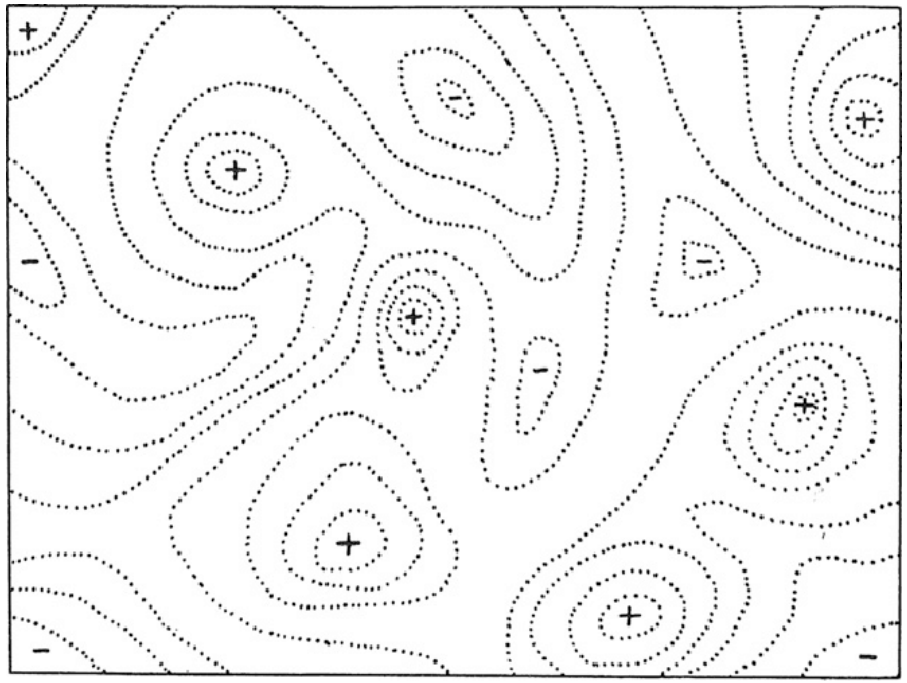


Evolution of quantitative traits

Hancock

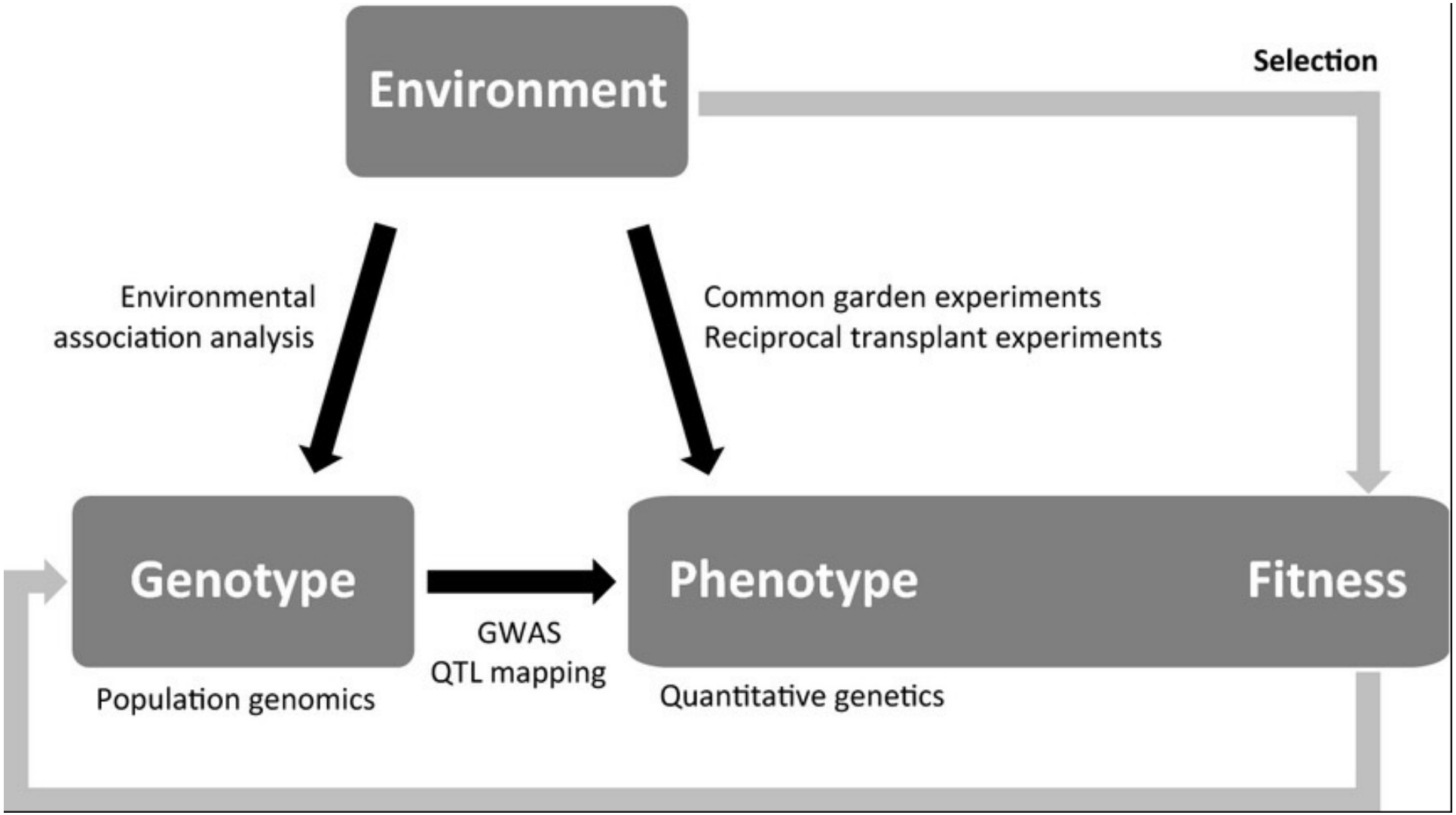
23. April 2024

Evolution of quantitative traits

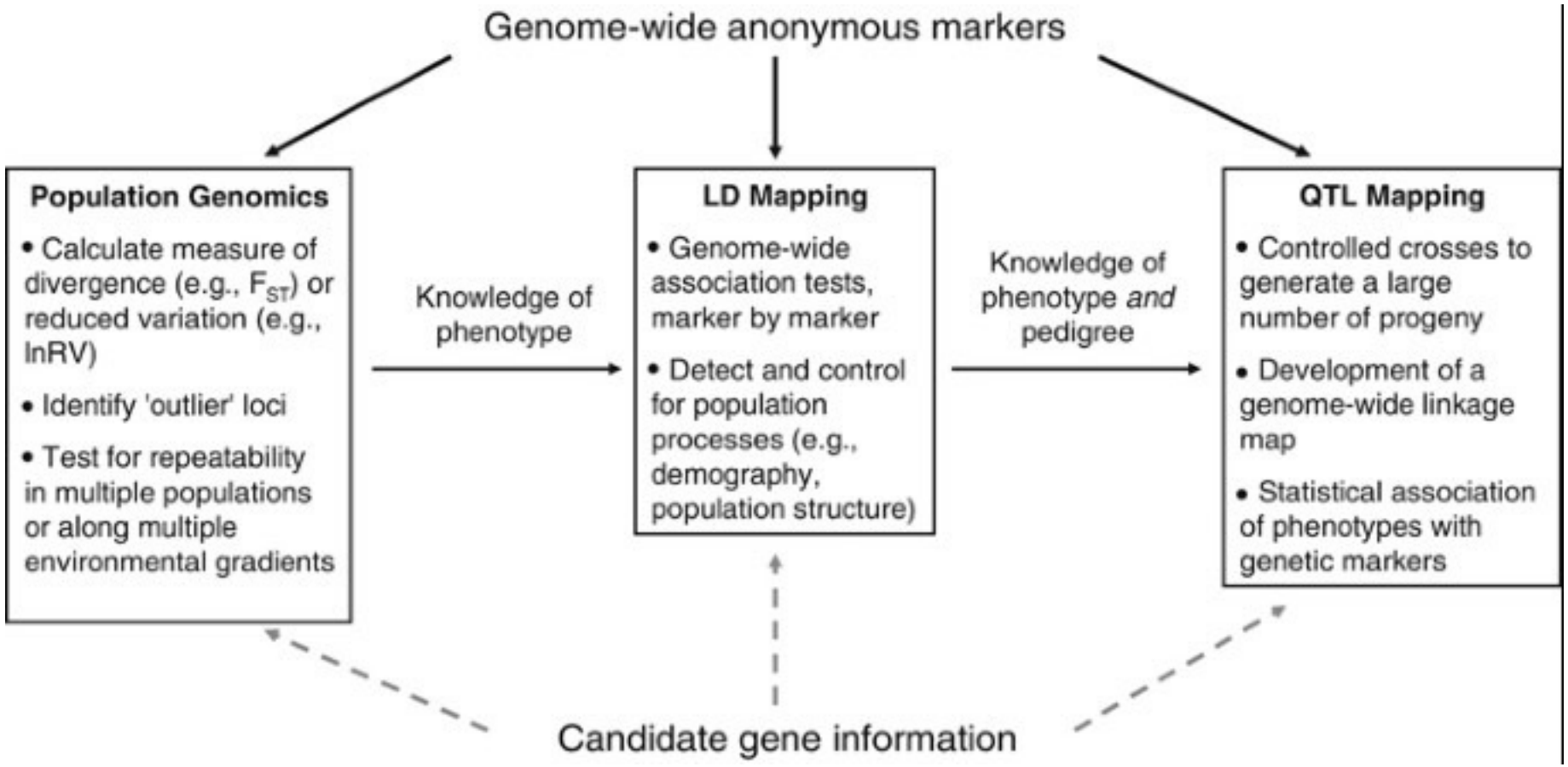
What is the genetic architecture of quantitative traits?

- Do common or rare variants underlie phenotypic variation?
- Do many or few loci contribute to traits?
- Are effects of individual variants strong or weak?
- How important are GxE and GxG effects?
- Where do trait-associated variants localize in biological networks?
- How does the architecture of adaptive traits differ from that of non-adaptive (neutral) traits?
- Do similar phenotypes arise via changes in the same genes, …via changes in the same pathways?

Combining population genomics and quantitative genetics to understand evolutionary mechanisms



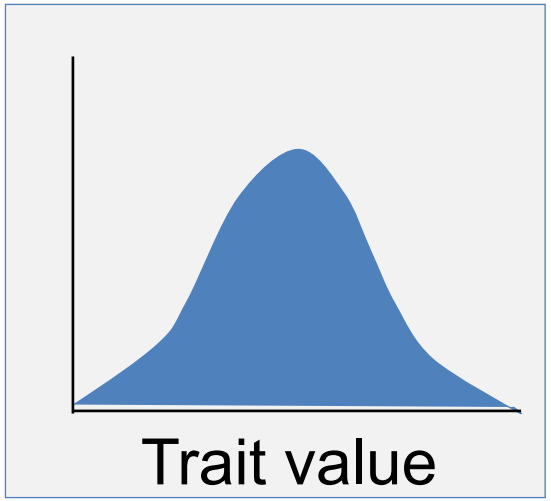
Rellstab et al., 2015



Stinchcombe and Hoekstra, 2007

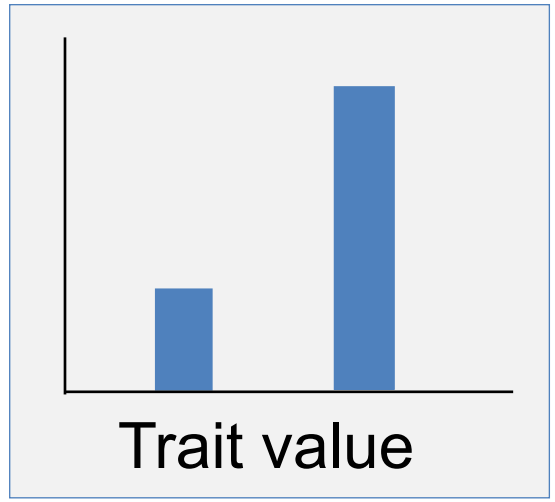
What is the nature of trait variation? Two views:

Continuous variation



Many variants, each with a small effect

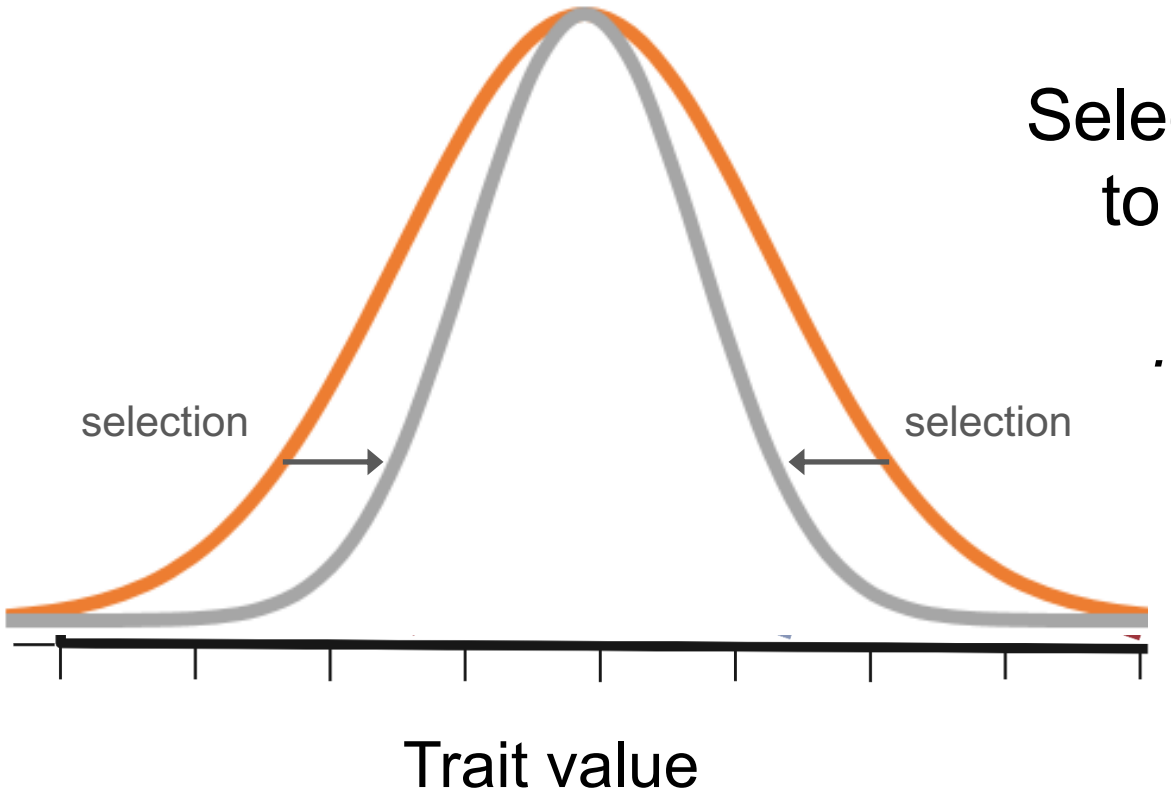
Discontinuous variation



Few variants, often with large effects

When a trait is already at the optimum

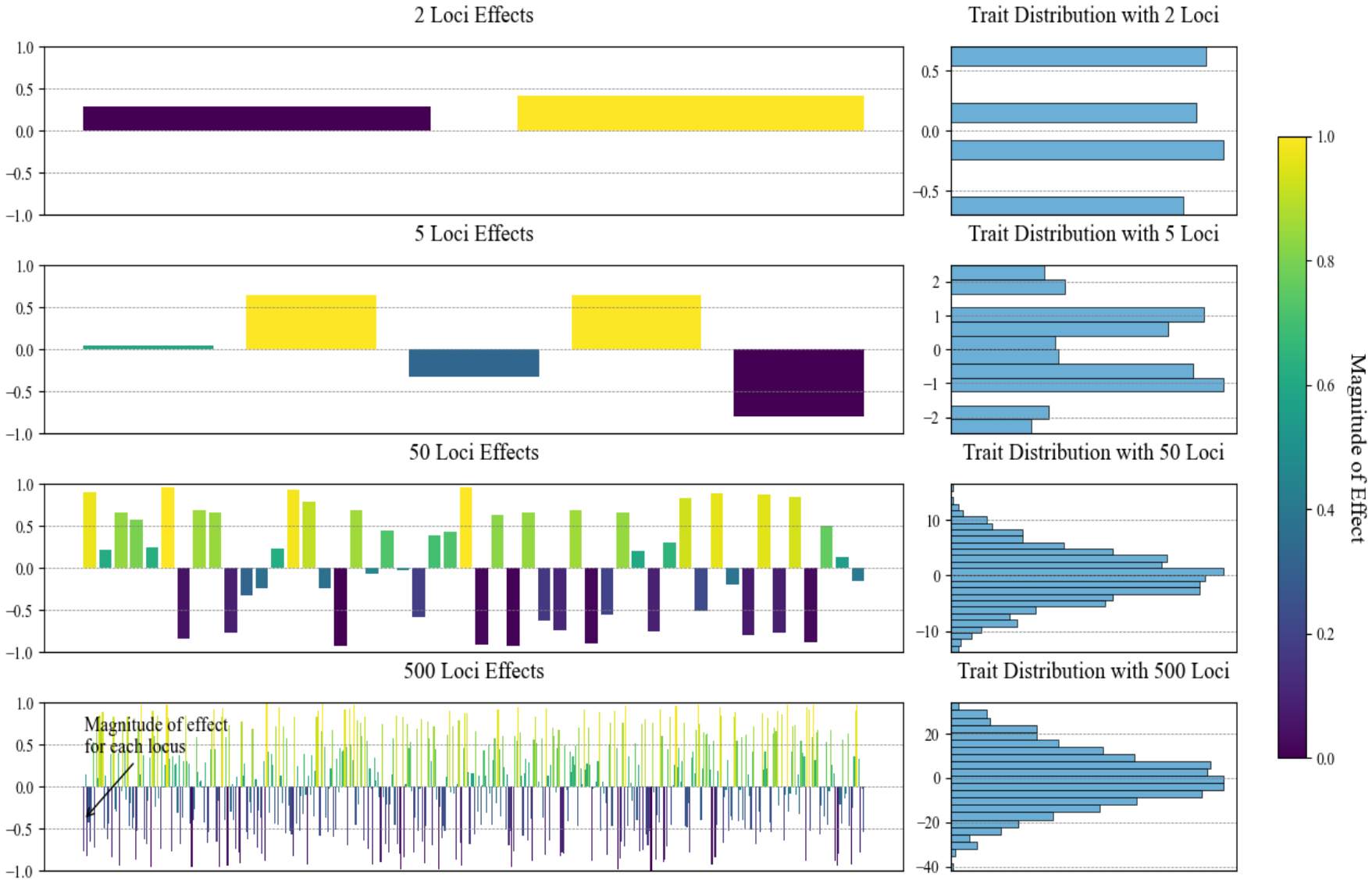
Stabilizing selection



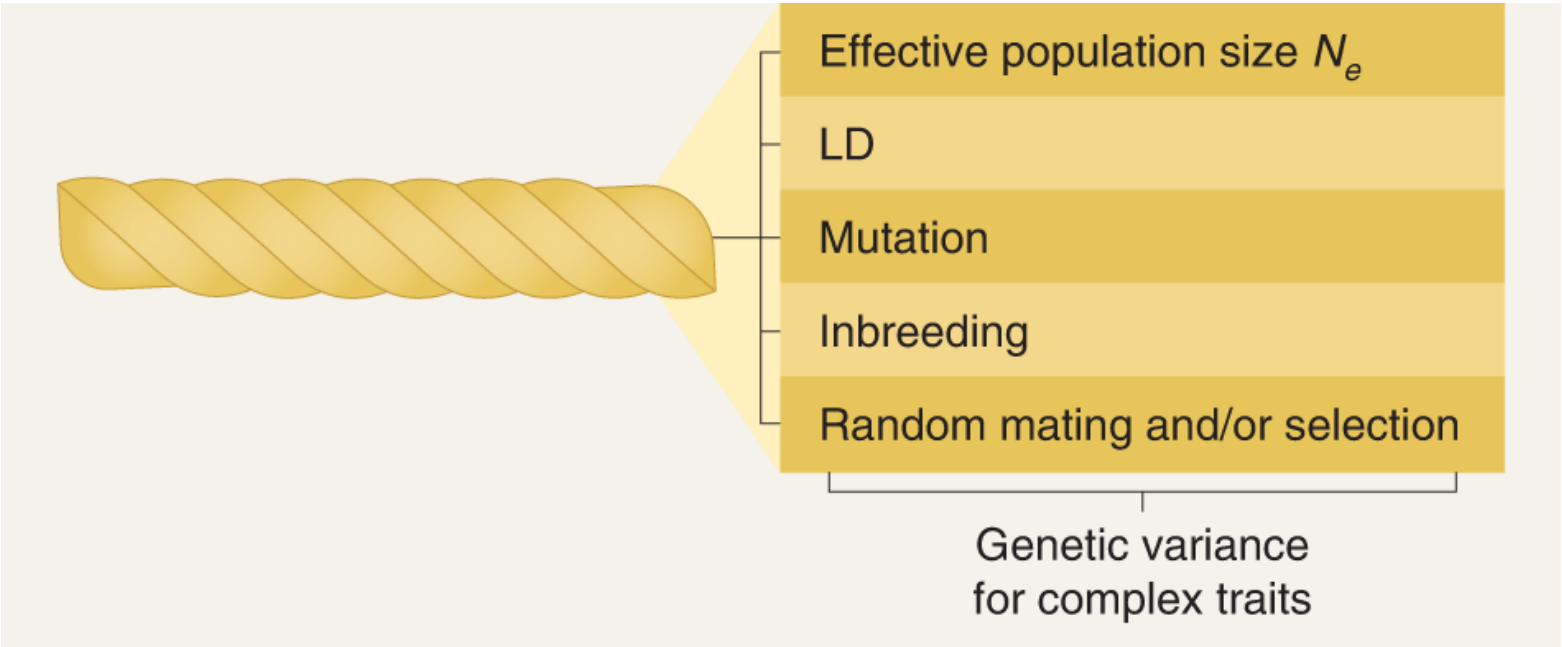
Selection against new mutations, which tend to push the trait away from the optimum

…results in an architecture where variation is due to low frequency variants with deleterious fitness effects

Fisher's infinitesimal model: quantitative variation can be due to many loci across the genome



Factors that shape genetic variance for complex traits within populations



With small N_e, inbreeding and LD increase and genetic variance decreases.
LD reduces the efficiency of selection.
Selection increases inbreeding and decreases genetic variance.
Mutation increases genetic variance.

Charlesworth, Goddard, Meyer, Visscher, Weir and Wray, 2022
https://www.nature.com/articles/s41588-022-01103-1?fromPaywallRec=false

Getting to the causal variants is hard

- Mapping using family-based designs (including crosses) and association studies only provide large-scale genomic regions
- To get to the causative variants a lot of additional work is usually necessary
- Especially if these are not due to obvious large-effect variants

Given these challenges, what can we learn from mapping approaches?

If traits are exceedingly complex, and influenced by a multitude of genetic and environmental factors, will mapping provide meaning?

Is the ‘QTN program’ a worthwhile approach?

A lot of effort has gone into identifying genetic variants that contribute to phenotypic variation

… with the idea that once we find these we can better understand the evolutionary process

But if small effect variants predominate then the gain of identifying QTN may not be worth the effort (*Rockman, 2012*)

Will a mapping approach be useful to answer *the questions we would like to ask*?

- Large effect variants are the easiest to identify from trait mapping studies, but these tend not to be responsible for most of the variation we observe in populations
- If most of the variance underlying quantitative traits is due to many variants, each with a very small effect, mapping will not allow us to identify these
- If we cannot identify the causal variants, can we ever answer the questions we would like to address?

Rockman, Evolution, 2011

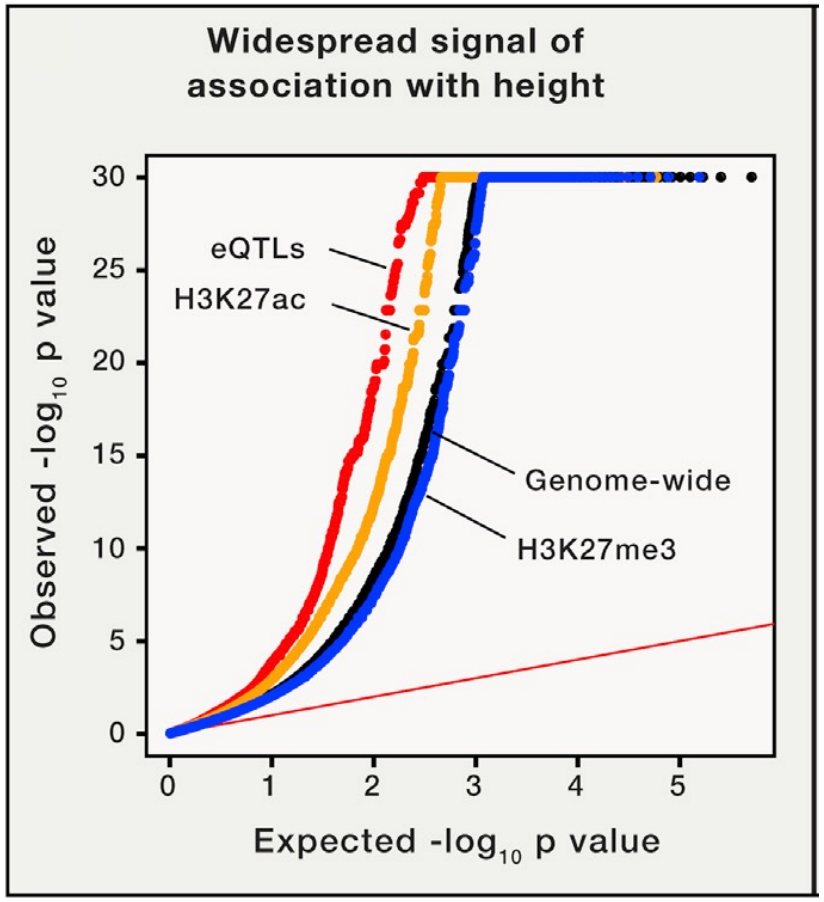
Empirical studies of adaptation: support for large effect QTNs in adaptation

There is a wealth of data suggesting that large effect mutations play important roles in adaptation, e.g.,

- Pigmentation in mice, flies, plants and humans
- Lactase persistence in humans
- Pesticide resistance in flies
- Time to first flowering in plants
- Microbial growth and resistance phenotypes

Moreover, there is a large number of cases of convergent evolution, where causative variants in the same genes arise in different populations or species

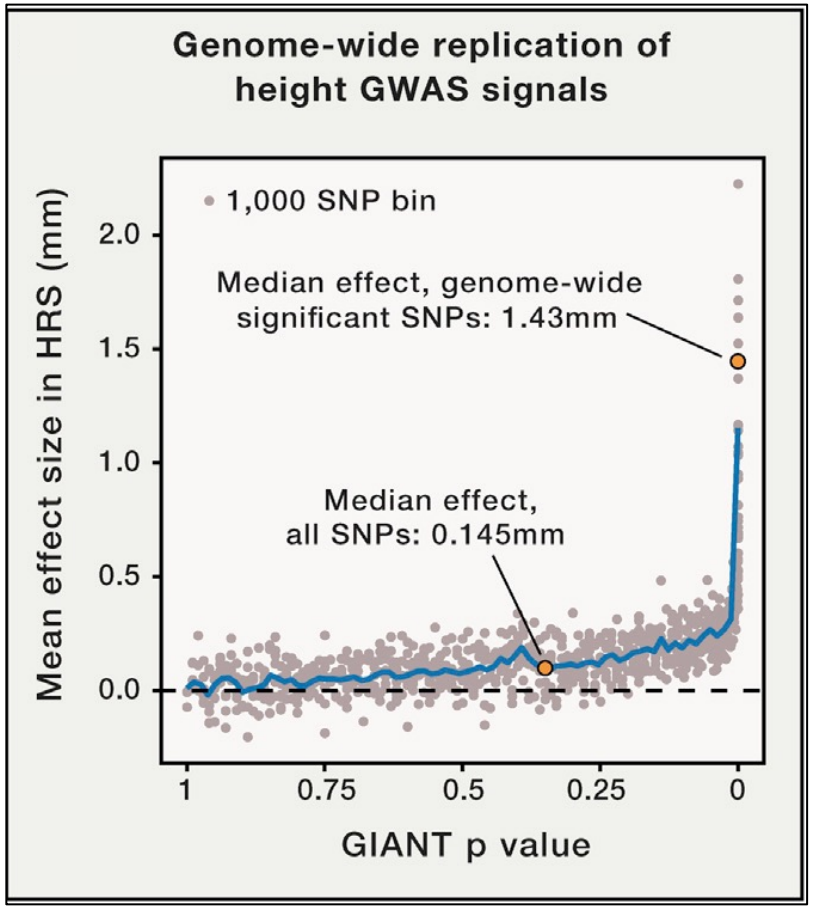
But for many traits, individual loci do not explain a large proportion of the heritability



Boyle and Prichard, 2017

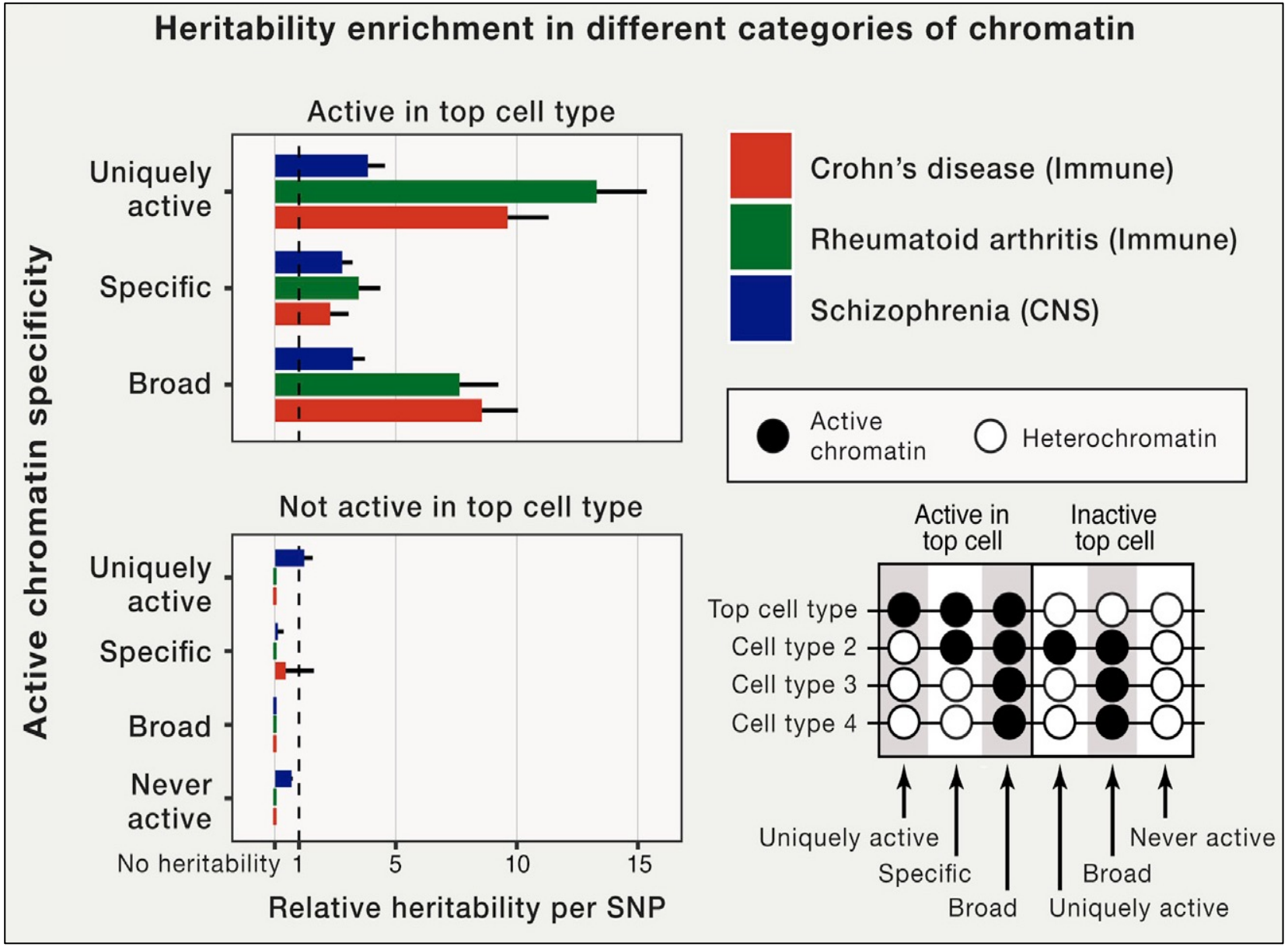
Genome-wide inflation of small p-values from the GWAS for height, with particular enrichment among expression quantitative trait loci and SNPs in active chromatin (H3K27ac)

But for many traits, individual loci do not explain a large proportion of the heritability



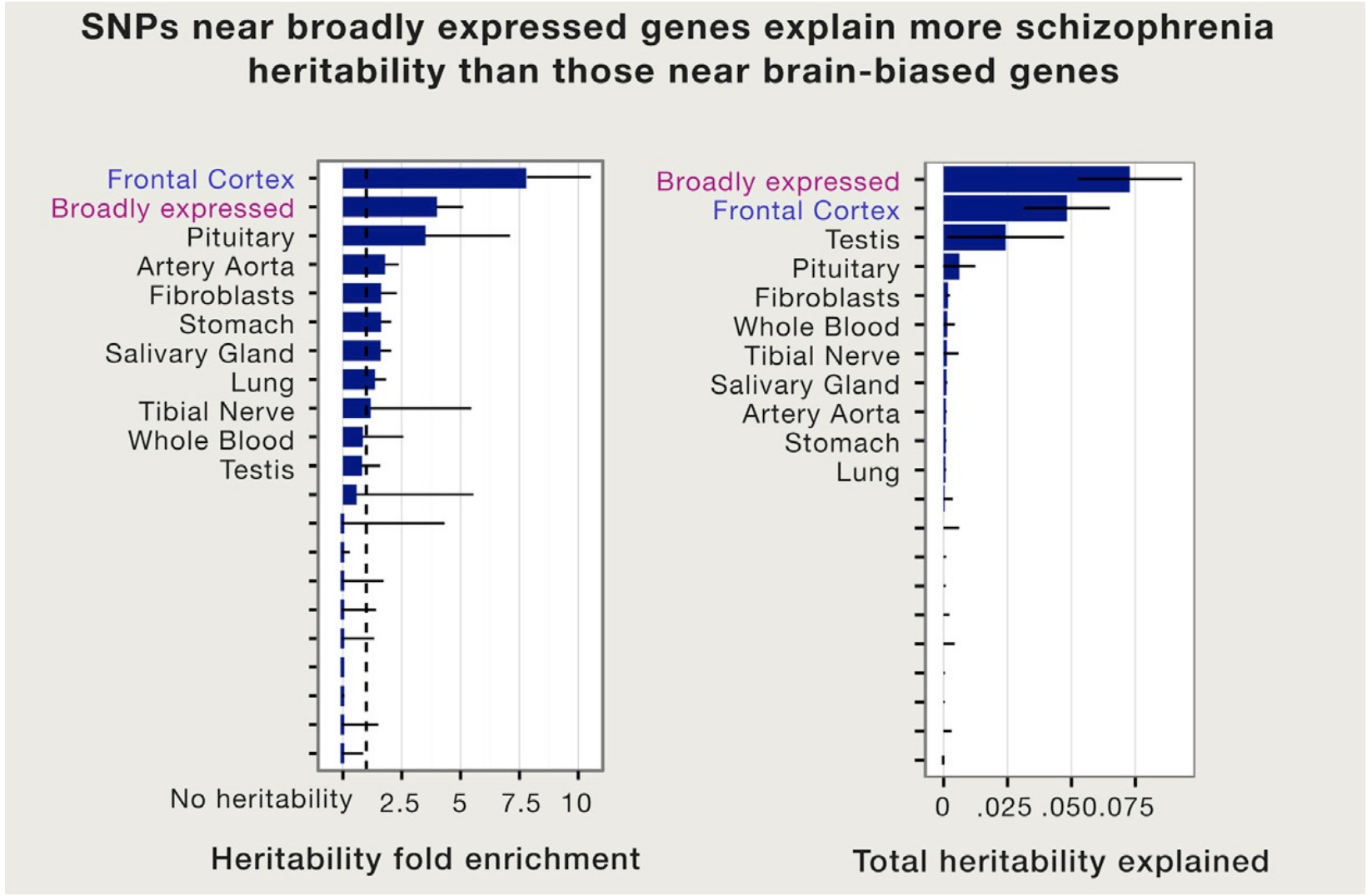
Boyle and Prichard, 2017

- Estimated mean effect size for SNPs, sorted by GIANT p-value with the direction (sign) of effect ascertained in the GIANT study, with replication effect sizes estimated using data from the Health and Retirement Study (HRS).
- The points show averages of 1,000 consecutive SNPS in the p-value-sorted list
- The effect size on the median SNP in the genome is about 10% of that for genome-wide significant hits.



Boyle and Prichard, 2017

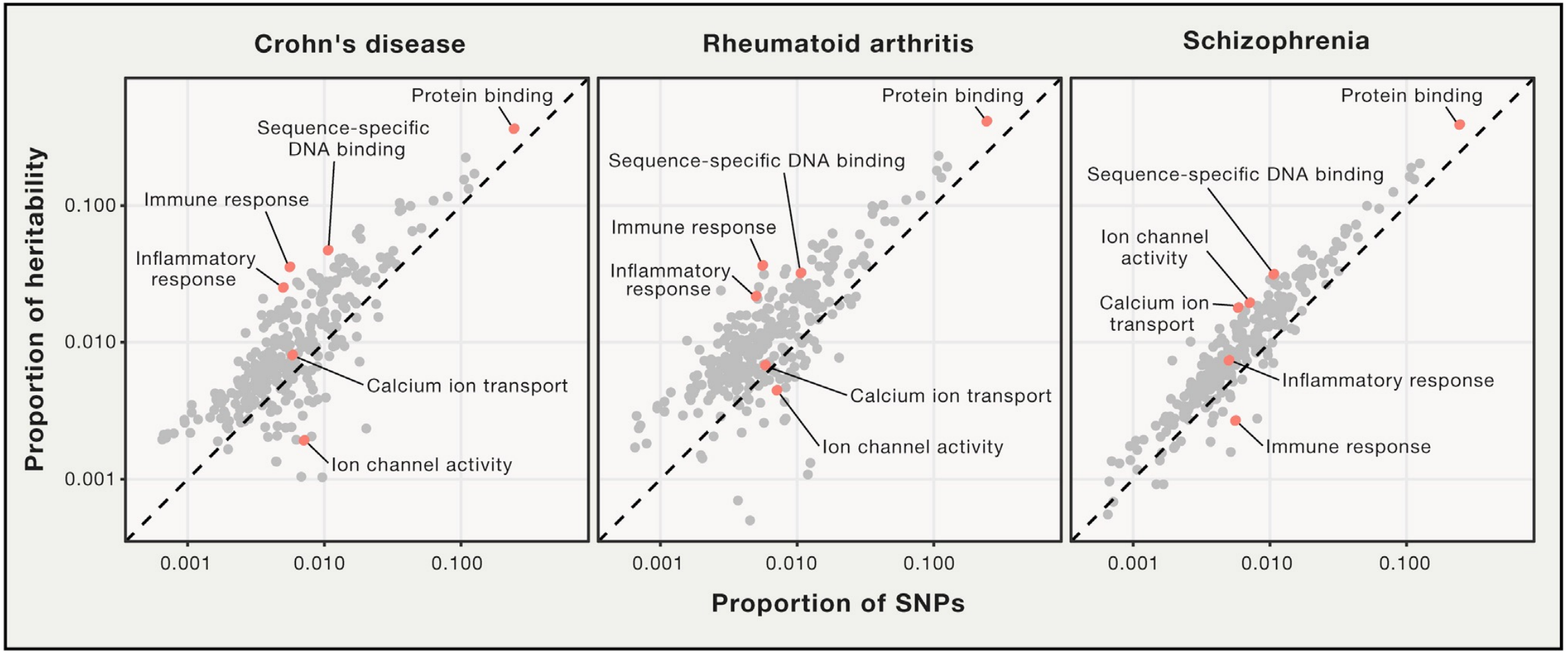
SNPs that contribute more to heritability are enriched in open chromatin, which is associated with active transcription



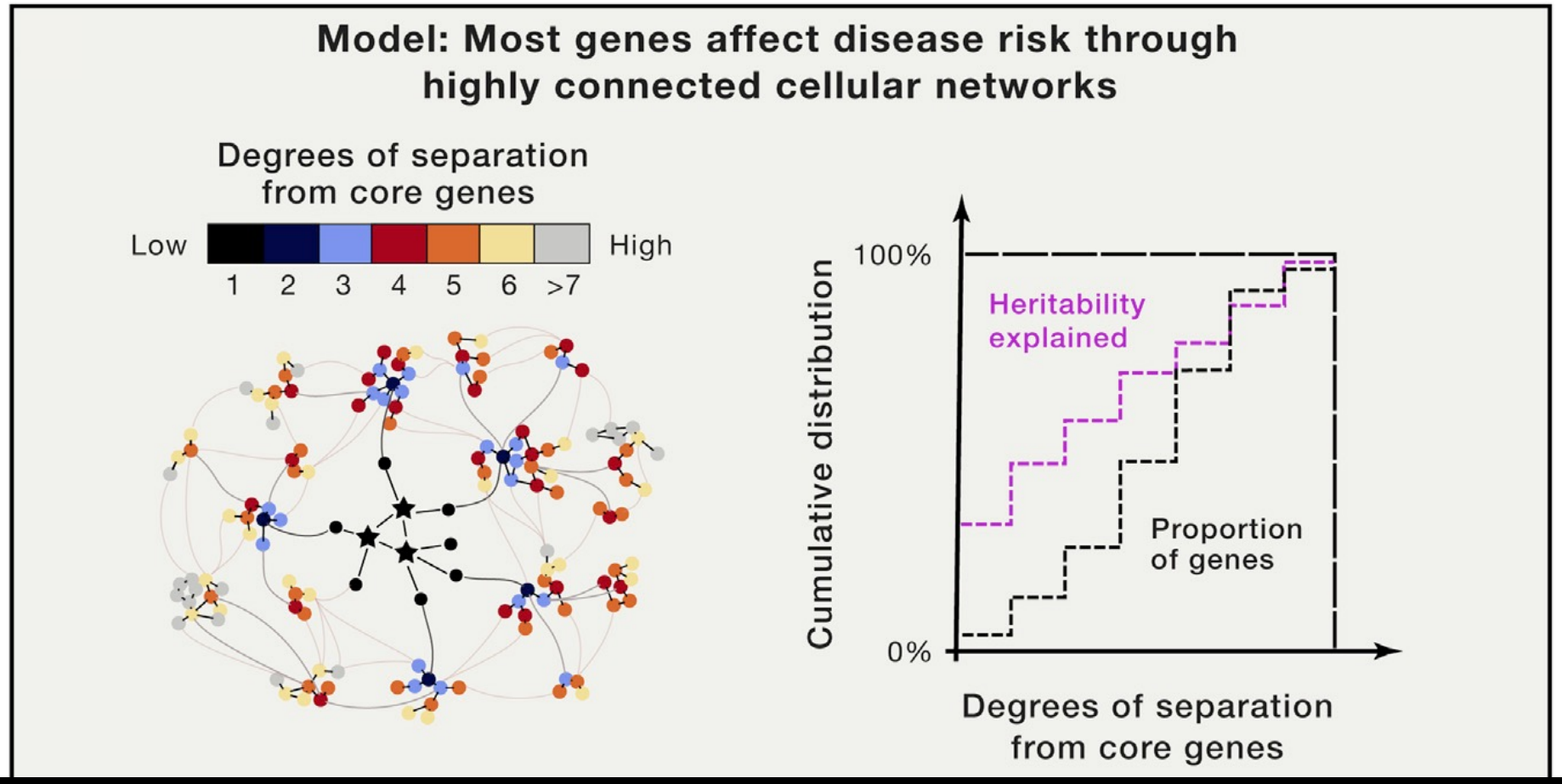
Genes with brain-specific expression show the strongest enrichment of schizophrenia signal (left), but broadly expressed genes contribute more to total heritability due to their greater number (right)

Loci in and around genes are enriched but many more loci are detected

Gene Ontology Enrichments for Three Diseases, with Categories of Particular Interest



The “omnigenic” model integrates network architecture into Fisher’s infinitesimal model



This model argues that most genes affect disease risk indirectly

Heritability for many disease traits appears to be spread across the genome

- Similar to the infinitesimal model, the omnigenic models suggests that most heritability in complex disease traits is due to genes outside core pathways
- **What other explanations might contribute to the observed patterns?**
  - Lack of power due to detect the true variants could lead to false positives or low accuracy in detection. Some contributors could be:
    - Allelic heterogeneity
    - Untyped rare or structural variants in incomplete LD with genotyped variants
    - Epistatic interactions
    - Genetic heterogeneity and across disease sub-populations could reduce power
  - Incomplete control for population structure could similarly lead to false positives and false negatives across the genome – recent evidence shows that this plays an important role in generating the pattern

Heritability for many disease traits appears to be spread across the genome

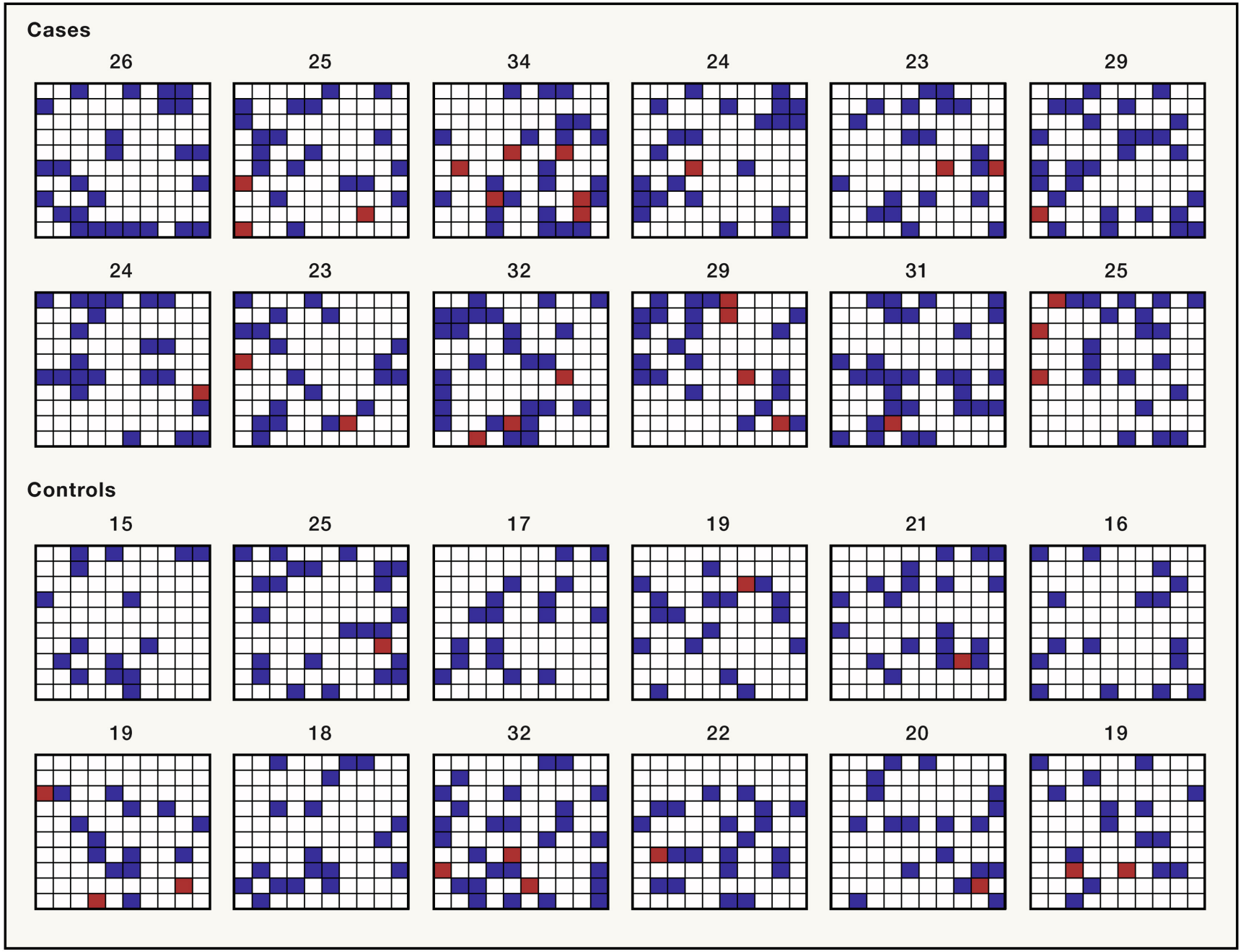
- Similar to the infinitesimal model, the omnigenic models suggests that most heritability in complex disease traits is due to genes outside core pathways
- **What other explanations might contribute to the observed patterns?**
  - Lack of power due to detect the true variants could lead to false positives or low accuracy in detection. Some contributors could be:
    - Allelic heterogeneity
    - Untyped rare or structural variants in incomplete LD with genotyped variants
    - Epistatic interactions
    - **Genetic heterogeneity and across disease sub-populations could reduce power**
  - Incomplete control for population structure could similarly lead to false positives and false negatives across the genome – recent evidence shows that this plays an important role in generating the pattern

A polygenic model results in genetic heterogeneity

Each box in the figure on the right represents 100 risk loci actoss sampled patients and controls. Blue represent heterozygotes for the risk allele and red homozygotes. Risk alleles are present in both cases and controls but tend to be more common in cases

Wray et al., argue that:

1. It will be important to improve patient stratification and disease sub-type analysis
2. There is a need to develop cell-based model system that can recapitulate aspects of complex traits



Wray et al., 2018

Heritability for many disease traits appears to be spread across the genome

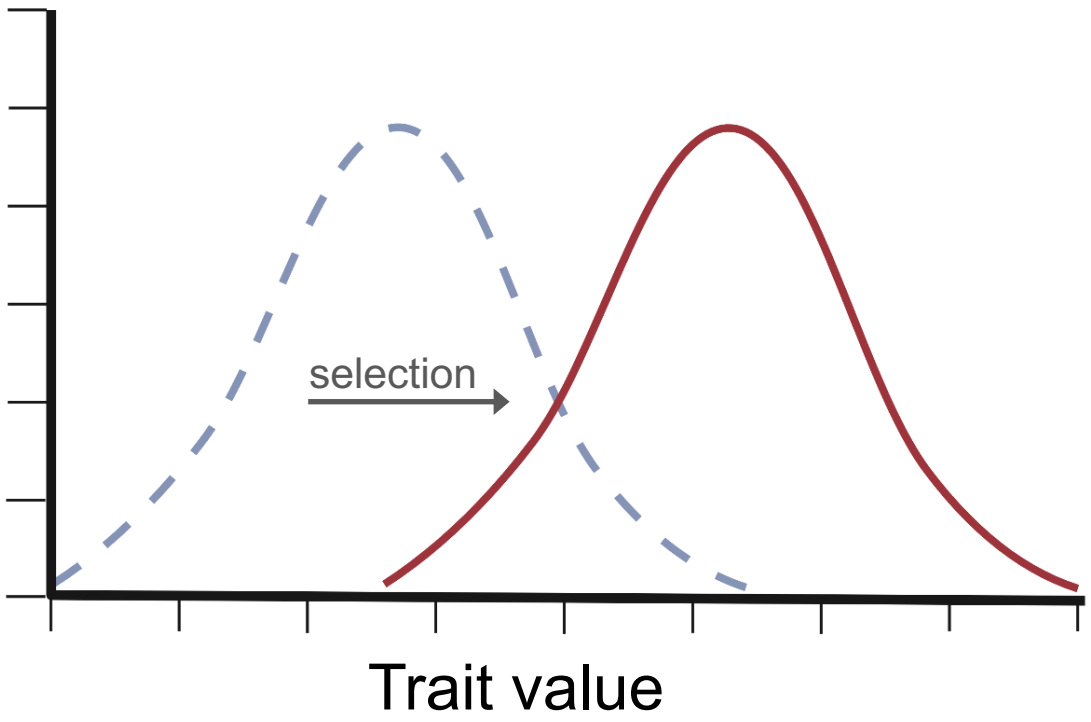
- Similar to the infinitesimal model, the omnigenic models suggests that most heritability in complex disease traits is due to genes outside core pathways
- **What other explanations might contribute to the observed patterns?**
  - Lack of power due to detect the true variants could lead to false positives or low accuracy in detection. Some contributors could be:
    - Allelic heterogeneity
    - Untyped rare or structural variants in incomplete LD with genotyped variants
    - Genetic heterogeneity of partially linked variants
    - Epistatic interactions
    - Heterogeneity across disease sub-populations could reduce power
  - Incomplete control for population structure could similarly lead to false positives and false negatives across the genome – recent evidence shows that this plays an important role in generating the pattern

Can we sufficiently control for all of the potential confounders to assess the impact of individual variants on traits?

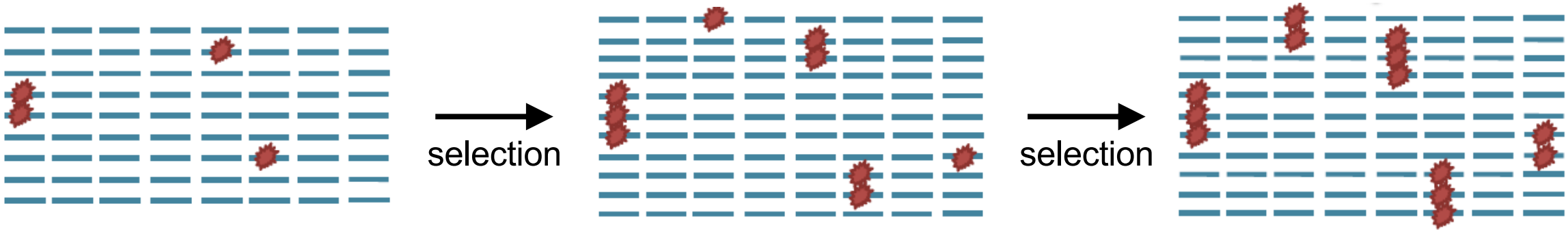
Positive selection

Selection to a shifting optimum

Directional selection



Modes of adaptation: polygenic adaptation

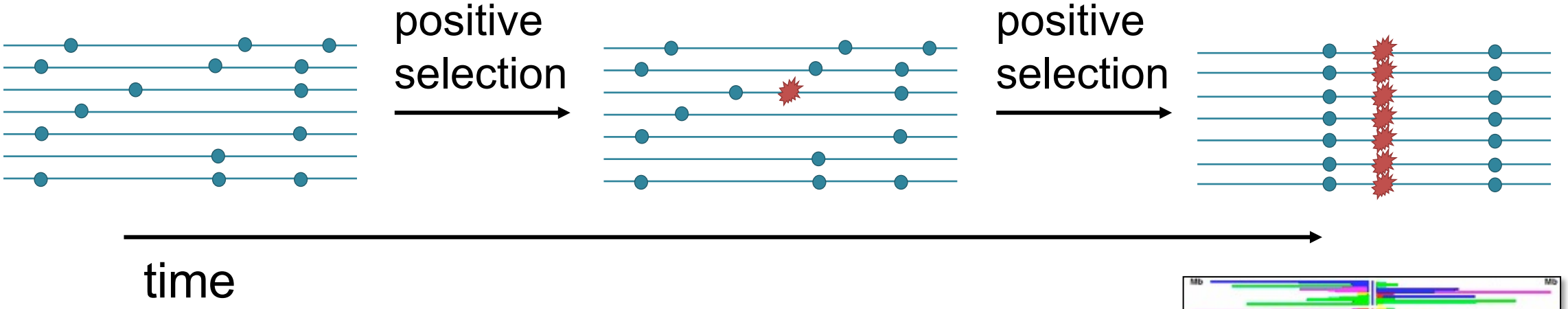


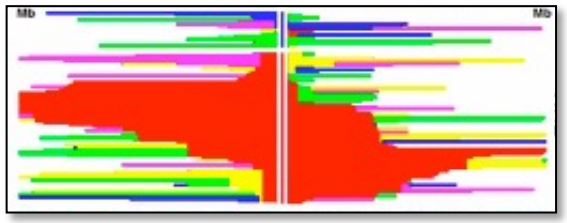
Selection results in subtle shifts in frequencies at many loci, most or all of which were already present in the population

graphic modified from Flood and Hancock, 2017

Modes of adaptation: hard sweep

Hard sweep mode of adaptation:

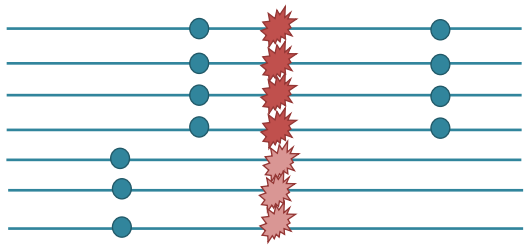




Maynard Smith and Haigh, 1974
graphic from Flood and Hancock, COPB, 2017

Modes of adaptation: soft sweep

Multiple common haplotypes



Soft sweep patterns have heterogeneous origins:

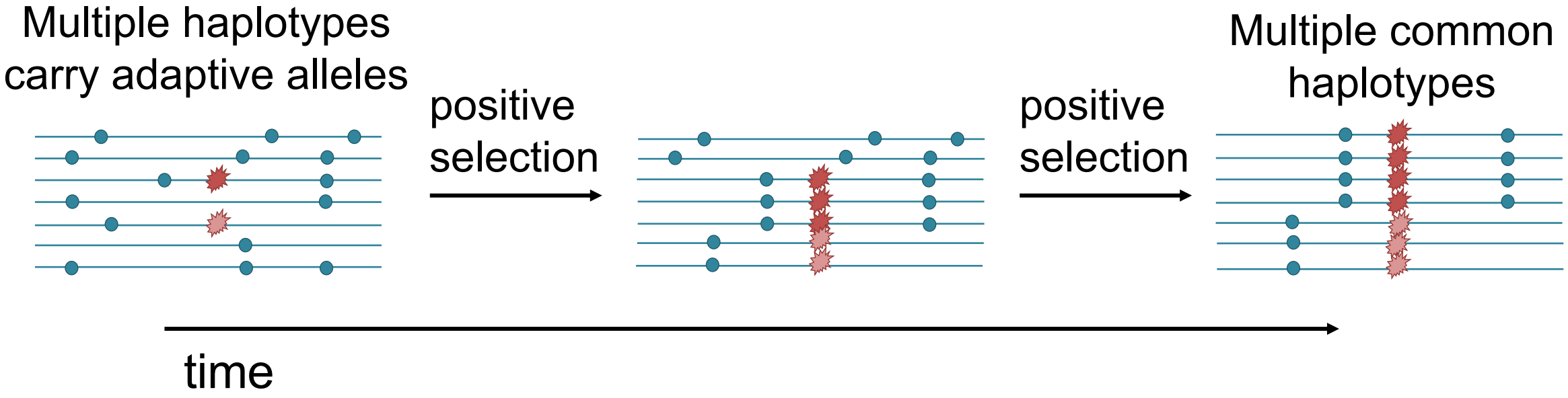
1. Selection from standing variation

2. Selection from recurrent (new) mutations

Hermisson and Pennings 2017
Harris et al., 2018
Paulose et al., 2019

Modes of adaptation: soft sweep

Soft sweep mode of adaptation:



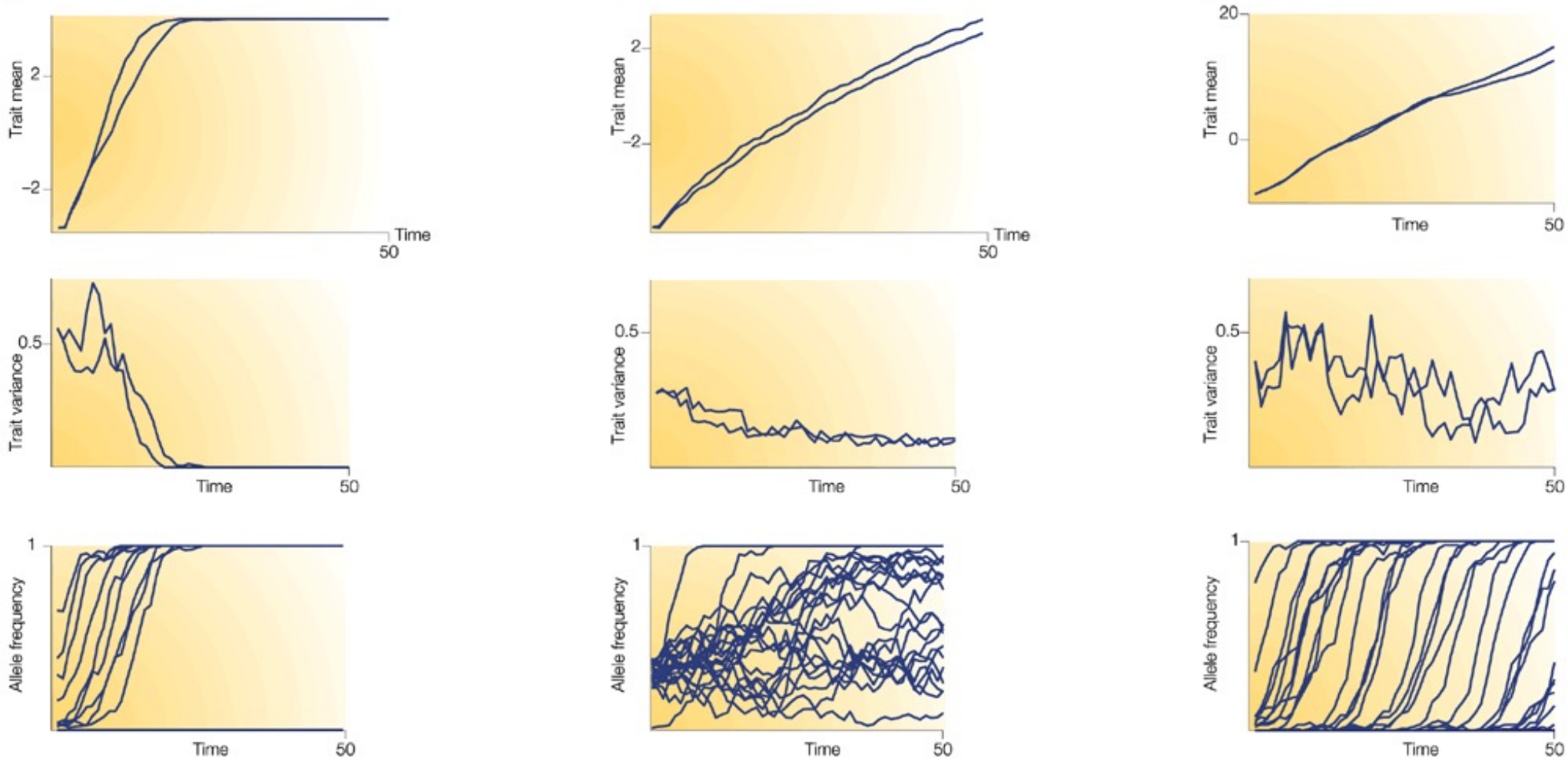
Orr and Bettancourt, 2001
Hermisson and Pennings, 2005
Pennings and Hermisson, 2006
Messer and Petrov, 2013
graphic from Flood and Hancock, COPB, 2017

The number of loci that contribute to a trait impact the expected mode of adaptation

10 unlinked loci with major alleles

400 minor effect loci + 2 major effect loci

100 minor effect loci



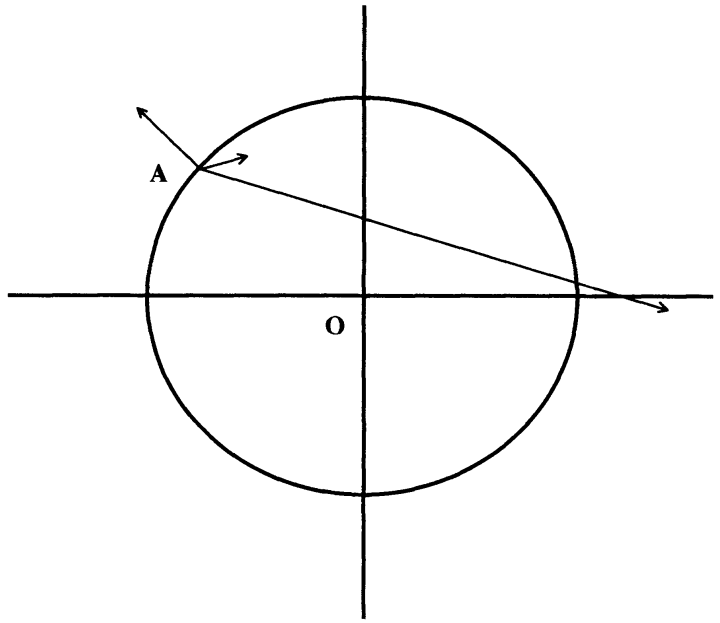
Barton and Keitley, NRG, 2007

Fisher proposed a geometric model for adaptive trait variation

Fisher's geometric model assumes:

- Many loci, each with a small effect, underlie quantitative trait variation
- Adaptation to a novel environment involves multiple traits
- After a shift to the new environment, many traits are far from their adaptive optimum
- Most variants that arise will move the population away from the optimum, but those with positive effects on fitness will accumulate over time
- Adaptation occurs by a step-wise process towards the new optimum

Fisher's geometric model of adaptation



Orr, 1998

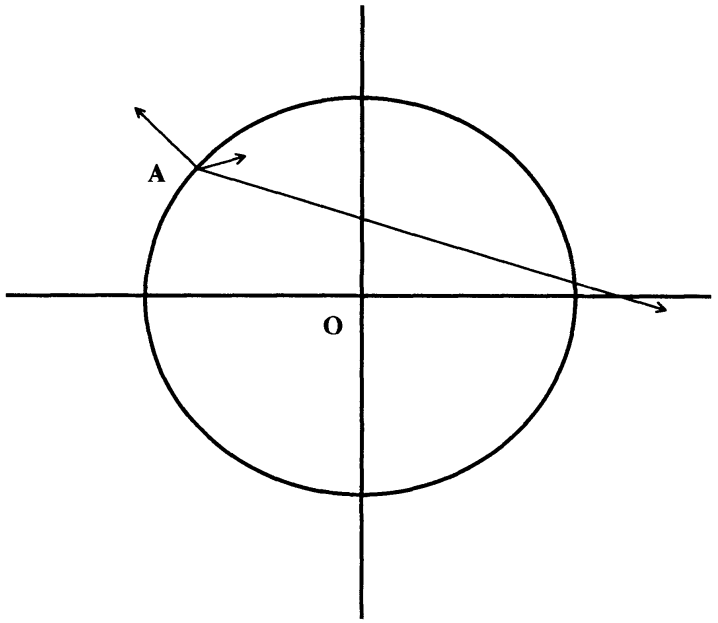
The statistical requirements of the situation, in which one thing is made to conform to another in a large number of different respects, may be illustrated geometrically. The degree of conformity may be represented by the closeness with which a point A approaches a fixed point O. In space of three dimensions we can only represent conformity in three different respects, but even with only these the general character of the situation may be represented. The possible positions representing adaptations superior to that represented by A will be enclosed by a sphere passing through A and centred at O. If A is shifted through a fixed distance, r, in any direction its translation will improve the adaptation if it is carried to a point within this sphere, but will impair it if the new position is outside. If r is very small it may be perceived that the chances of these two events are approximately equal, and the chance of an improvement tends to the limit $\frac{1}{2}$ as r tends to zero; but if r is as great as the diameter of the sphere or greater, there is no longer any chance whatever of improvement, for all points within the sphere are less than this distance from A. For any value of r between these limits the actual probability of improvement is

$$\frac{1}{2}\left(1-\frac{r}{d}\right),$$

where d is the diameter of the sphere.

Fisher, 1930

Fisher's geometric model of adaptation



Orr, 1998

The chance of improvement thus decreases steadily from its limiting value $\frac{1}{2}$ when r is zero, to zero when r equals d. Since A in our representation may signify either the organism or its environment, we should conclude that a change on either side has, when this change is extremely minute, an almost equal chance of effecting improvement or the reverse; while for greater changes the chance of improvement diminishes progressively, becoming zero, or at least negligible, for changes of a sufficiently pronounced character.

Fisher, 1930

Fisher's geometric model of adaptation (1930)

Effect size vs. probability of fixation

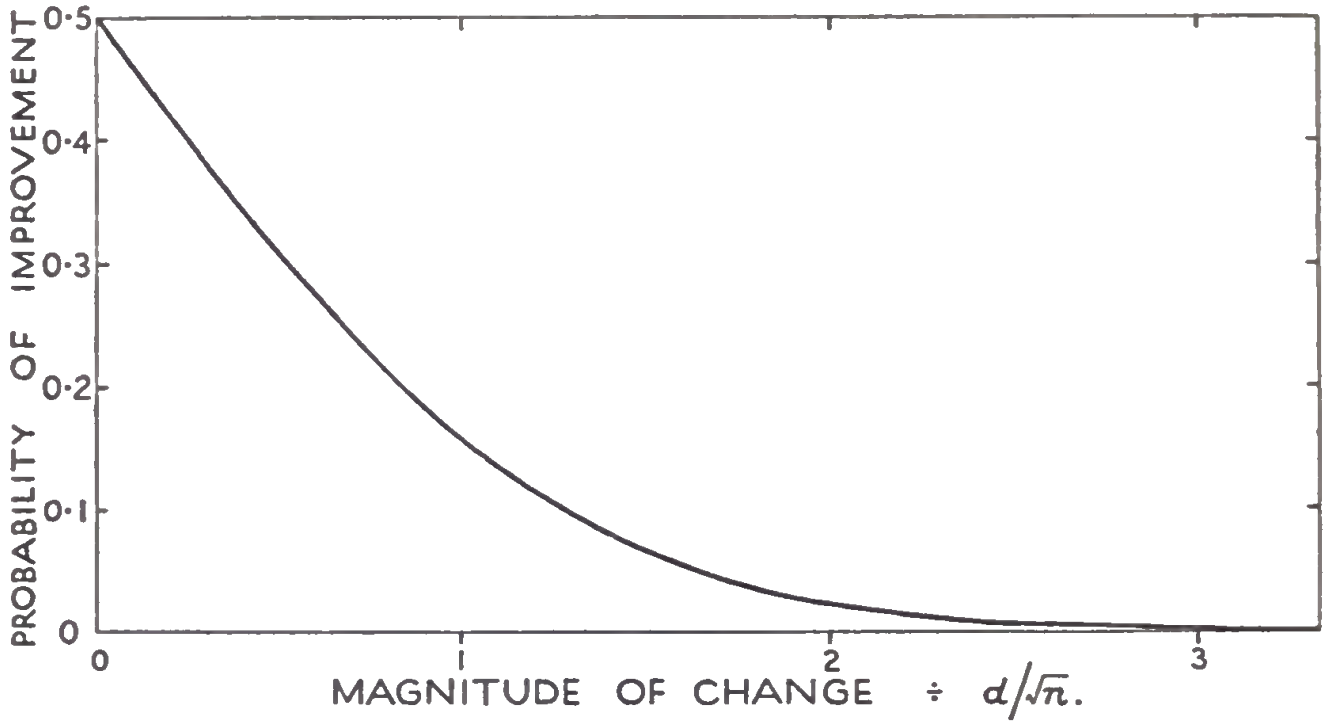


FIG. 3. The relation between the magnitude of an undirected change and the probability of improving adaptation, where the number of dimensions (n) is large

- In Fisher's model, the probability a mutation would improve a trait decreased as a function of the magnitude of its effect
- Fisher assumed that very small effect mutations had a 50% probability of being beneficial whereas larger effect mutations were more likely deleterious
- This led to the conclusion that adaptation progresses by small (infinitesimal) effect mutations

Fisher, 1930

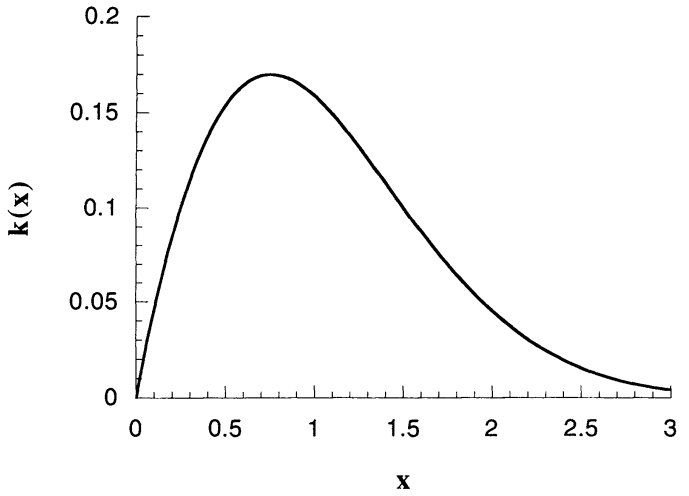
Fisher's geometric model of adaptation

The representation in three dimensions is evidently inadequate; for even a single organ, in cases in which we know enough to appreciate the relation between structure and function, as is, broadly speaking, the case with the eye in vertebrates, often shows this conformity in many more than three respects. It is of interest therefore, that if in our geometrical problem the number of dimensions be increased, the form of the relationship between the magnitude of the change r and the probability of improvement, tends to a limit which is represented in Fig. 3. The primary facts of the three dimensional problem are conserved in that the chance of improvement, for very small displacements tends to the limiting value $\frac{1}{2}$, while it falls off rapidly for increasing displacements, attaining exceedingly small values, however, when the number of dimensions is large, even while r is still small compared to d.

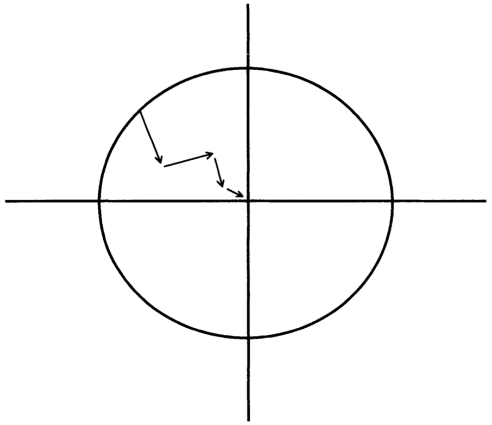
Fisher, 1930

Orr's version of the FGM (Fisher-Orr model)

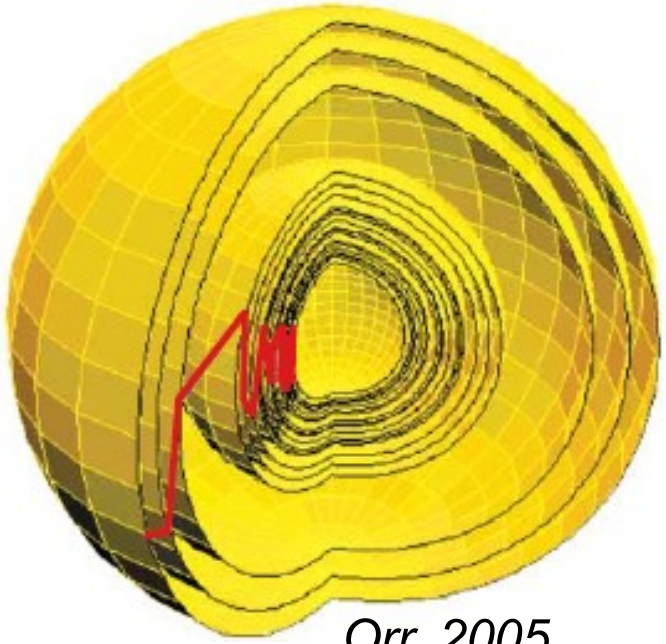
- Orr produced a revised version of FGM that assumed a different distribution of mutational effects
- Orr thought Kimura's distribution (right) was more realistic



- In Orr's resulting model, the distribution of fitness effects after an adaptive walk is exponential, with larger effects occuring earlier in the walk and smaller effects occuring later
- A schematic of a two-dimensional walk is shown on the right



Strong directional selection to a divergent optimum: an adaptive walk



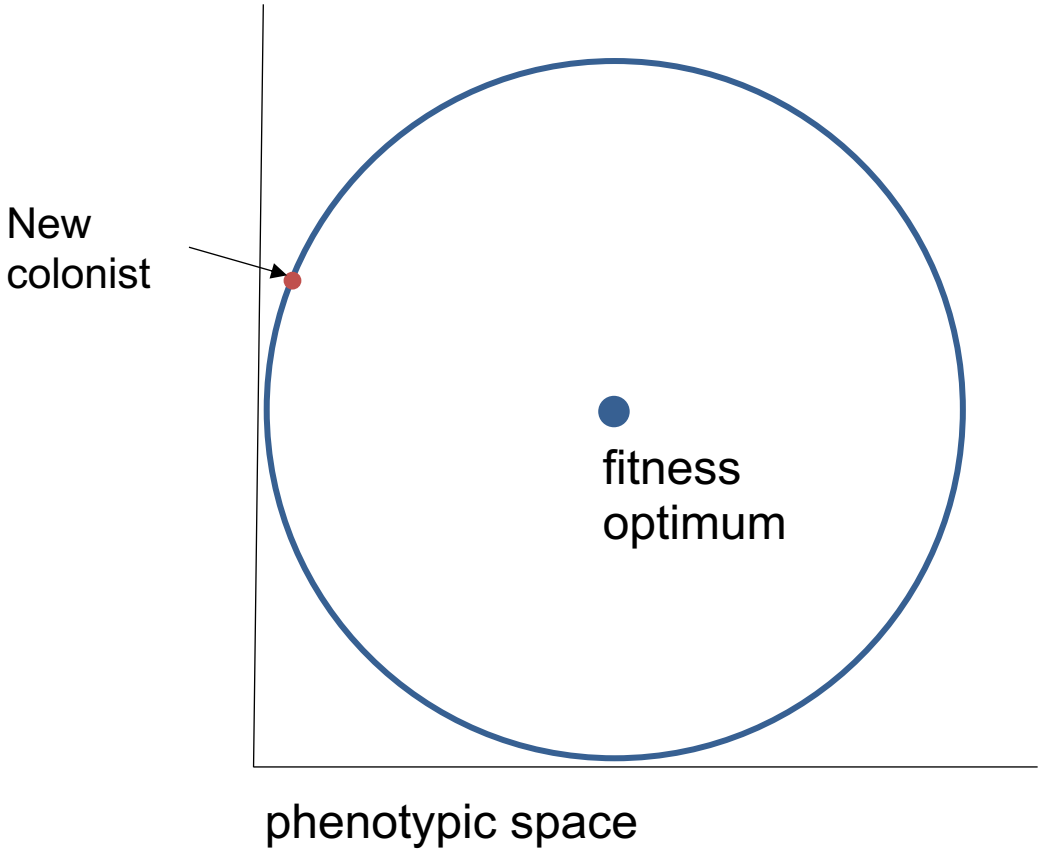
Orr, 2005

Fisher-Orr Geometric Model

- Sudden multivariate shift in the selective regime
- New fitness optimum is far from the previous optimum
- New mutations underlie adaptation
- Effect size distribution is exponential

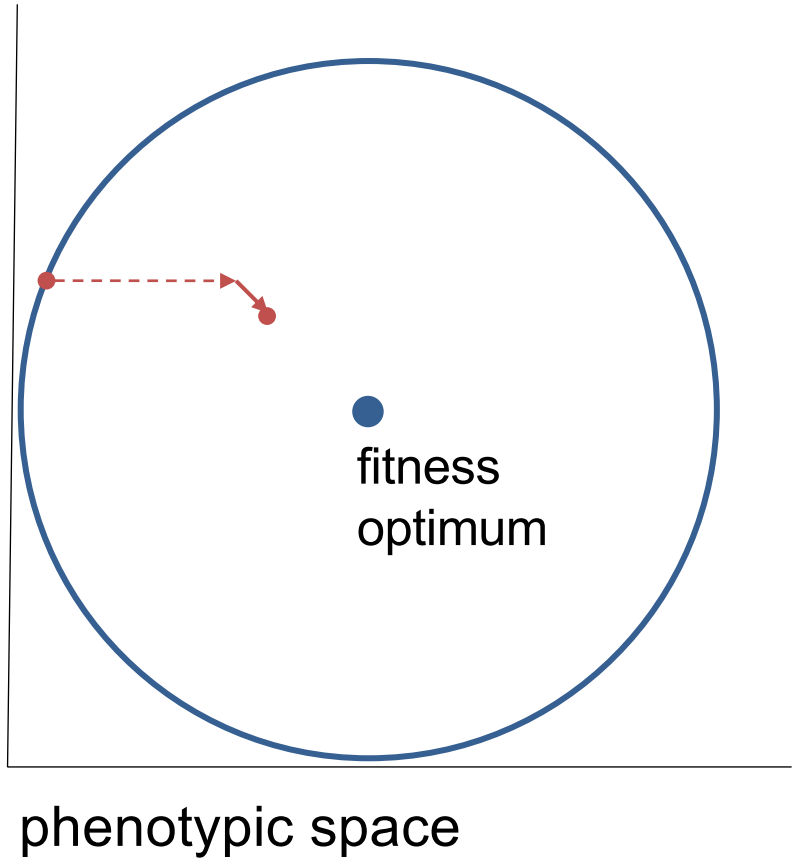
Fisher 1930
Orr and Coyne, 1992
Orr 1998
Orr 2005

Adaptive walk: populations reach their fitness optima through a series of genetic steps



Adaptive walk: populations reach their fitness optima through a series of genetic steps

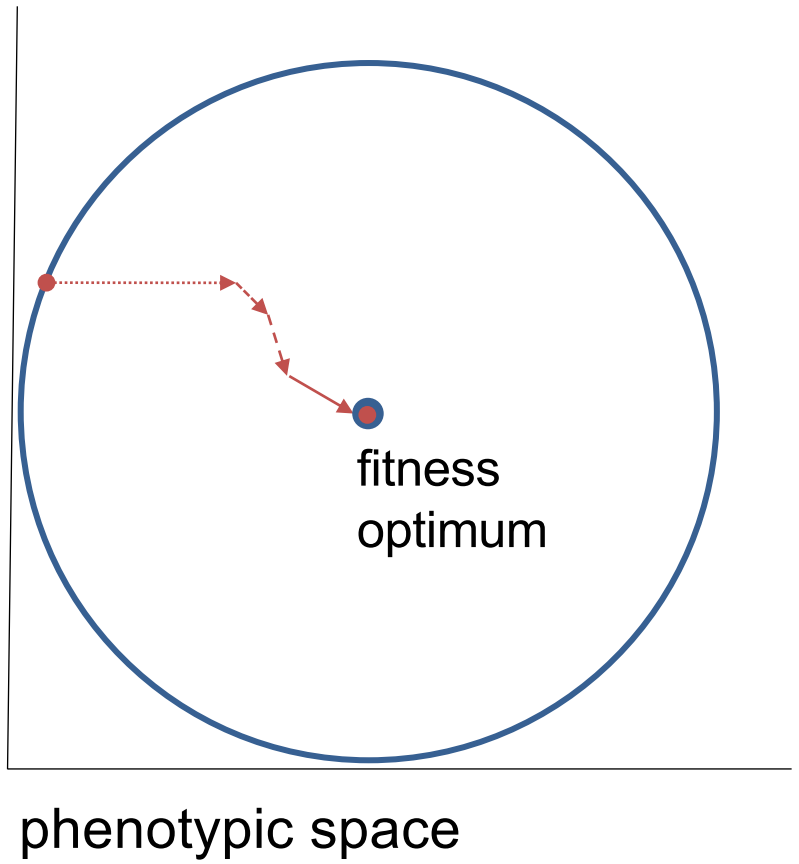
Over time, the population adapts through an accumulation of genetic changes



Adaptive walk: populations reach their fitness optima through a series of genetic steps

Over time, the population adapts through an accumulation of genetic changes

…that are positively selected and move it toward the fitness optimum



Population size and the architecture of adaptation

- Gillespie described potential adaptive regimes, with finite populations versus approximately infinite (very large) population size and mutation availability
- The extremes are the strong selection weak mutation (SSWM) regime and the weak selection strong mutation (WSSM) regime.
- In the SSWM, available beneficial mutations are limited in the population
- In the WSSM regime, the potential beneficial mutations are not limiting
- Whether a population falls into the SSWM or WSSM regime for a particular trait is a function of the population size and the beneficial mutation rate, which depends on the number of potential loci and types of mutations that can contribute to a trait

Population size and the architecture of adaptation

The strong selection weak mutation (SSWM) regime is expected to hold when the total number of new mutations that enter a diploid population each generation is small, i.e., $4NU_b \ll 1$ and when selection is strong, i.e., $4Ns >> 1$,

where N is the population size and U_b is the genome-wide beneficial mutation rate for the focal trait

U_b depends on u, the per base mutation rate and the mutational target size, i.e., the number of nucleotides that affect the trait

Such a population is mutation-limited; i.e., only a subset of possible beneficial mutations are available for adaptation

In this scenario, when single new beneficial mutations overcome genetic drift they are likely to create a hard-sweep architecture of adaptation

Population size and the architecture of adaptation

Alternatively, in the weak selection, strong mutation (WSSM) regime, $4NU_b \gg 1$

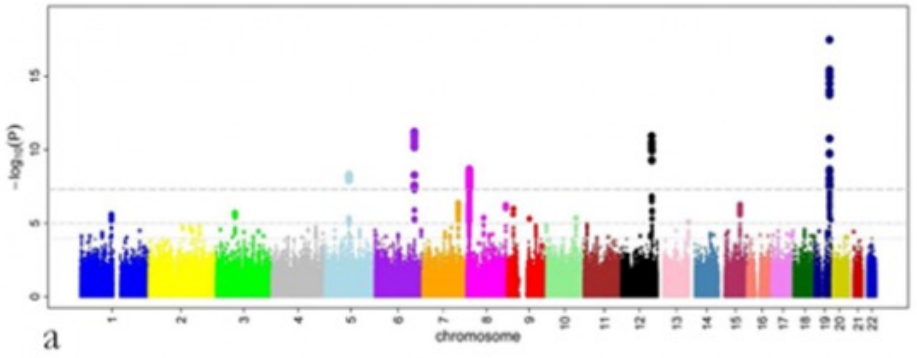
In this regime, the population is expected to have a large mutational supply. In other words, the population is not mutation-limited, and adaptation may occur largely through small frequency shifts at a large number of alleles, each with small individual effects.

The key result from the SSWM vs WSSM comparison is that $4NU_b = 1$ is the expected threshold that separates the sweep-like from highly polygenic architectures

The role of epistatic interactions (GxG) in an adaptive walk

Top-down: mapping components of fitness

- Can use trait mapping approaches to identify variants and GxG effects associated with fitness and specific components of fitness
- These tend to be underpowered and LD interferes with localization over short scales



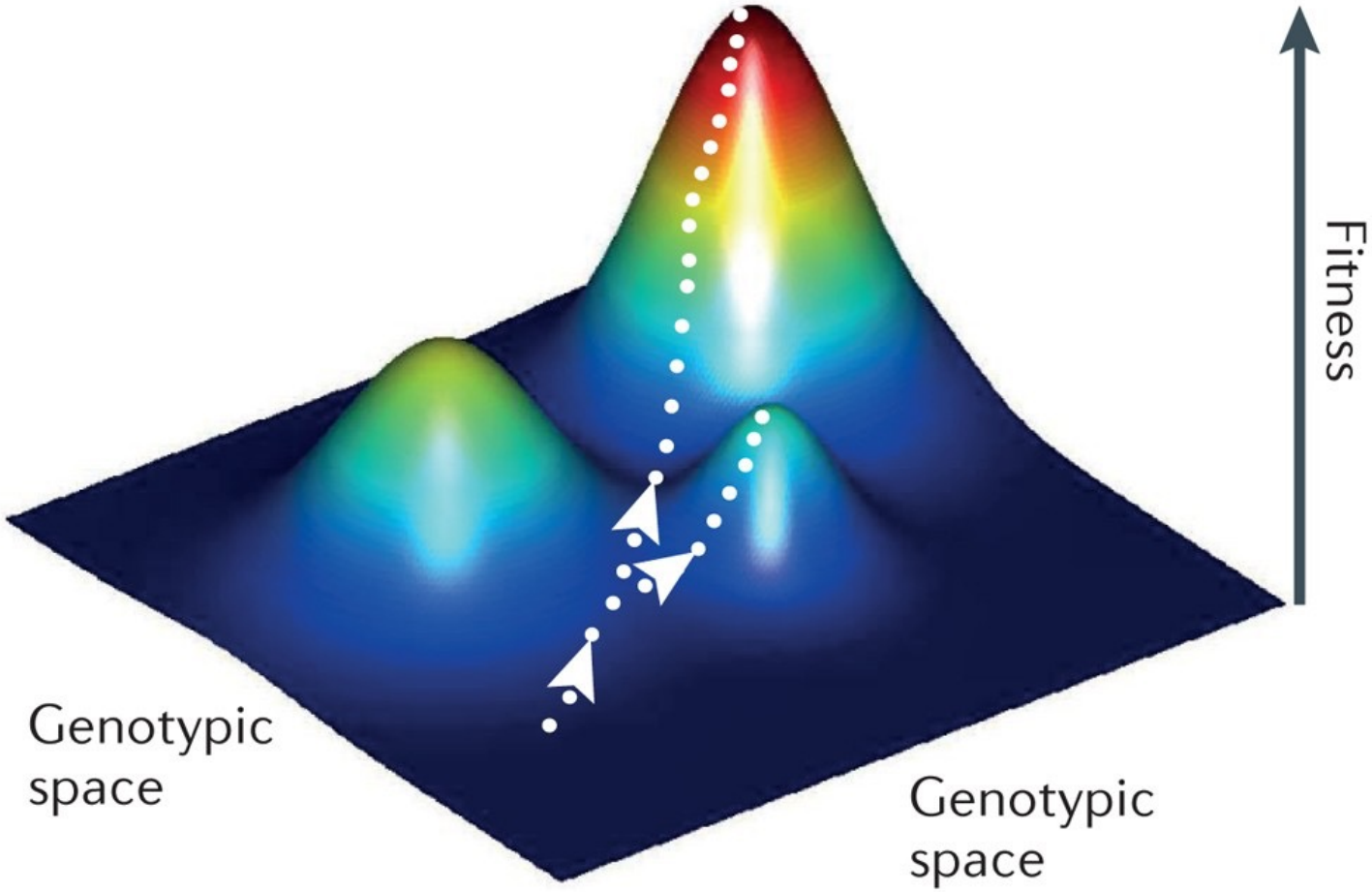
Bottom-up: make all possible combinations of mutations to understand fitness landscape

- Create mutants with all possible combinations of a set of nucleotide changes

Brief history of adaptive landscape models

- In contrast to Fisher, Sewell Wright thought that epistasis was important to evolutionary trajectories and pervasive
- Wright represented the landscape of potential fitness much like map-makers represent geographical topography
- Wright realized the genotypic space would be immense
- His models depict rugged fitness landscapes with high peaks separated by valleys of low-fitness genotypes
- In recent years, mutant analysis has allowed systematic analysis of the topology of adaptive landscapes

A multi-dimensional fitness landscape

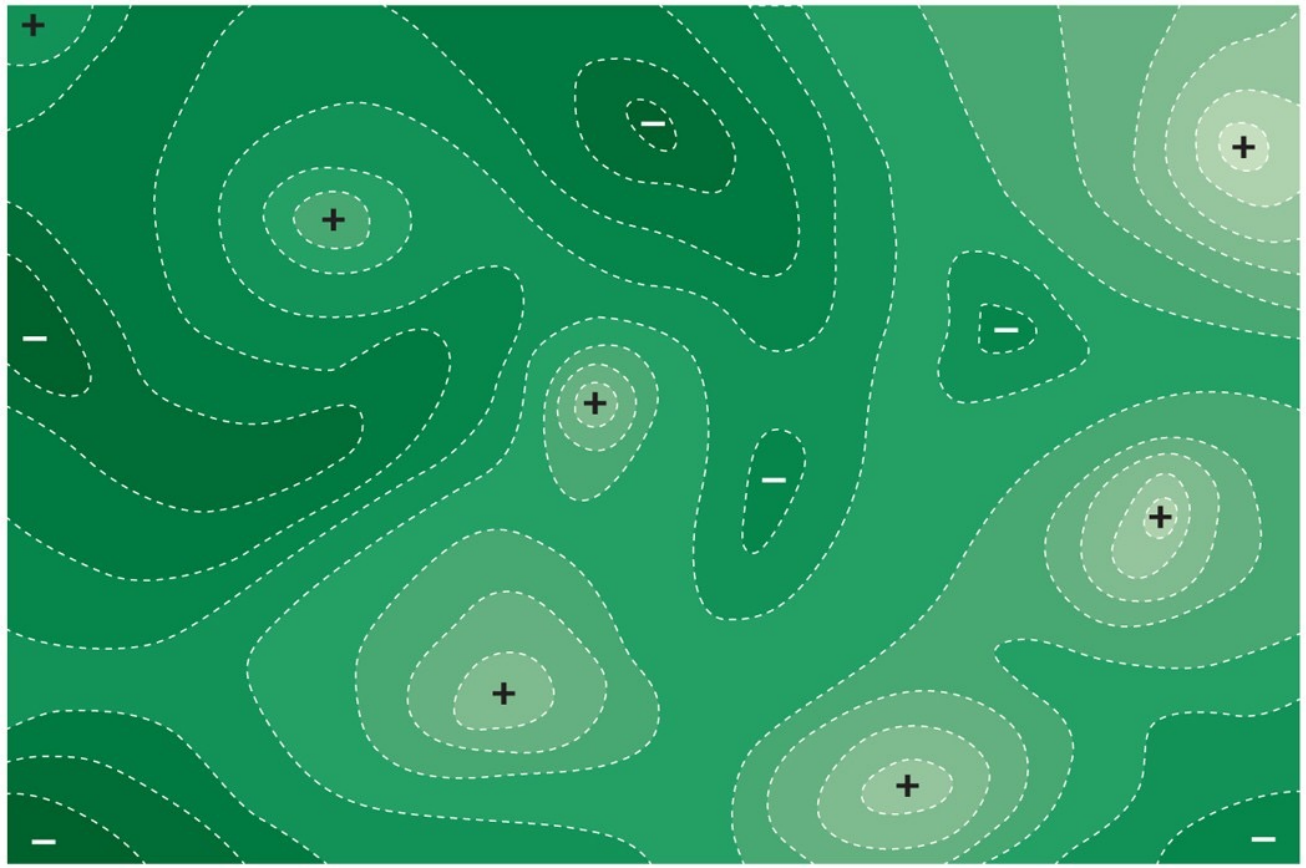


A rugged fitness landscape with peaks and valleys

De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Multi-dimensional fitness landscapes

Wright's two-dimensional "field of gene combinations"

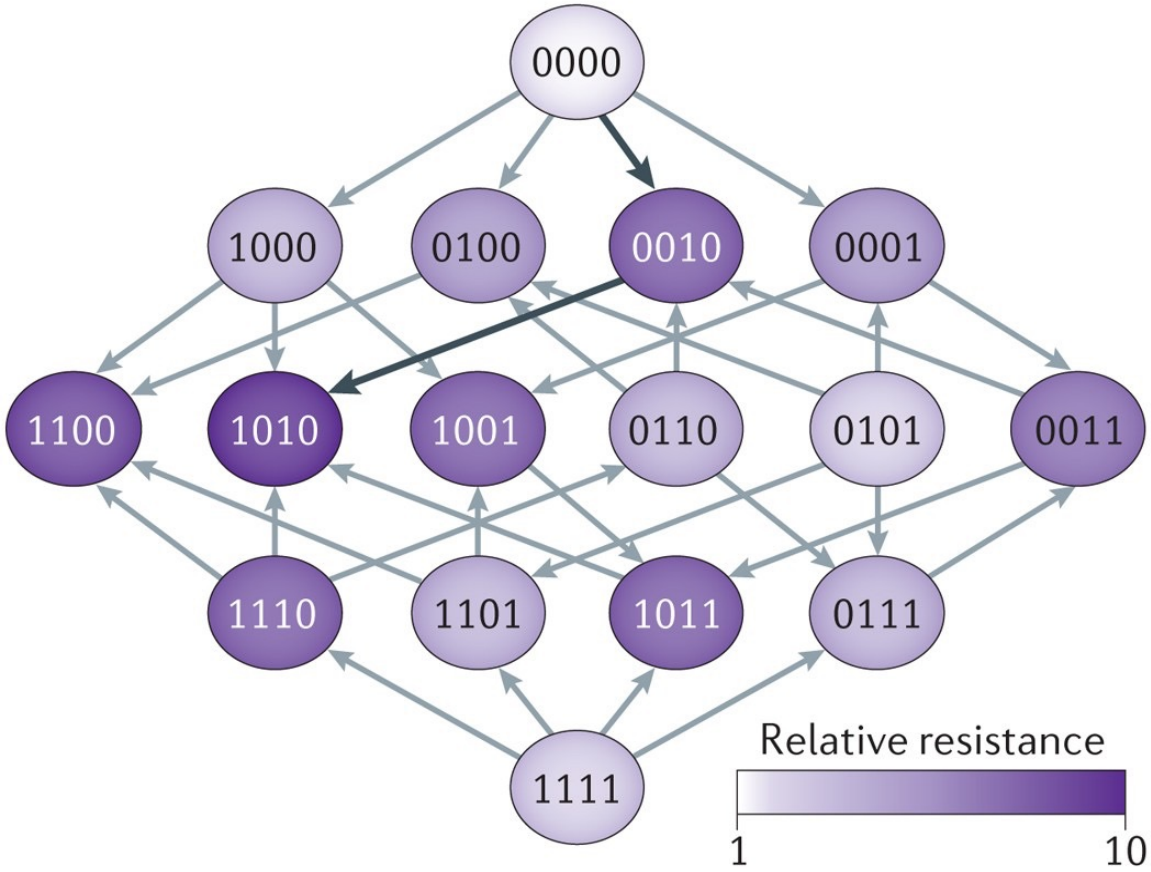


- Wright (1931, 1932) envisioned peaks and valleys of fitness across potential genetic combinations
- He wrote: "*it may be taken as certain that there will be an enormous number of widely separated harmonious combinations*"
- In this context, evolution can be seen as "walks" across the landscape and adaptation as "climbs" to higher positions

De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Multi-dimensional fitness landscapes

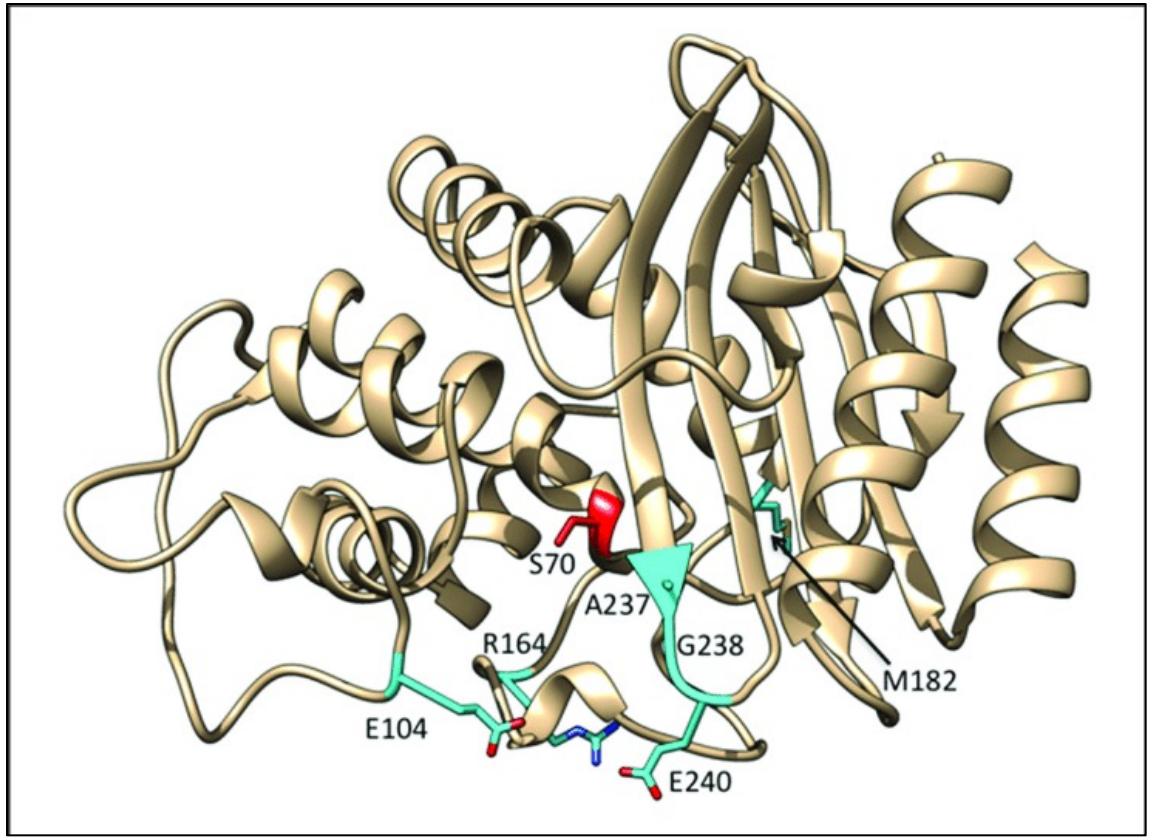
Fitnesses of genotypes in the β-lactamase TEM1



De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Evolution of antibiotic resistance

TEM1 β-lactamase

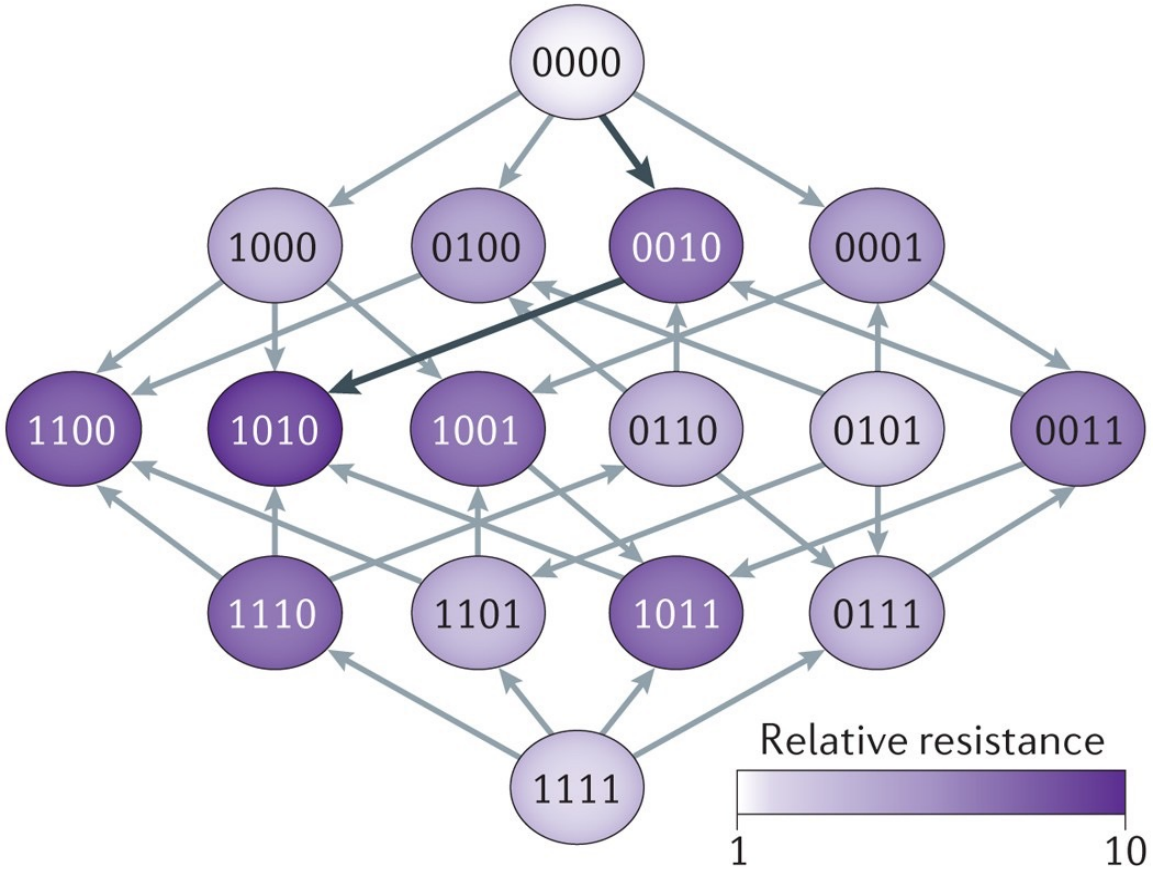


- Evolution of antibiotic resistance is a major challenge for human health
- A common mechanism of resistance in Gram-negative bacteria involves the production of β-lactamases that hydrolyze β-lactam antibiotics
- Some variants of the antibiotic resistance enzyme TEM1 allow resistance to the antibiotic cefotaxime
- Microbial adaptation to hydrolyze cefotaxime has been the subject of several studies, including those to understand the mutational landscape of antibiotic resistance

Palzkill, 2018

Multi-dimensional fitness landscapes

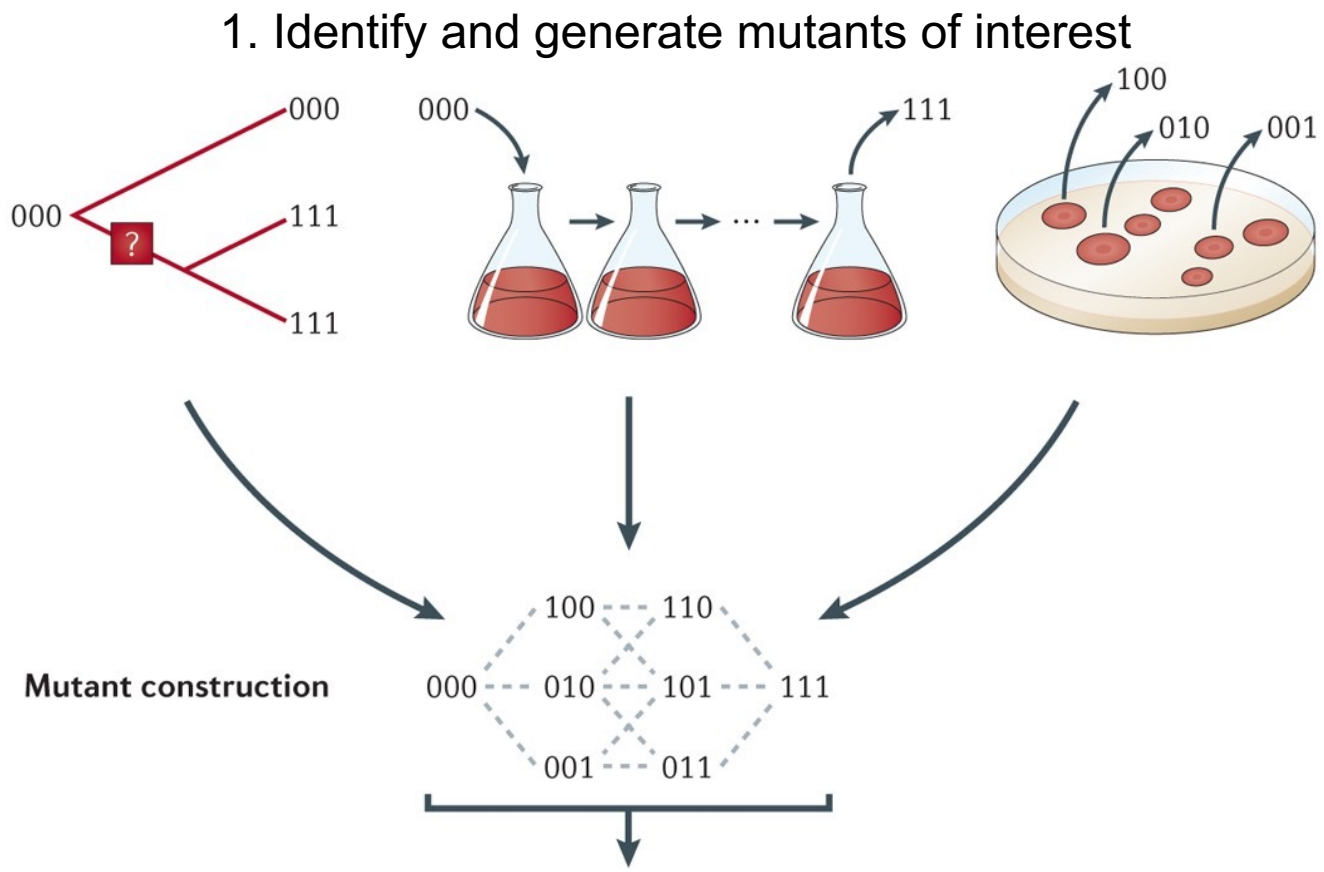
Fitnesses of genotypes in the β-lactamase TEM1



- An empirical fitness landscape involving 4 mutations in the TEM1 gene is shown
- Nodes represent the different possible amino acid allelic combinations ($2^4 = 16$)
- Wild type is represented by ’0000’ and ‘1’s represent derived alleles
- Bold black arrows indicate the ‘greedy’ walk in which the existing genotype is substituted with the largest-benefit mutation among those available

De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

The empirical study of fitness landscapes



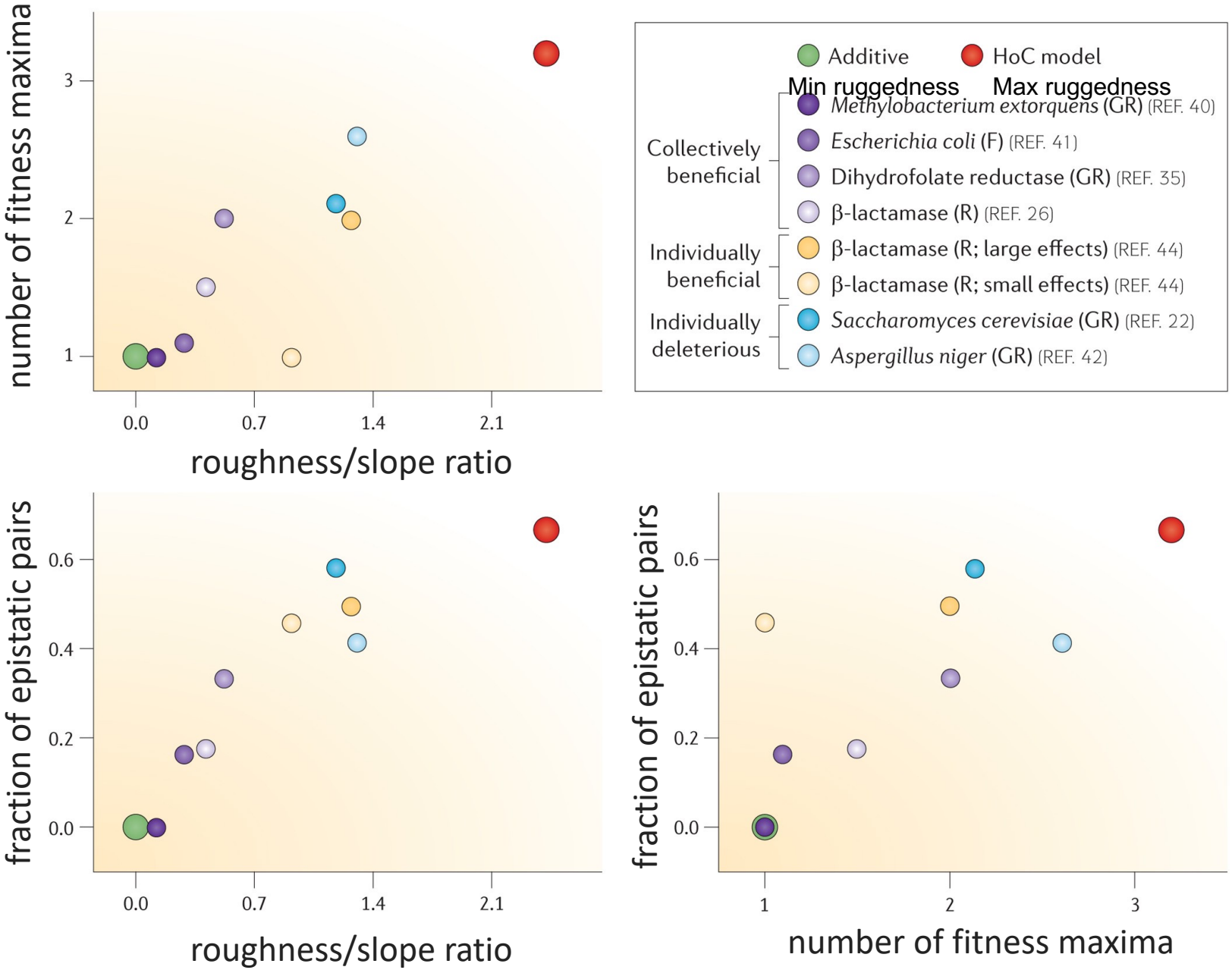
- Identify mutants of interest and construct mutants to contain all 2L possible combinations of L selected mutation
- Test each combination to assess the importance of the marginal and epistatic effects on fitness

De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Ruggedness across 8 fitness landscapes

House of cards (HoC) model:
Maximally rugged model in which fitness values of different genotypes are independent and identically distributed random variables

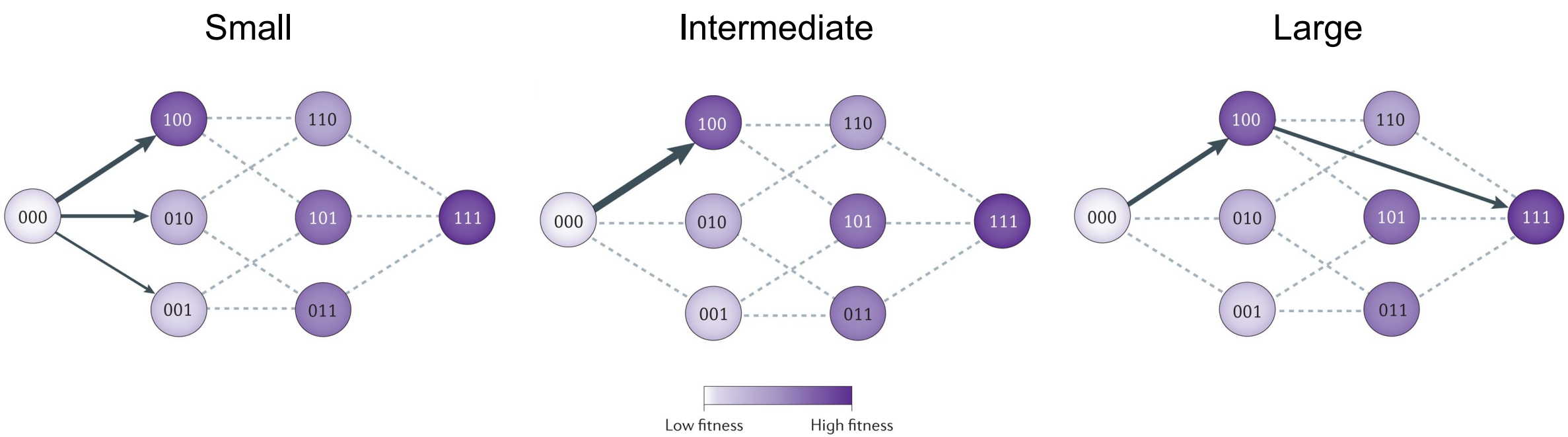
Additive model:
No epistasis



De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Mutational trajectories for populations of different size

Population size:



Large populations more often adapt through ideal trajectories and to higher fitness peaks

De Visser and Krug, NRG, 2014
https://www.nature.com/articles/nrg3744

Is evolution predictable?

- Some studies have found evidence for convergent or parallel evolution using the same genes, but others have shown evolution is often contingent on random events
- There are higher frequencies of shared adaptive changes at the level of genes and pathways relative to specific nucleotide substitutions
- Some factors that can impact repeatability:
  - Pathway accessibility and number of potential paths, which are due to characteristics of the fitness landscape (single peak vs multiple peaks, ruggedness)
  - Mutational availability, which is due to the population size and mutational target size at a locus

https://www.nature.com/articles/nrg3744

Summary

- Complex disease trait architectures appear to be highly polygenic, with contributions from some expected genes as well as from many other genes across the genome
- Many factors have the potential to impact the genetic architecture of adaptation, including the size of a population, the strength of selection on a trait, the genome-wide beneficial mutational target size and beneficial mutation rate, and population structure