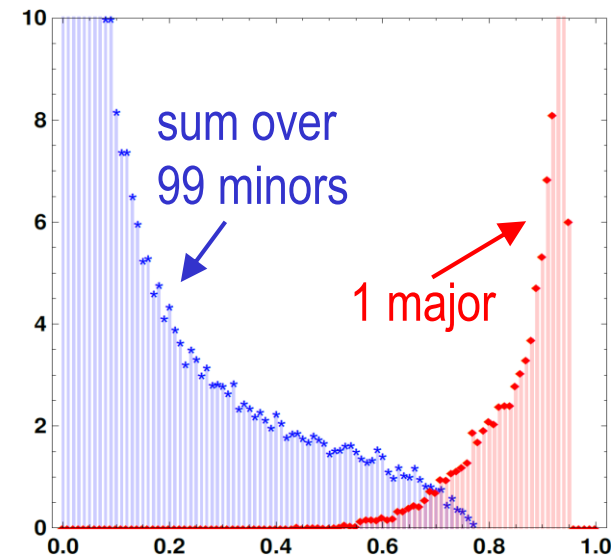
2 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

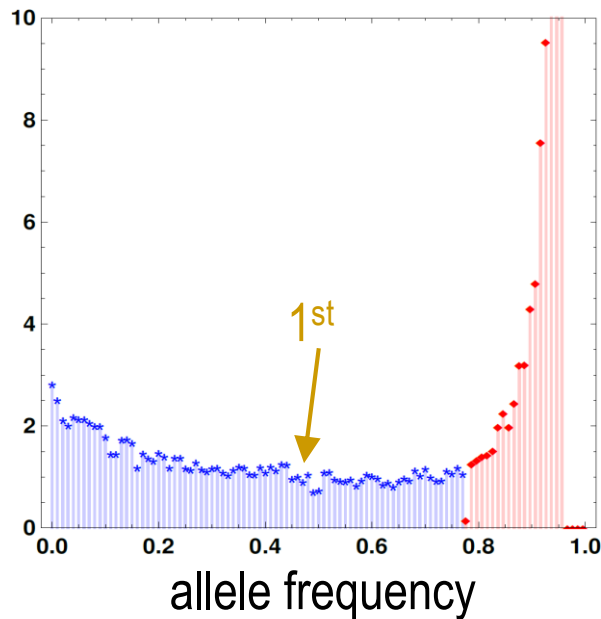
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

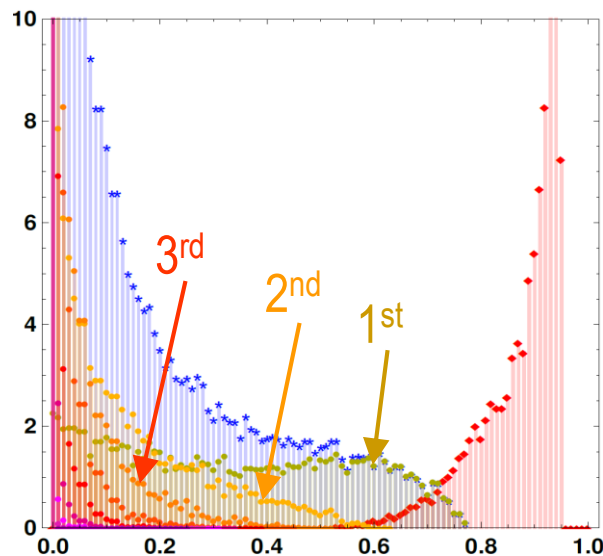
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta_{Bg} = (L - 1)\theta = 1$$

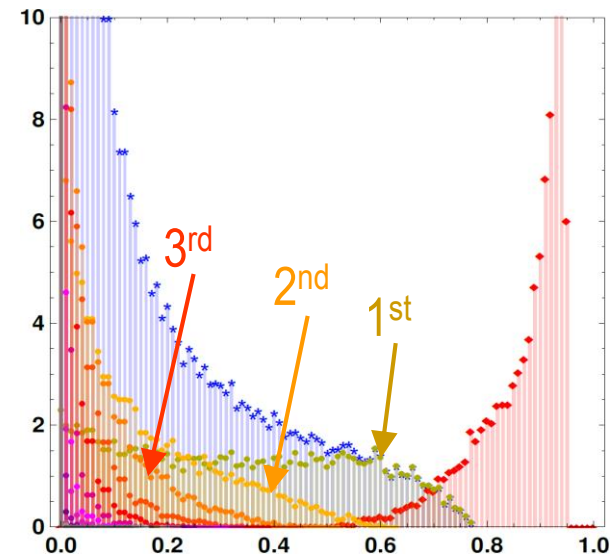
2 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

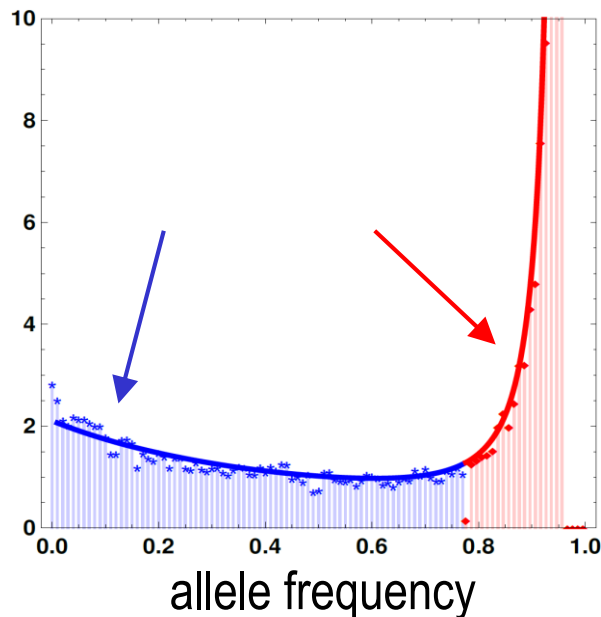
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

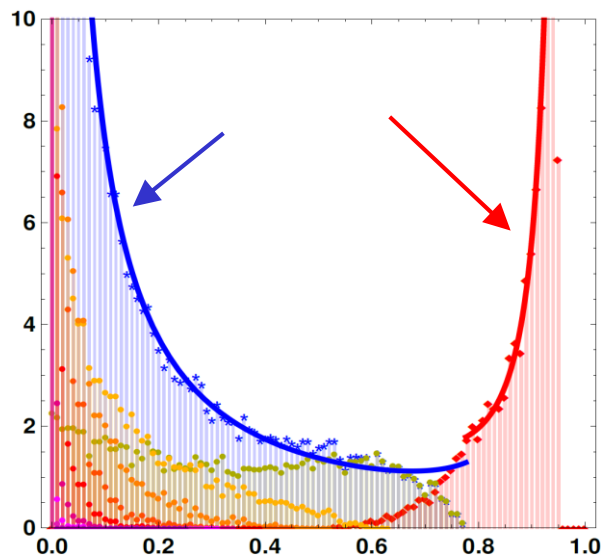
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta_{Bg} = 1, \text{Yule approximation}$$

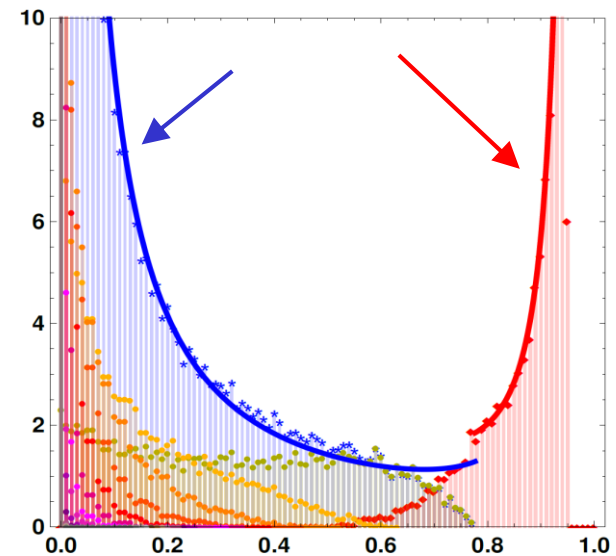
2 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

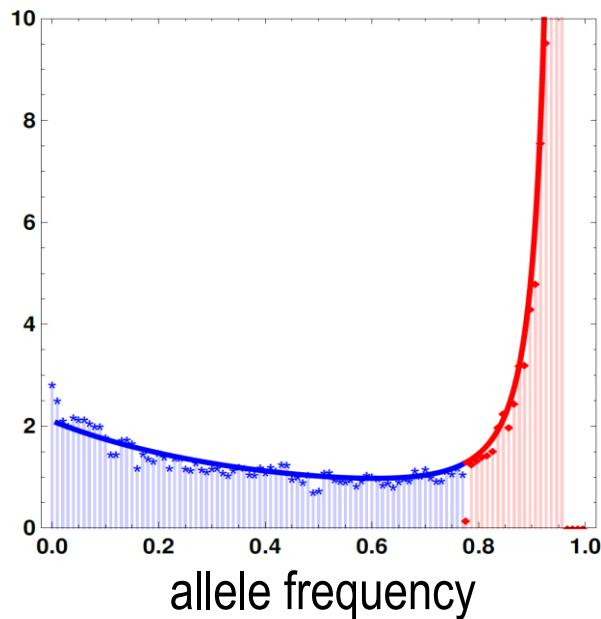
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

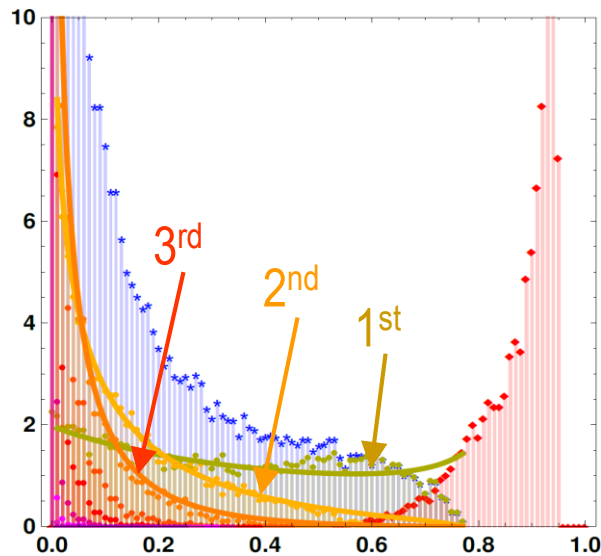
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta_{Bg} = 1, \text{Yule approximation}$$

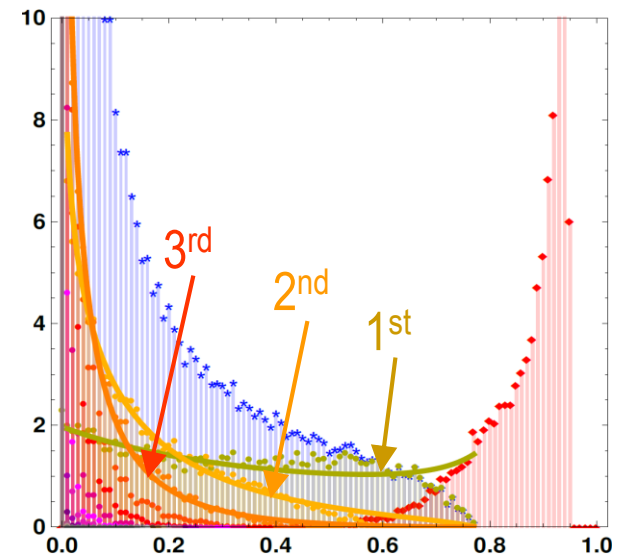
2 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

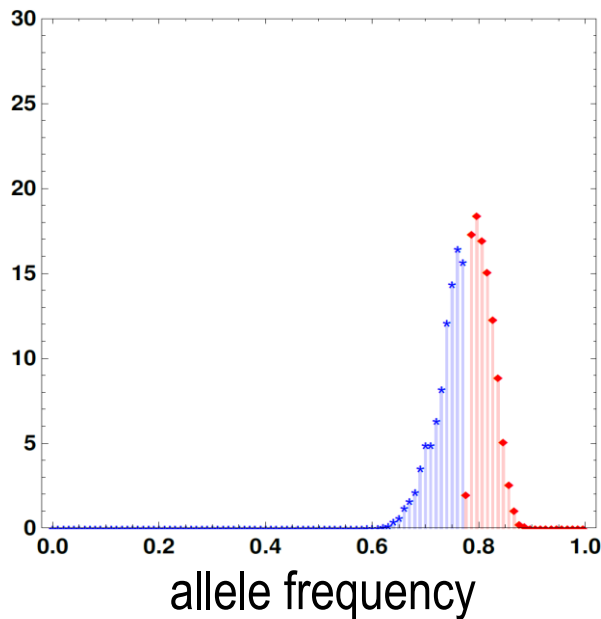
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

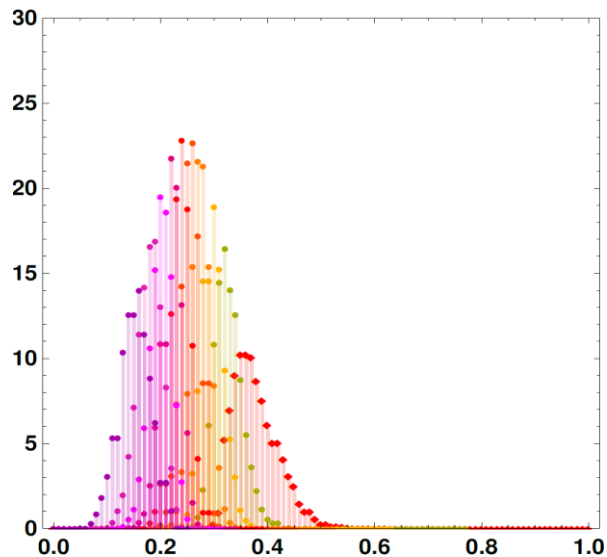
Architecture of polygenic adaptation allele frequencies with complete redundancy

$$\theta_{Bg} = (L - 1)\theta = 100$$

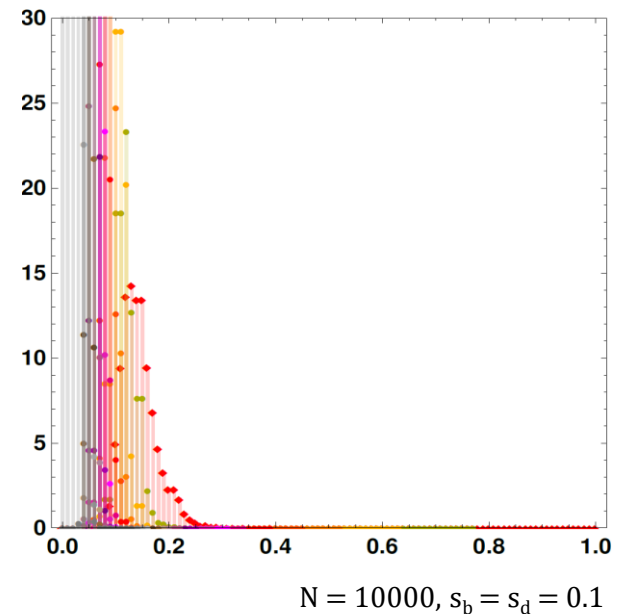
2 loci



10 loci



100 loci

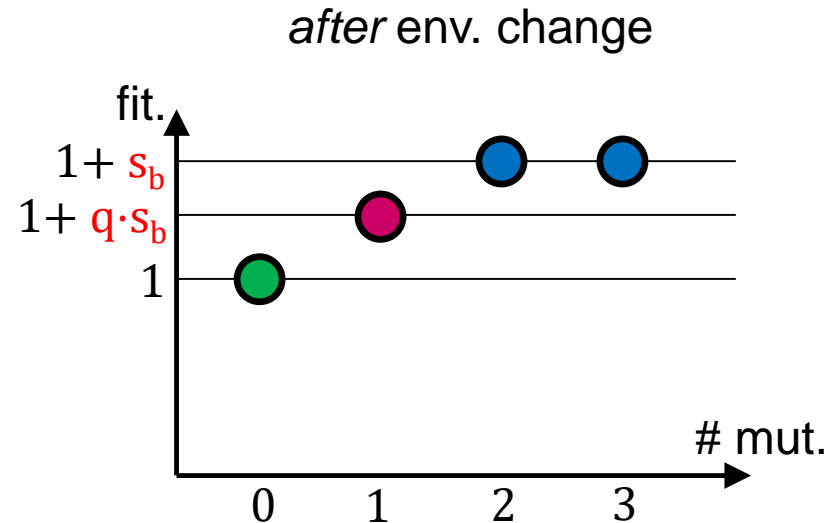
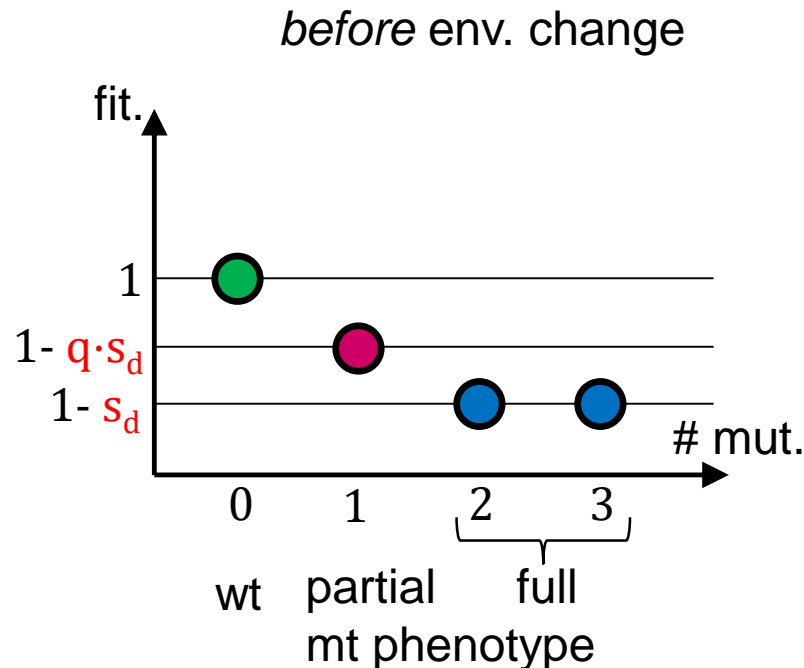


shifts @ many loci

Architecture of polygenic adaptation

Generalized model: *relaxed redundancy*

- multiple mutations needed for complete adaptation
- selection (e.g. saturated response):
 - $q \leq 1$ measures effect of single mutants



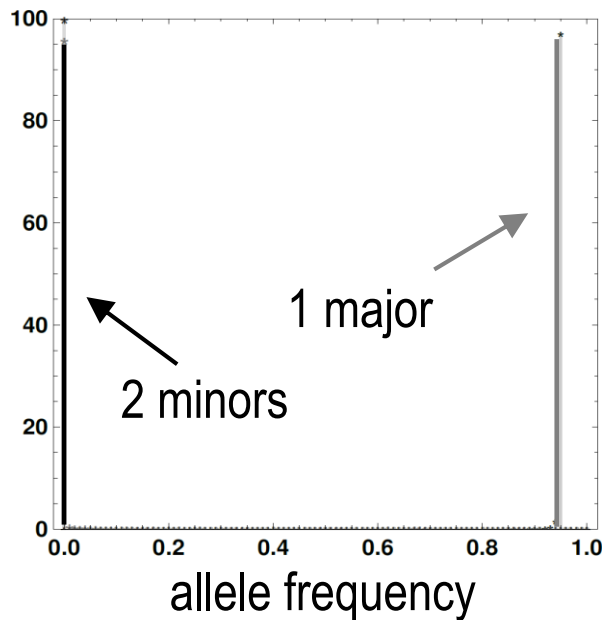
➤ *sample at 95% max. mean fitness*

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

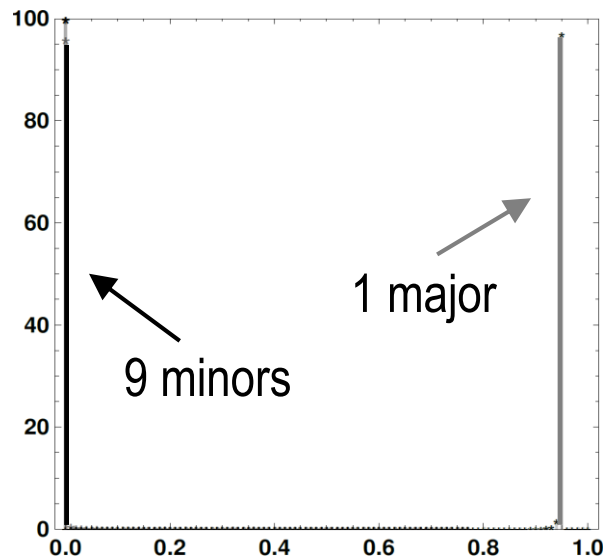
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 0.01; q = 1$$

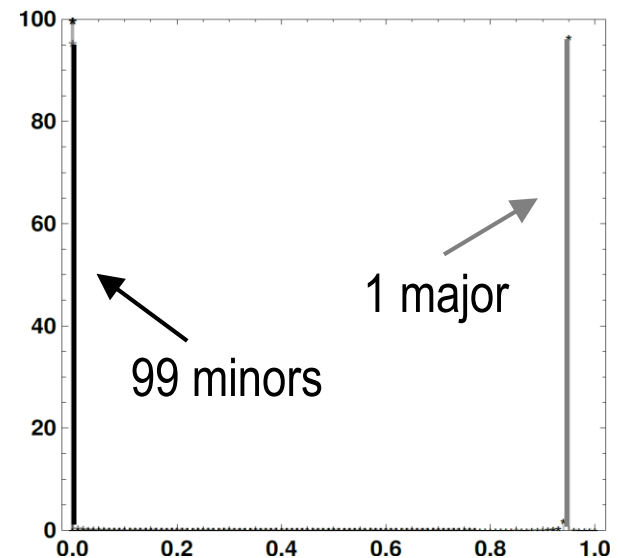
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

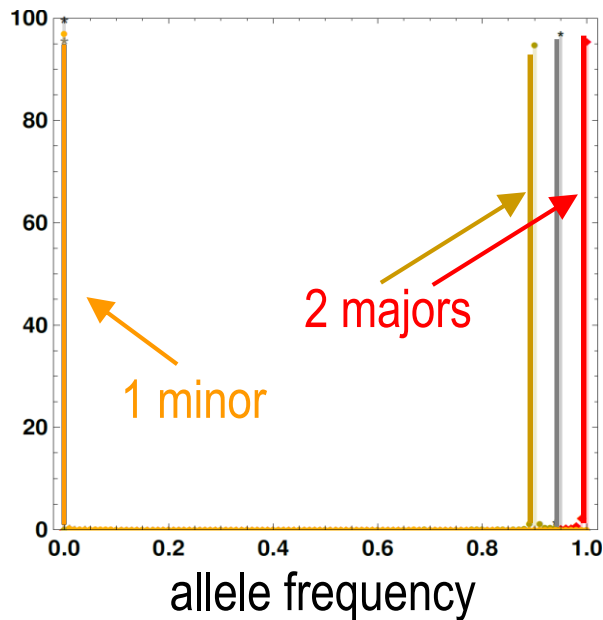
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

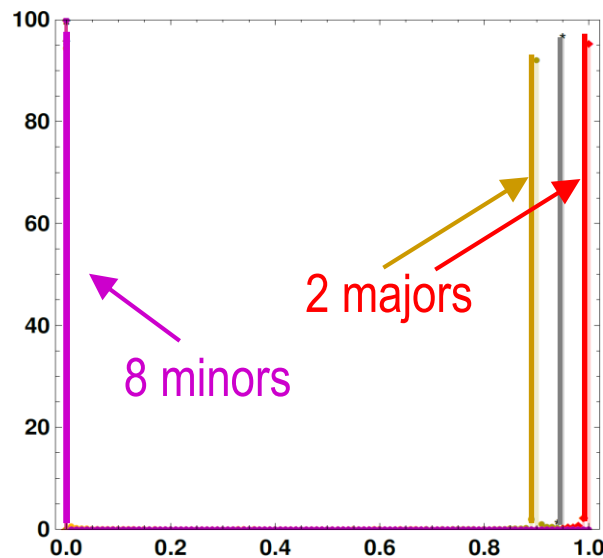
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 0.01; q = 1 \text{ \& } q = 0.5$$

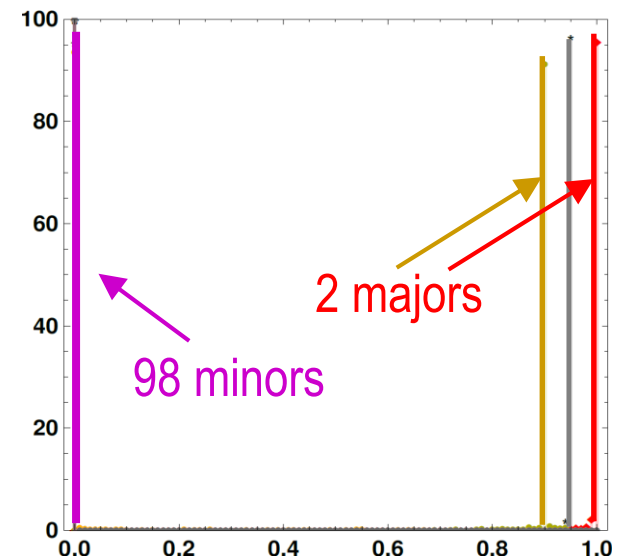
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

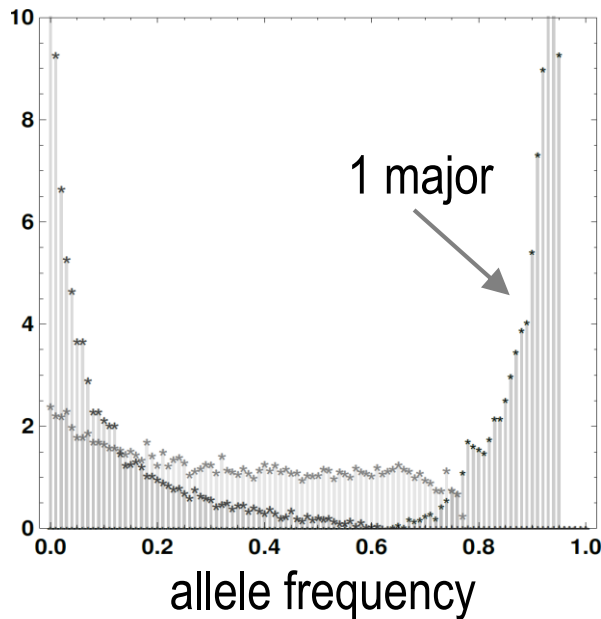
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

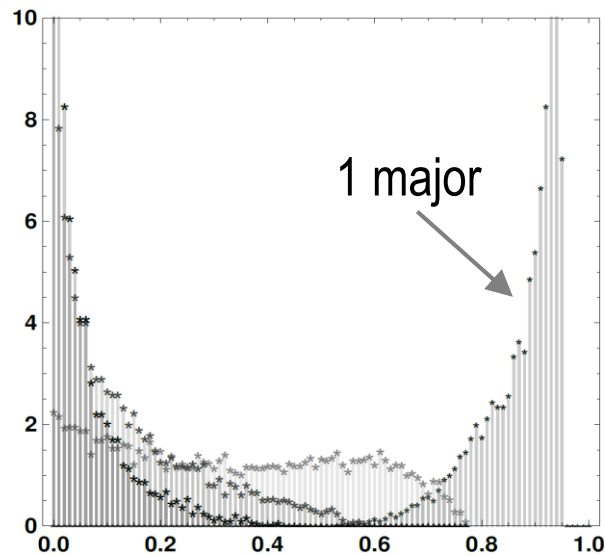
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 1; q = 1$$

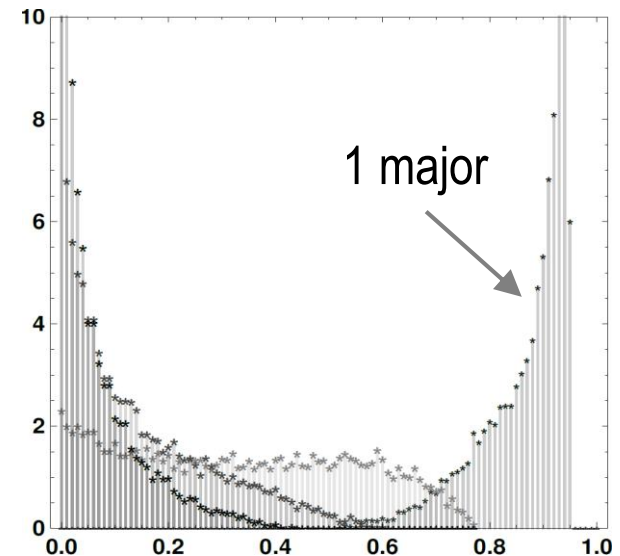
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

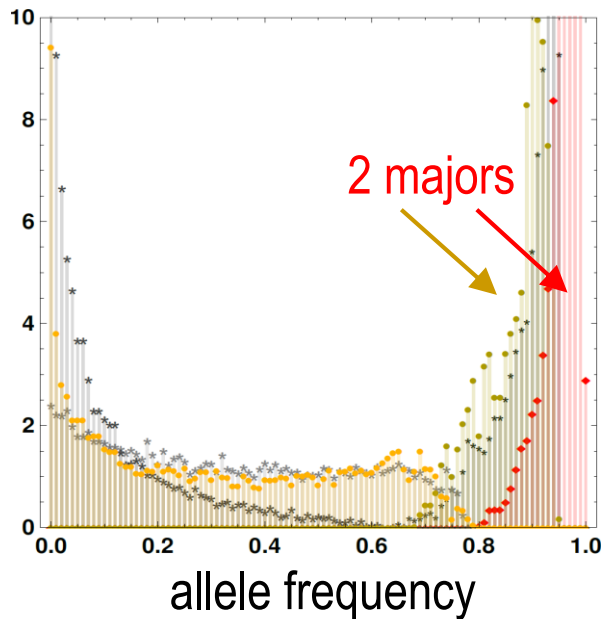
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

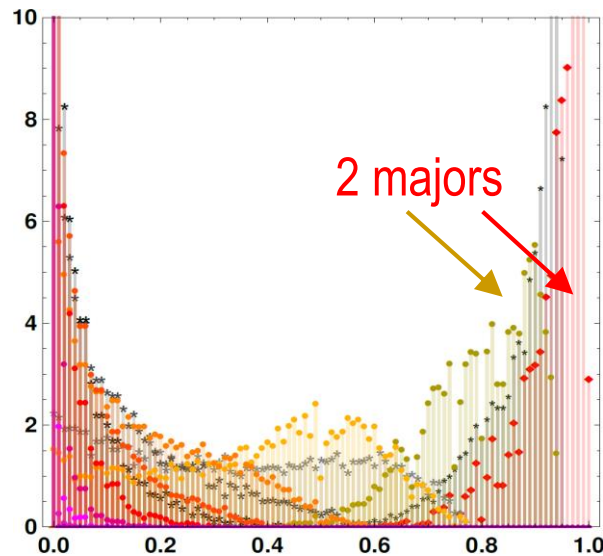
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 1; q = 1 \text{ \& } q = 0.5$$

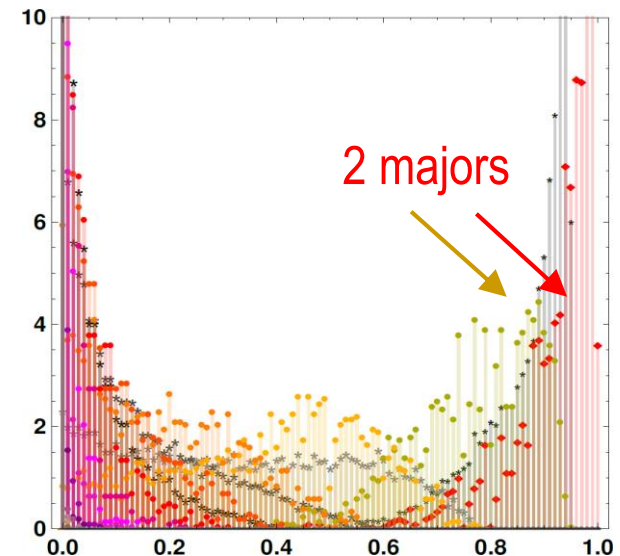
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

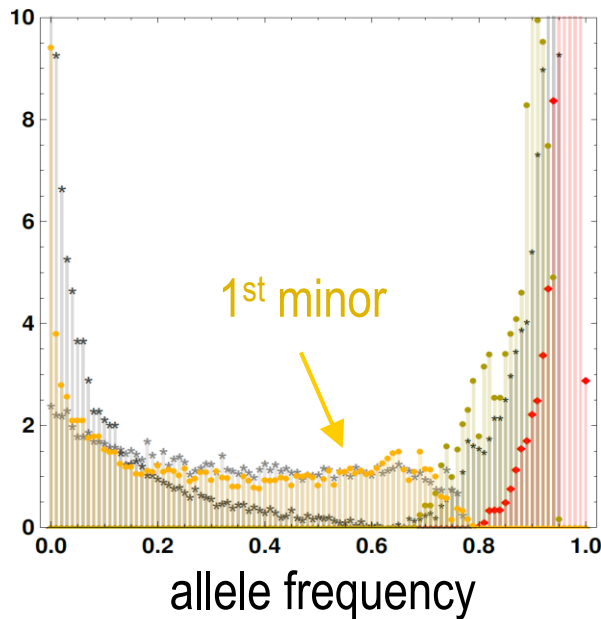
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

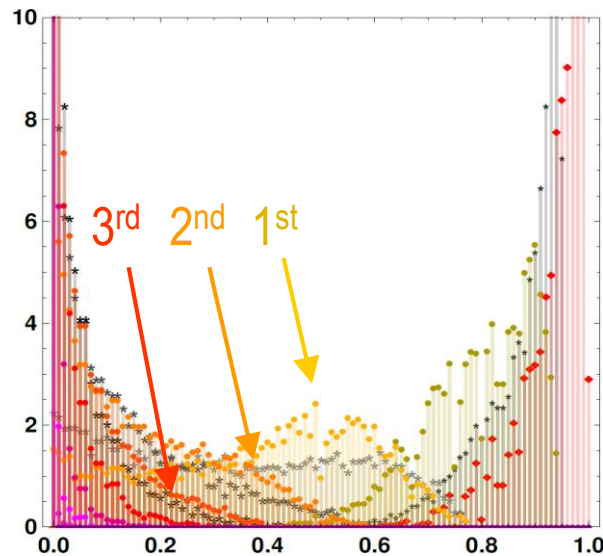
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 1; q = 1 \text{ \& } q = 0.5$$

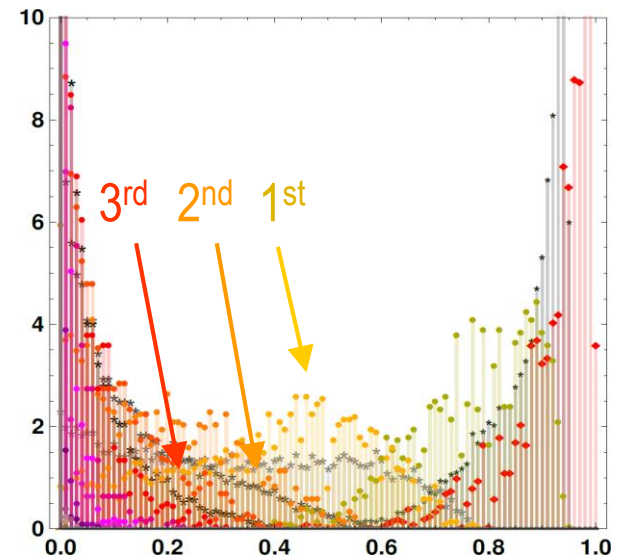
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

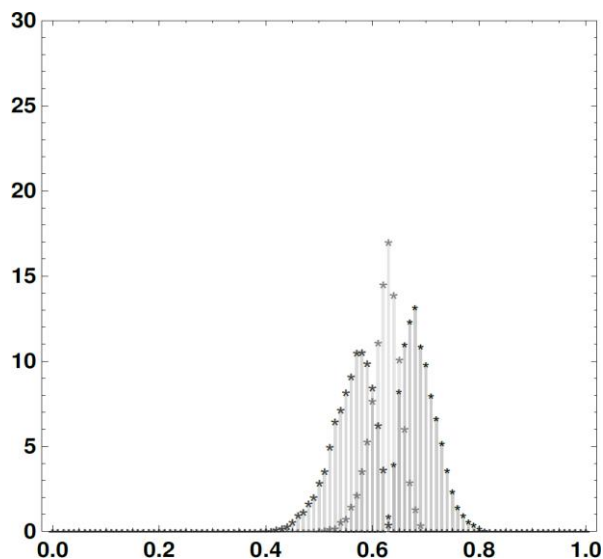
partial sweeps & strong major/minor structure

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

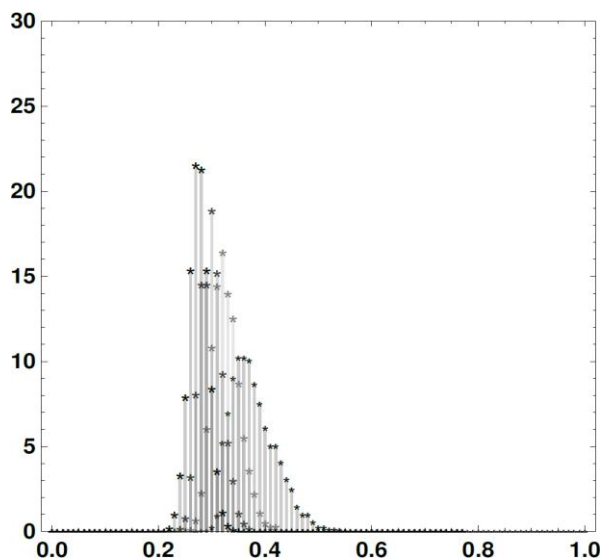
$$\theta_{Bg} = 100; q = 1$$

3 loci

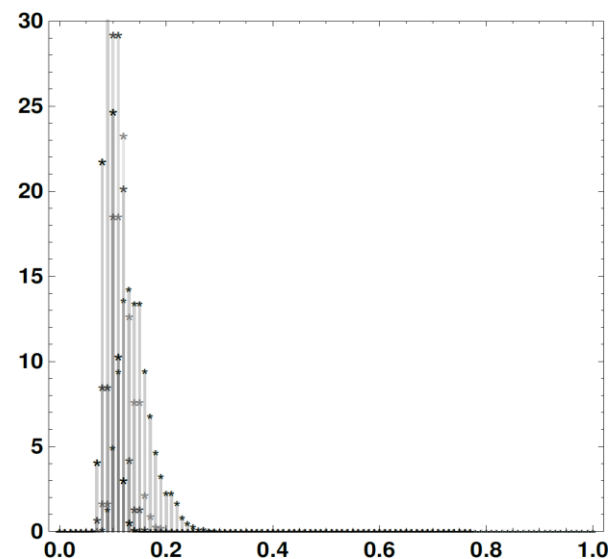


allele frequency

10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

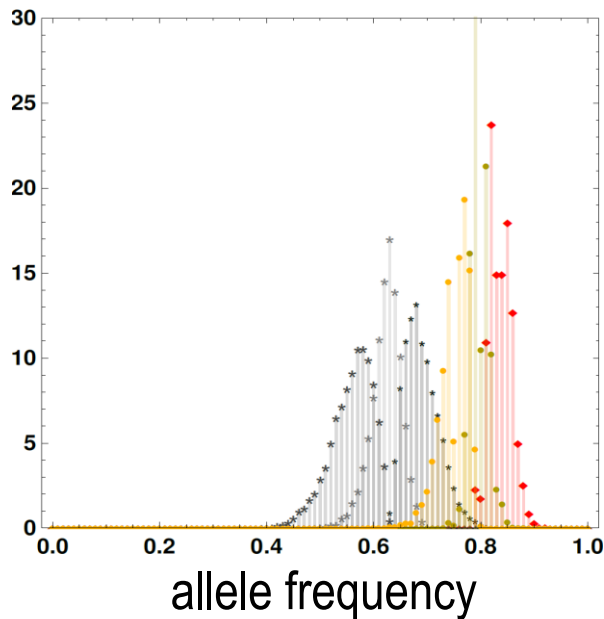
shifts @ many loci

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

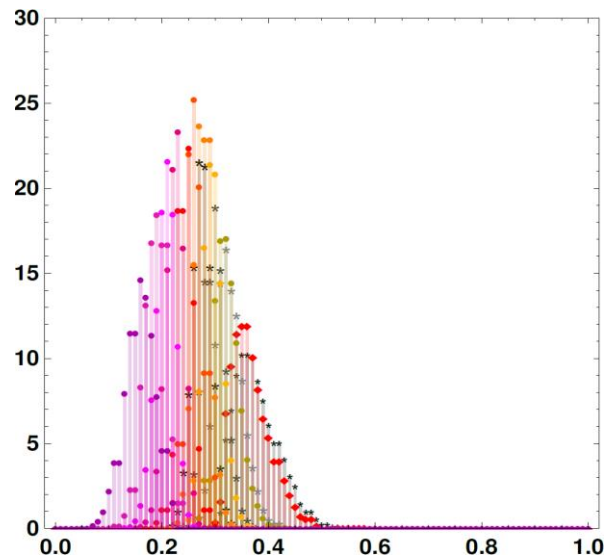
Architecture of polygenic adaptation allele frequencies with **relaxed** redundancy

$$\theta_{Bg} = 100; q = 1 \text{ \& } q = 0.5$$

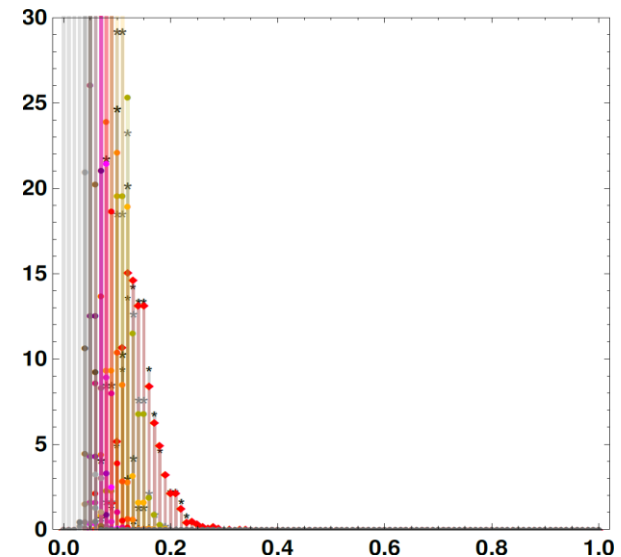
3 loci



10 loci



100 loci



$N = 10000, s_b = s_d = 0.1$

shifts @ many loci

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

What determines the pattern of polygenic adaptation?

- Pattern (largely) independent of selection
 - positive selection during adaptive phase s_b
 - deleterious selection on standing variation s_d
 - independent of linkage among selected loci D
- Redundancy
 - more sweeps for relaxed redundancy q
- Almost all depends on the mutation rate
 - total mutational input across loci $[(L-1)\theta]$ θ_{Bg}

POLYGENIC ADAPTATION: SWEEPS & SHIFTS

Three scenarios of polygenic adaptation for traits with strong redundancy

Assume: high level of redundancy

- diminishing returns epistasis for fitness

$\theta_{Bg} \lesssim 0.1$: **sweep** @ single major locus

- usually hard sweep from new mutation

$0.1 \lesssim \theta_{Bg} \lesssim 10$: **major-minor locus pattern** of adaptation

- (almost) completed sweep from SGV at major locus
- partial (hard or soft) sweeps at several minor loci

$\theta_{Bg} \gtrsim 10$: **small frequency shifts** @ many loci

- no clear selection footprint in linked variation

Three scenarios of polygenic adaptation for traits with strong redundancy

How should we **explain evidence for** adaptation by **small shifts** (size / weight / yield traits)?

- really “small shifts”?
 - slow / incomplete sweeps at small-effect loci ?
- $\theta_{Bg} \approx 2LN_e u$ is large:
 - large “omnigenic” basis $L > 10000$?
 - large “short-term N_e ” ?
- initial allele frequencies more homogeneous than predicted by mutation-selection-drift balance
 - **balancing selection** ? (but implies **constraint**)
 - **spatial structure** or **admixture** ?

Three scenarios of polygenic adaptation for traits with strong redundancy

More than **sweeps** or **shifts**:

- pattern of stalled partial sweeps
 - predicted in large and relevant parameter region
 - strongly heterogeneous even among loci with identical effect
 - should also be heterogeneous among replicates / for parallel adaptation (“zoomed-up stochasticity”)

Evidence ?

- a lot of evidence for partial sweeps
 - strong completed sweeps are rare
- plateauing of allele trajectories in experimental evolution
 - strongly heterogeneous among replicates

Thanks!



Ilse Hölliger



Pleuni Pennings

