

# General Multilocus Results

# Recombination and linkage disequilibrium

- $n$  loci,  $N = \{1, 2, \dots, n\}$
- $m_k$  alleles  $A_{i_k}^{(k)}$  (with  $i_k = 1, \dots, m_k$ ) at locus  $k$
- Gamete  $A_{i_1}^{(1)} A_{i_2}^{(2)} \dots A_{i_n}^{(n)}$  is abbreviated by  $i = (i_1, \dots, i_n)$
- Frequency of gamete  $i : p_i$
- $\mathbf{p} = (p_1, \dots, p_n)$

- Fitness of genotype  $ij$  and gamete  $i$ :  $W_{ij}$ ,  $W_i$
- Mean fitness:  $\bar{W}$
- Let  $I, J$  be a nontrivial decomposition of the set of loci,  $N$ , i.e.,

$$I \cup J = N, I \cap J = \phi, 1 \in I.$$

- Let  $c_I$  denote the probability of reassociation of the genes at the loci in  $I$ , inherited from one parent, with the genes at the loci in  $J$ , inherited from the other.

# Dynamics under selection and recombination

$$p_i' = p_i \frac{W_i}{\bar{W}} - D_i,$$

where

$$D_i = \frac{1}{\bar{W}} \sum_j \sum_I c_I (W_{ij} p_i p_j - W_{i_1 j_1, j_1 i_1} p_{i_1 j_1} p_{j_1 i_1})$$

represents a measure of the linkage disequilibrium in gamete  $i$ .

# The linkage-equilibrium manifold

$$\Lambda_0 = \{\mathbf{p} : p_i = p_{i_1}^{(1)} p_{i_2}^{(2)} \cdot \dots \cdot p_{i_n}^{(n)} \text{ for every } i\}$$

Points on  $\Lambda_0$  are uniquely determined by the vector of gene frequencies  $\rho$ . If there is no position effect,  $W_{ij} = W_{i_1 j_1, j_1 i_1}$ , then

$$\Lambda_0 \subseteq \{\mathbf{p} : \mathbf{D} = \mathbf{0}\}$$

## Recombination alone

$$p_i' = p_i - D_i$$

Then

$$\Lambda_0 = \{\mathbf{p} : \mathbf{D} = \mathbf{0}\}$$

- is invariant and globally attracting at the uniform geometric rate  $1 - c_{\min}$ , where  $c_{\min} = \min c_{kl} > 0$ .
- All points on  $\Lambda_0$  are fixed points.

## Recombination alone

- The recombination dynamics can be linearized by a procedure called Haldane linearization. This, however, is a recursive algorithm.
- Explicit transformations to coordinates that linearize the dynamics exist only for special cases (only one recombination event).
- Many other measures of linkage disequilibrium have been used, depending on the context.

# Weak selection (loose linkage)

- $W_{ij} = 1 + sr_{ij}$  with  $|r_{ij}| \leq 1$  and  $s/c_{\min} \ll 1$
- For weak selection, a smooth invariant manifold  $\Lambda_s$  close to  $\Lambda_0$  exists, which is globally attracting at a geometric rate.

## Quasi linkage equilibrium

- $\Lambda_s$  is given by equations of the form

$$D_i = s\psi_i(\rho, s),$$

and is called the *quasi-linkage equilibrium manifold*.

- Thus, after some time, the linkage disequilibria are  $O(s)$ . Their between-generation change is then of order  $O(s^2)$ .

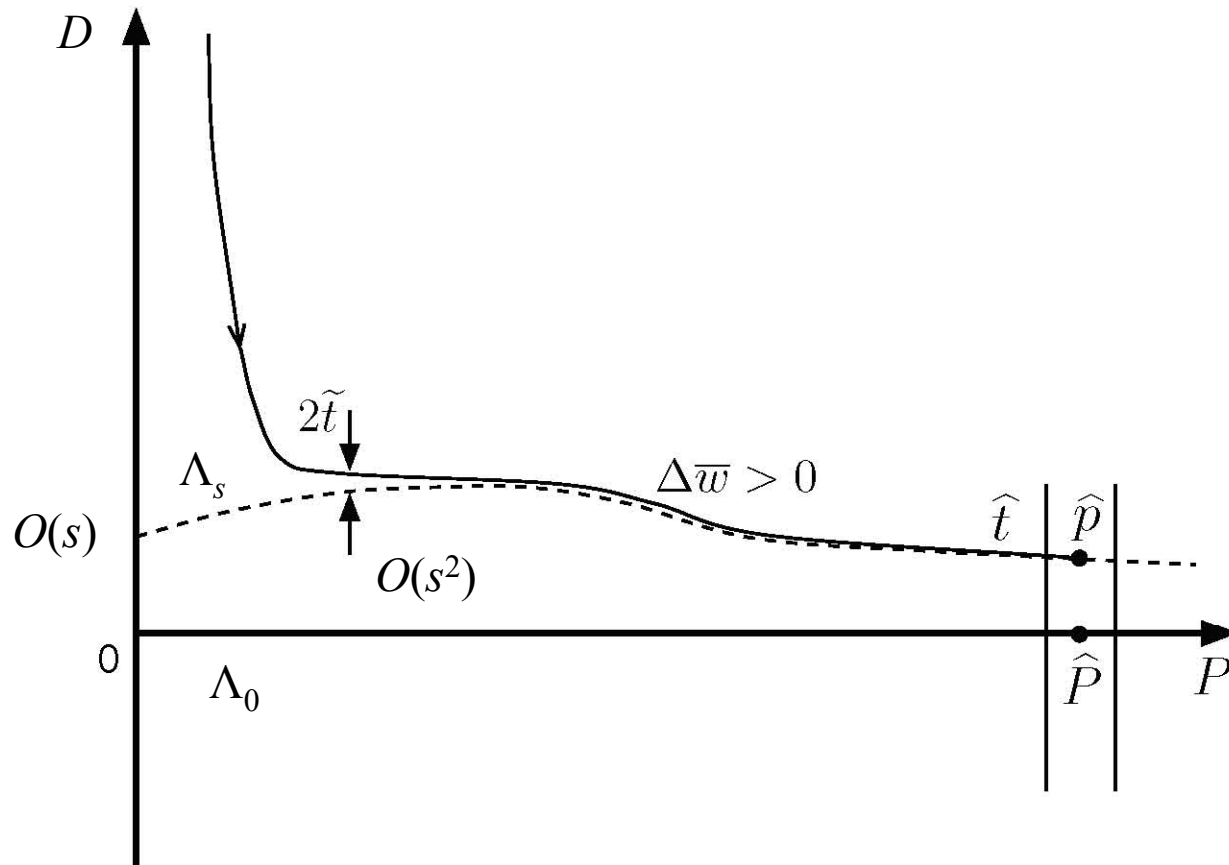
## Main results

- The approximate dynamics on  $\Lambda_0$  (by *assuming* linkage equilibrium) is gradient like.
- This is also true for the weak-selection limit on  $\Lambda_s$ , which is continuous in time.
- If  $s$  is sufficiently small and all equilibria of the weak-selection limit are hyperbolic, then each solution of the full dynamics converges to a fixed point (on  $\Lambda_s$ ) as  $t \rightarrow \infty$ .

## Further results

- The fixed points of the full dynamics are small perturbations of those of the weak-selection limit on  $\Lambda_0$ .
- Mean fitness may decrease during the short period, of order  $\log(1/s)$  generations, of approach to  $\Lambda_s$ . It may also decrease close to equilibrium, generically after a long time of order  $1/s$ .
- For intermediate times, mean fitness must increase.

# The approach to quasi linkage equilibrium



# Weak epistasis

$$W_{ij} = \sum_{k=1}^n \alpha_{i_k j_k}^{(k)} + \varepsilon r_{ij},$$

where  $\varepsilon$ , the strength of epistasis, is sufficiently small.

- If  $\varepsilon = 0$ , there is no epistasis and fitness is additive (between loci).

## No epistasis

- The gametic fitnesses in the next generation are independent of the recombination rates.
- Mean fitness is nondecreasing and depends only on the gene frequencies.
- $\Delta \bar{W} = 0 \iff p_i(W_i - \bar{W}) = 0 \quad \forall i.$
- A point  $\mathbf{p}$  is an equilibrium if and only if it is both a selection equilibrium for each locus and it is in linkage equilibrium.

## Main results for weak epistasis

- If for  $\varepsilon = 0$ , each equilibrium is hyperbolic, then for sufficiently small  $\varepsilon$ , each trajectory converges to an equilibrium point.
- These equilibria are small, i.e.,  $O(\varepsilon)$ , perturbations of those with  $\varepsilon = 0$ .
- $\mathbf{D} = O(\varepsilon)$  and mean fitness differs also only by  $O(\varepsilon)$  from its value without epistasis.

# Exact recursions for quantitative genetic models

- Are very cumbersome, but have been accomplished.
- Needs the (widely made) assumptions of additive effects of loci (on the trait, not fitness!).
- Uses multivariate moments or cumulants.

# The weak-selection response for multivariate (within-gamete) cumulants

$$\begin{aligned}
 \tilde{\Delta}_s \kappa_i &= s_1 \kappa_{i.} + s_2 [\kappa_{i..} + 2C_1 \kappa_{i.}] \\
 &\quad + s_3 [\kappa_{i...} + 3C_1 \kappa_{i..} + 3(C_1^2 + C_2) \kappa_{i.}] \\
 &\quad + s_4 [\kappa_{i....} + 4C_1 \kappa_{i...} + 6(C_1^2 + C_2) \kappa_{i..} + 4(C_1^3 + 3C_1 C_2 + C_3) \kappa_{i.}] + \dots, \\
 \tilde{\Delta}_s \kappa_{ij} &= s_1 \kappa_{ij.} + s_2 [\kappa_{ij..} + 2C_1 \kappa_{ij.} + 2\kappa_{i.} \kappa_{j.}] \\
 &\quad + s_3 [\kappa_{ij...} + 3(\kappa_{i..} \kappa_{j.} + \kappa_{j..} \kappa_{i.}) + 3C_1 (\kappa_{ij..} + 2\kappa_{i.} \kappa_{j.}) + 3(C_1^2 + C_2) \kappa_{ij.}] \\
 &\quad + s_4 [\kappa_{ij....} + 4(\kappa_{i.} \kappa_{j...} + \kappa_{j.} \kappa_{i...}) + 6\kappa_{i..} \kappa_{j..} \\
 &\quad\quad + 4C_1 (\kappa_{ij...} + 3\kappa_{i.} \kappa_{j..} + 3\kappa_{j.} \kappa_{i..}) + 6(C_1^2 + C_2) (\kappa_{ij..} + 2\kappa_{i.} \kappa_{j.}) \\
 &\quad\quad + 4(C_1^3 + 3C_1 C_2 + C_3) \kappa_{ij.}] + \dots, \\
 \tilde{\Delta}_s \kappa_{ijk} &= s_1 \kappa_{ijk.} + s_2 [\kappa_{ijk..} + 2C_1 \kappa_{ijk.} + 2(\kappa_{ij.} \kappa_{k.} + **)] \\
 &\quad + s_3 \{ \kappa_{ijk...} + 6\kappa_{i.} \kappa_{j.} \kappa_{k.} + 3(\kappa_{ij.} \kappa_{k..} + \kappa_{ij..} \kappa_{k.} + **) \\
 &\quad\quad + 3C_1 [\kappa_{ijk..} + 2(\kappa_{ij.} \kappa_{k.} + **)] + 3(C_1^2 + C_2) \kappa_{ijk.} \} + \dots, \\
 \tilde{\Delta}_s \kappa_{ijkl} &= s_1 \kappa_{ijkl.} + s_2 [\kappa_{ijkl..} + 2C_1 \kappa_{ijkl.} + 2(\kappa_{ijk.} \kappa_{l.} + ***) + 2(\kappa_{ij.} \kappa_{kl.} + **)] \\
 &\quad + s_3 \{ \kappa_{ijkl...} + 3(\kappa_{ij..} \kappa_{kl.} + ****) + 6(\kappa_{ij.} \kappa_{k.} \kappa_{l.} + ****) \\
 &\quad\quad + 3[(\kappa_{ijk..} \kappa_{l.} + \kappa_{ijk.} \kappa_{l..}) + **] \\
 &\quad\quad + 3C_1 [\kappa_{ijkl..} + 2(\kappa_{ij.} \kappa_{kl.} + **) + 2(\kappa_{ijk.} \kappa_{l.} + **)] \\
 &\quad\quad + 3(C_1^2 + C_2) \kappa_{ijkl.} \} + \dots,
 \end{aligned}$$

where  $\kappa_{i.} = \sum_{j=1}^{\ell} \kappa_{ij} = c_{e_i.}$ ,  $\kappa_{i..} = \sum_{j,k=1}^{\ell} \kappa_{ijk} = c_{e_i..}$ , etc., and \*\* (\*\*\*, \*\*\*\*, \*\*\*\*\*) denotes two (three, five) additional terms, each obtained by permuting the subscripts.

# The change of multivariate cumulants across generations (selection + recombination)

$$\begin{aligned}
 \Delta \kappa_i &= \Delta_s \kappa_i, \\
 \Delta \kappa_{ij} &= -r_{ij} \kappa_{ij} + (1 - r_{ij}) \Delta_s \kappa_{ij} + r_{ij} \Delta_s \kappa_{i,j}, \\
 \Delta \kappa_{ijk} &= -r_{ijk} \kappa_{ijk} + (1 - r_{ijk}) \Delta_s \kappa_{ijk} \\
 &\quad + r_{i,jk} \Delta_s \kappa_{i,jk} + r_{j,ki} \Delta_s \kappa_{j,ki} + r_{k,ij} \Delta_s \kappa_{k,ij}, \\
 \Delta \kappa_{ijkl} &= -r_{ijkl} \kappa_{ijkl} + (1 - r_{ijkl}) \Delta_s \kappa_{ijkl} \\
 &\quad + (r_{i,jkl} \Delta_s \kappa_{i,jkl} + ***) + (r_{ij,kl} \Delta_s \kappa_{ij,kl} + **) \\
 &\quad + (\kappa_{ij} + \Delta_s \kappa_{ij})(\kappa_{kl} + \Delta_s \kappa_{kl}) R(ij; kl) + ** \\
 &\quad + \Delta_s \kappa_{i,j} (\kappa_{kl} + \Delta_s \kappa_{kl}) R(i, j; kl) + ***** \\
 &\quad + \Delta_s \kappa_{i,j} \Delta_s \kappa_{k,l} R(i, j; k, l) + **,
 \end{aligned}$$

where \*\* (\*\*\*, \*\*\*\*\*) denotes two (three, five) similar terms with permuted subscripts and  $R(ij; kl) = (1 - r_{ijkl}) + r_{ij,kl} - (1 - r_{ij})(1 - r_{kl})$ ,  $R(i, j; kl) = r_{i,jkl} + r_{j,ikl} - r_{ij}(1 - r_{kl})$ ,  $R(i, j; k, l) = r_{ik,jl} + r_{il,jk} - r_{ij}r_{kl}$ . The factor  $R(ij; kl)$  is a measure of association between the pair of loci  $\{ij\}$  and the pair  $\{kl\}$ ; it is zero if the event that  $\{ij\}$  stay together is independent of the event that  $\{kl\}$  stay together.

## Good news

- Kirkpatrick, Johnson & Barton (2002) have automated and implemented (as *Mathematica* packages) a much more general version.
- There a number of interesting, general insights from this exercise.
- There are several simple consequences.

## Some simple consequences

- Response of the mean phenotype:

$$\begin{aligned}\Delta \bar{P} &= \Delta \bar{G} = \Delta_s \bar{G} = \tilde{\Delta}_s \bar{G} / \bar{W} \\ &= \frac{1}{\bar{W}} \sum_{k=0}^K s_k M_{k+1}^0 - \bar{G} \\ &= \frac{1}{\bar{W}} \text{Cov}_G(G, W)\end{aligned}$$

(Robertson's secondary theorem of natural selection).

## Some simple consequences

- Quasi linkage equilibrium approximation for the response of the genetic variance:

$$\Delta\sigma_G^2 = s_1 C_3 + s_2 (C_4 + 2C_1 C_3 + 4\sum_{i=1}^n \kappa_{ii}^2) + \dots$$
$$+ \sum_{i=1}^n \mu_i \gamma_i^2 + O(s^2).$$

- Also the breeder's equation and Lande's equation can be derived.

# Hot Topics

# Maintenance of quantitative genetic variation

# The problem

- Almost all traits exhibit substantial genetic variation, with heritabilities typically in the range  $0.15 \leq h^2 \leq 0.80$ .
- By which mechanisms can these be explained, given that many, if not most traits, are under stabilizing selection?
- If there is a single, main cause, it must be mutation-selection balance.

# Findings

- Extensive mathematical analyses have produced good approximations for the variance maintained under mutation-stabilizing selection balance.
- Unless available measurements for mutation rates of genes affecting quantitative traits, are much too low, heritabilities of  $h^2 > 0.2$  cannot be explained.

## Alternative hypothesis

- Apparent stabilizing selection and pleiotropy
- Balancing selection
  - Heterozygote advantage
  - Migration (geographic structure)
  - Frequency-dependent selection
  - Temporally fluctuating selection
- Genotype-environment interaction
- Populations are rarely at equilibrium

# Epistasis and the evolution of genetic architecture

# Epistasis

- Refers to the interaction of individual gene effects.
- Is involved in the explanation of many evolutionary phenomena, e.g.,
  - reproductive isolation of species
  - evolution of mutational robustness (canalization)
  - evolution of sex and recombination

# Epistasis

can occur on

- the level of fitness, i.e., fitness effects can be (and mostly are) epistatic, and on
- the level of traits, i.e., genetic effects on a trait can be epistatic.

Most of quantitative genetics assumes absence of epistasis for traits but not for fitness.

# Concepts of epistasis

- *Statistical definition:*
  - Residual effects of an analysis of variance, i.e., noise arising from deviation from additivity
- *Functional definition:*
  - Functional epistasis occurs whenever the effect of a genetic substitution (on one or multiple loci) depends on the genetic background

## The simplest realization

$$P = G + E,$$

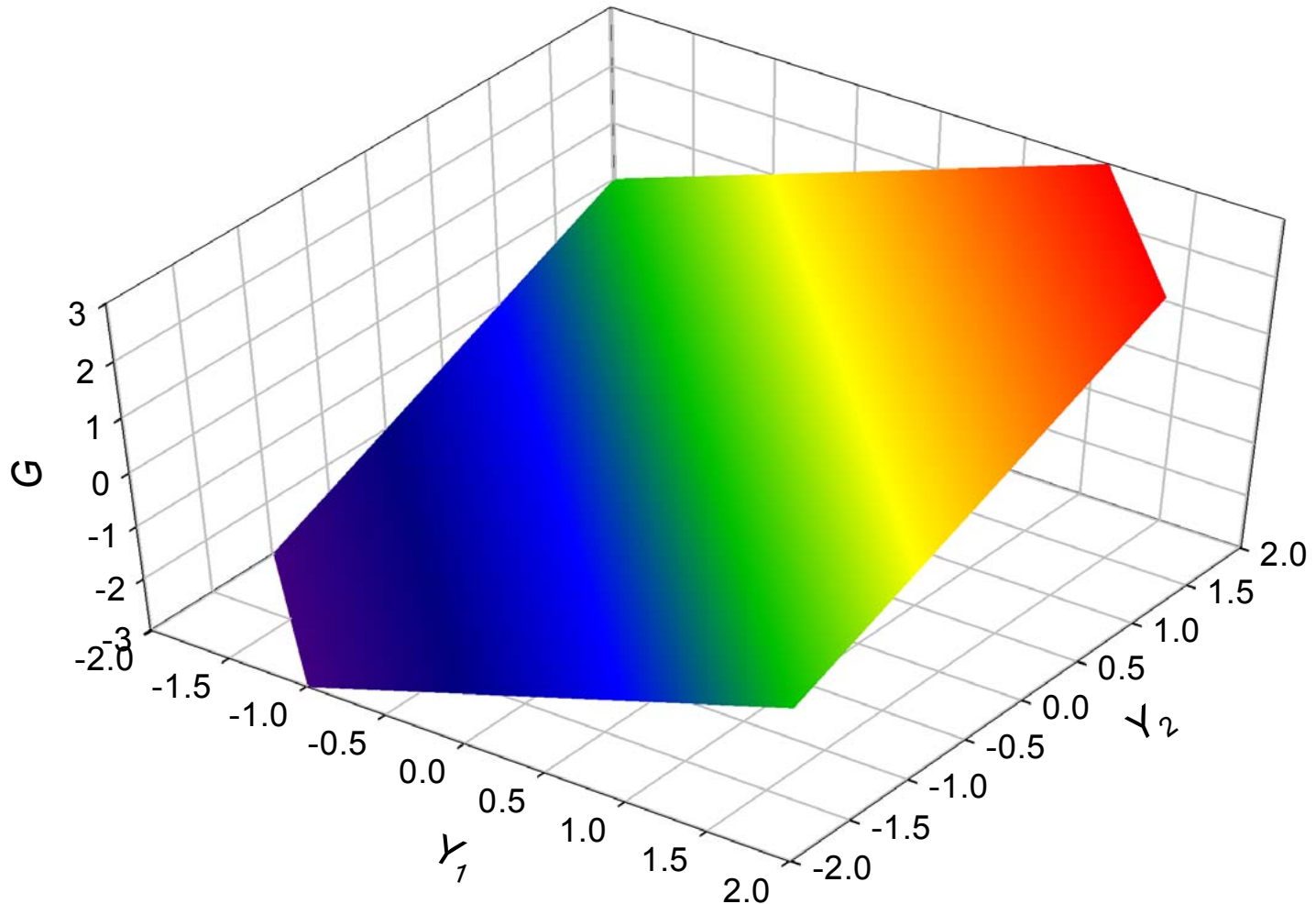
$$G = G_r + \sum_i Y_i + \sum_{i,j>i} \varepsilon_{ij} Y_i Y_j.$$

- This is a special case of the multilinear model of gene interactions developed by Hansen, Hermisson, and Wagner (2001+).
- $G_r$  is the genotypic value of the reference genotype, relative to which all allelic effects are measured.

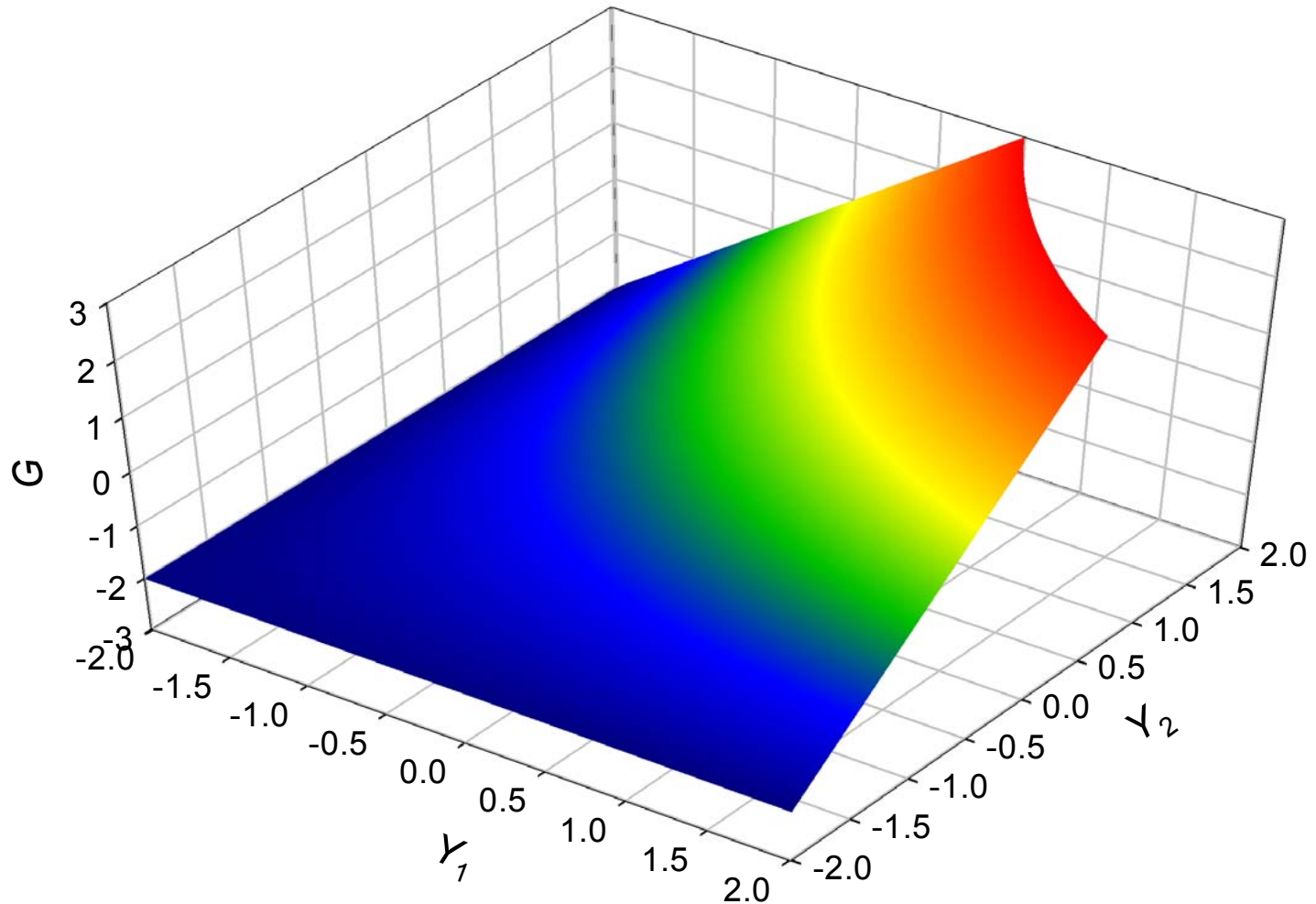
Phenotype landscapes for

$$G = Y_1 + Y_2 + \varepsilon Y_1 Y_2$$

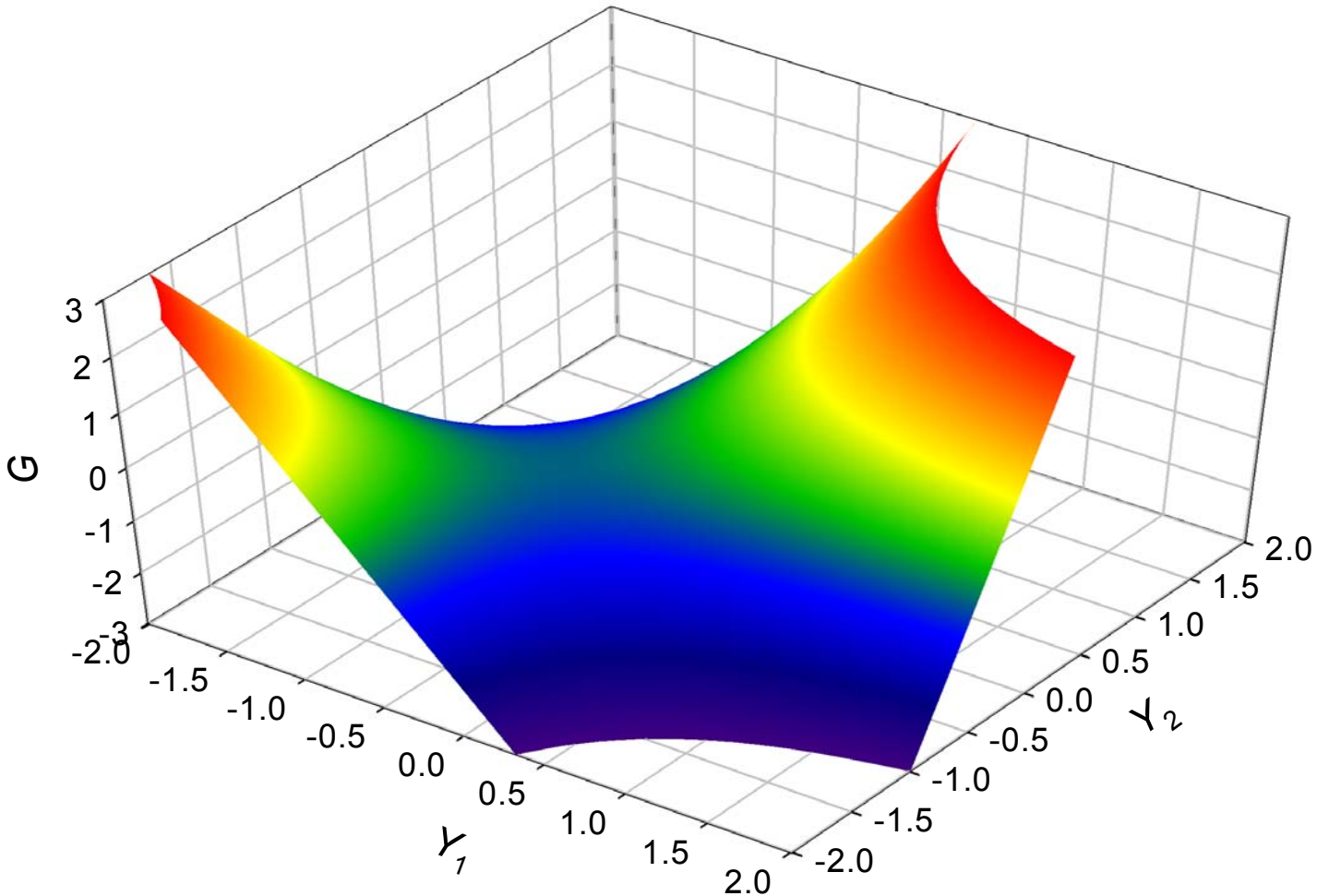
# The phenotype landscape without epistasis ( $\epsilon = 0$ )



# The phenotype landscape with weak epistasis ( $\varepsilon = 0.5$ )



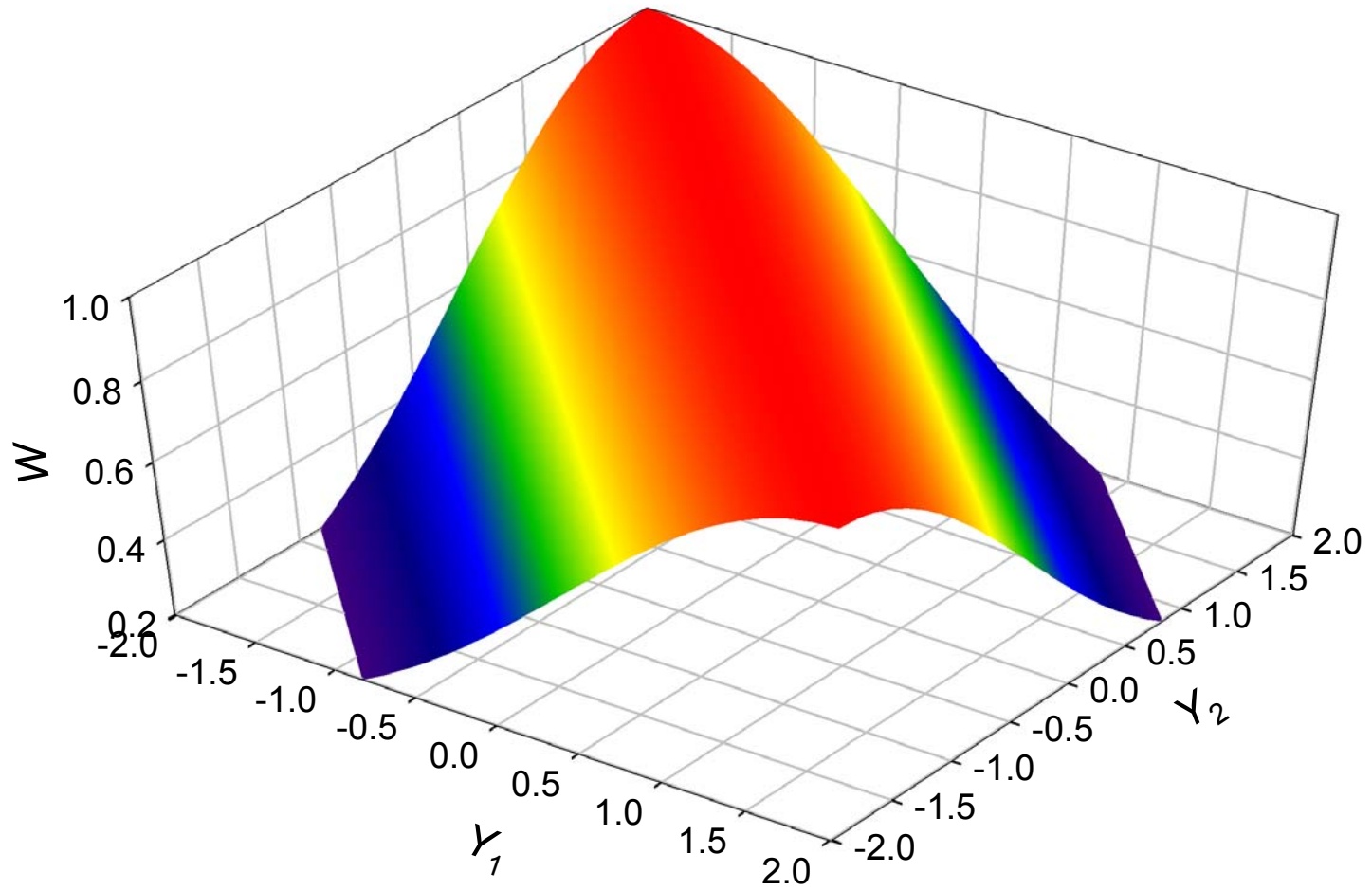
# The phenotype landscape with strong epistasis ( $\varepsilon = 2$ )



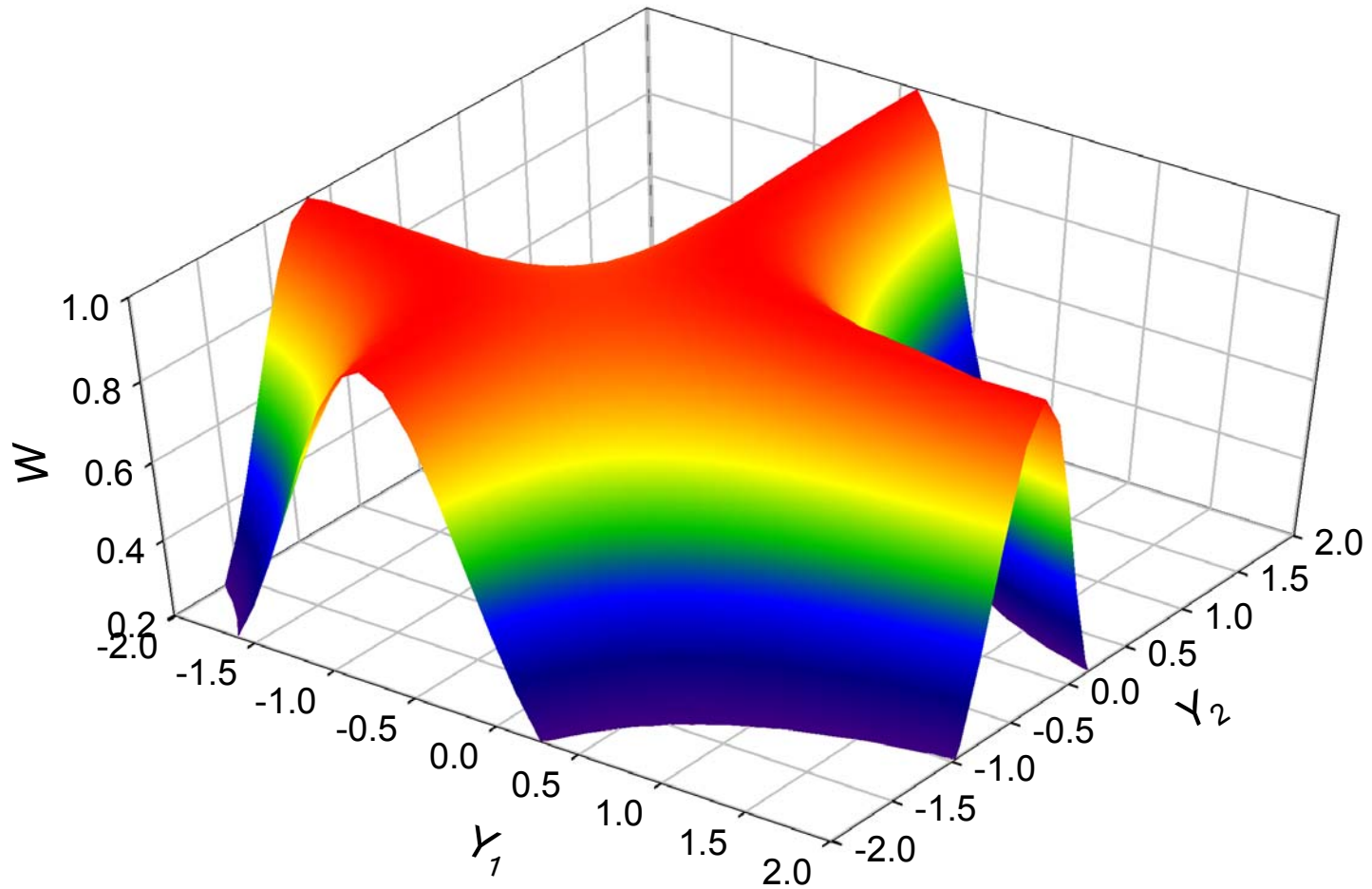
Fitness landscapes for

$$W(G) = \exp(-G^2)$$

# The fitness landscape without epistasis



# The fitness landscape with epistasis



## Consequences of epistasis

- The important point that distinguishes the epistatic landscape from a purely additive model is that *the variational properties* (i.e., the effects of single and multiple mutants on phenotype) *are able to change* as the population evolves on the landscape.
- Variational properties may change even when the population evolves on isophenotype contours!

## Mutation–stabilizing selection balance

- Loci with different mutation rates evolve large differences in their variances of mutational effects, with the locus with the higher mutation rate evolving the smaller variance.
- The epistatic force acts to reduce the phenotypic effect of genetic variation.
- Loci with high mutation rates evolve buffering (robustness), others not.

# Evolvability and Robustness

# Evolvability

- Evolvability can be viewed as the intrinsic capacity of a genome to produce adaptive variants (Wagner and Altenberg, 1996)
- In quantitative genetics, evolvability is closely related to the variability and standing genetic variance of phenotypic traits.
- For multivariate phenotypes, the relevant parameter is the mutational matrix, which describes the effects of new mutations on trait variances and covariances.

# Phenotypic robustness

- A character state that has evolved under natural selection is phenotypically robust if the variability of the character under a given source of variation is significantly reduced in this state as compared to a set of alternative states.
- Phenotypes can be buffered against heritable and nonheritable variables.
- Canalization against macro-environmental perturbations is one classical example.

# Pleiotropy and the evolution of multivariate complex traits

# Problems

- Recall:

$$\Delta \bar{\mathbf{P}} = \mathbf{G} \nabla \ln \bar{W}$$

- Therefore, the ability of existing theory to reconstruct the evolutionary history of selection or to predict the future trajectory of evolution of a population depends crucially on the dynamics of the **G**-matrix.

# Problems

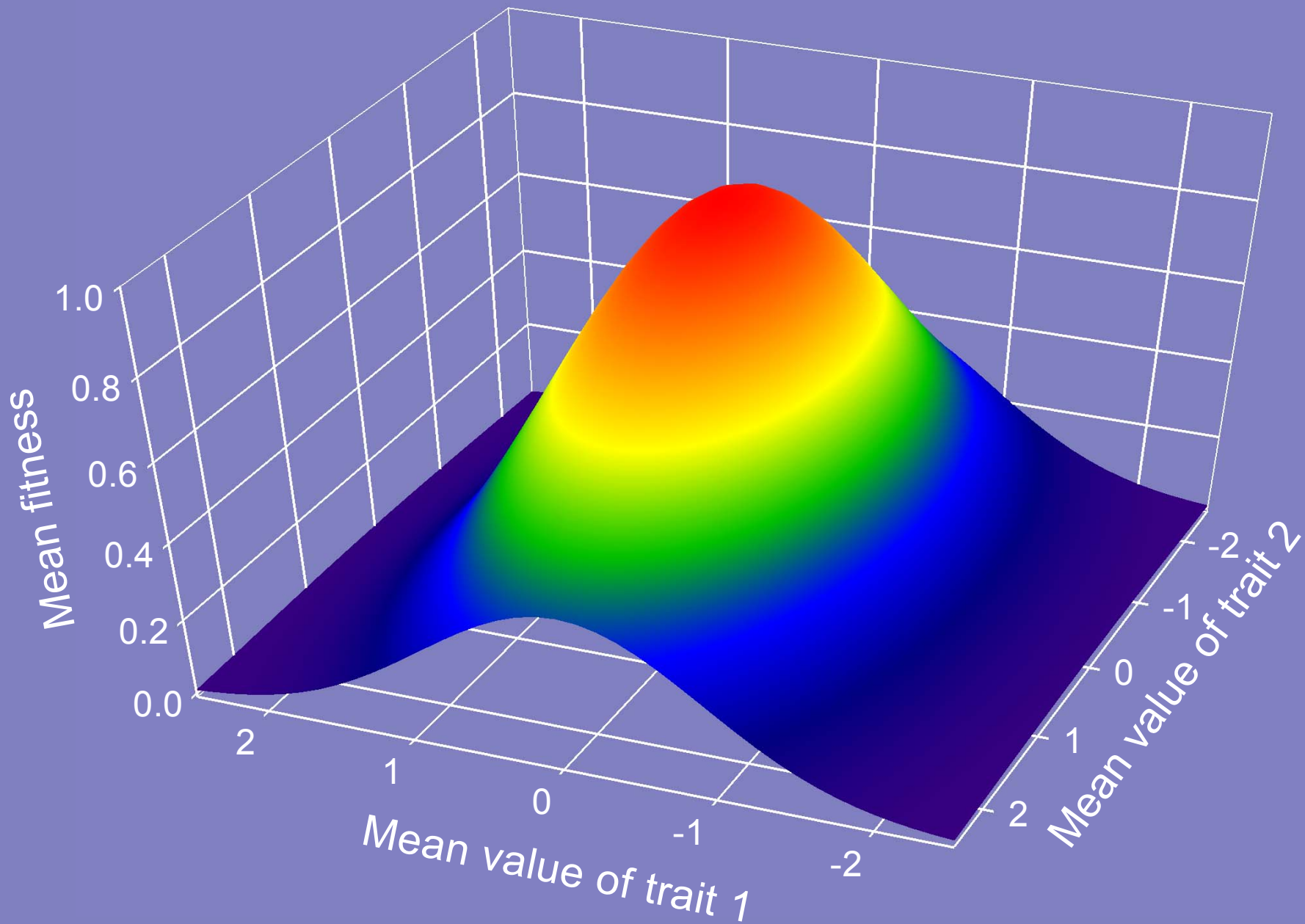
- Several structural features are remarkably constant among related populations or species.
- This suggests that persistent selection regimes or other factors promote stability.
- Existing analytical theory reveals little about the dynamics of  $G$ , because it is too complex to be mathematically tractable.

## One possible solution

- Extensive Monte-Carlo simulation studies
  - each individual has a diploid genome with many (e.g., 50) loci
  - at each locus mutations are drawn from a multivariate distribution
  - recombination is modeled explicitly
  - evolution in different adaptive landscapes is explored

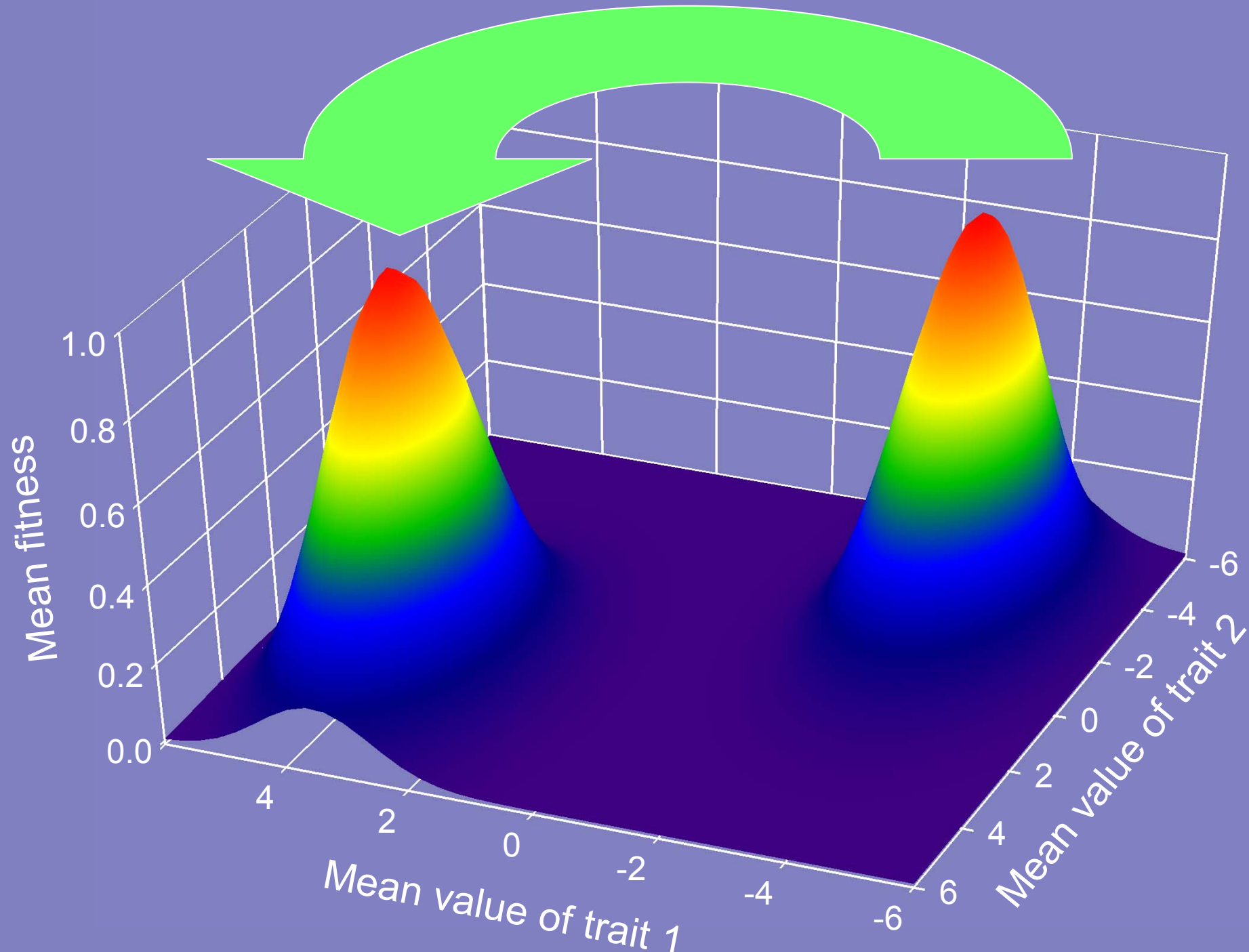
# Adaptive landscapes

- A single (multivariate) peak



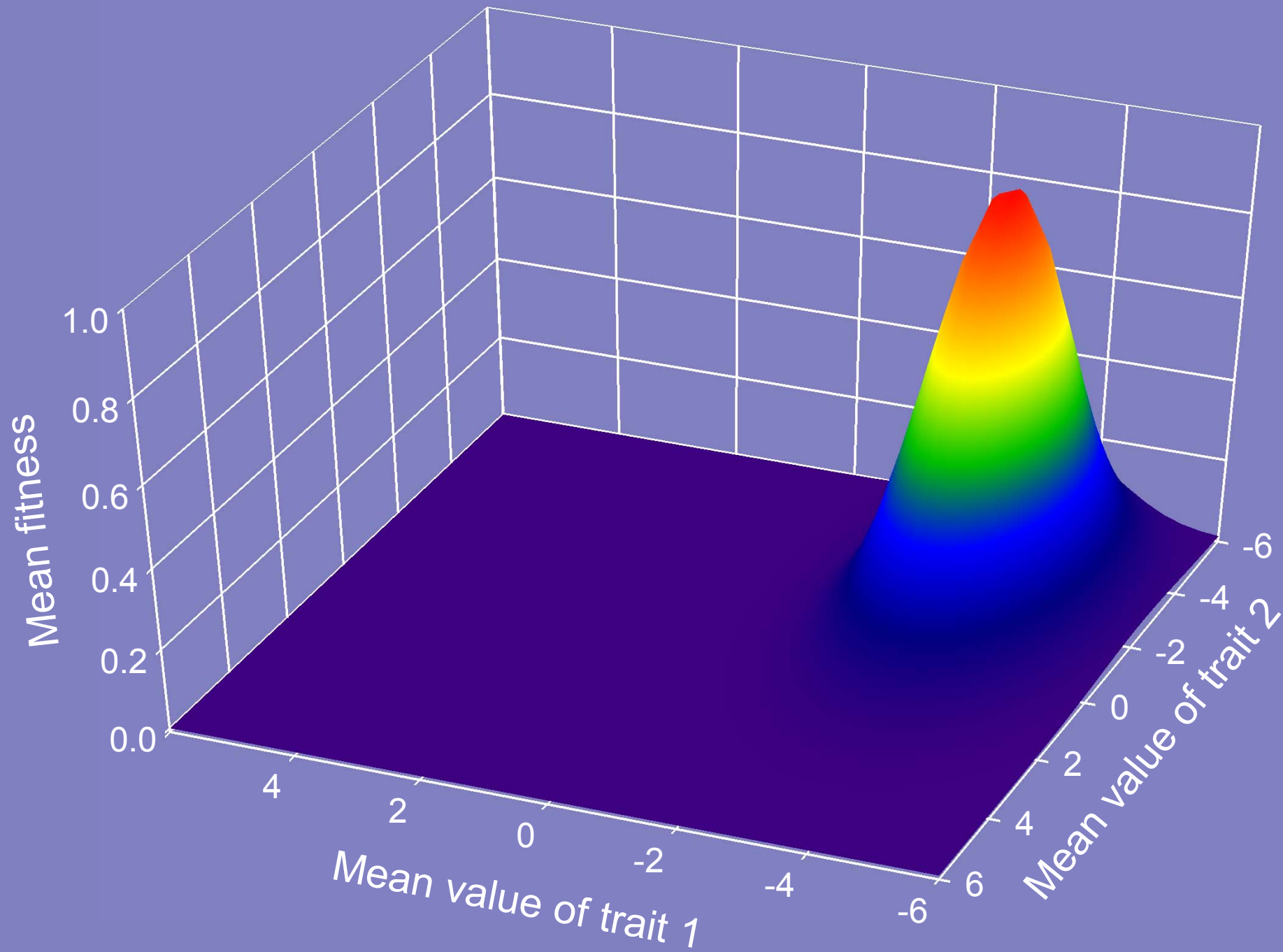
# Adaptive landscapes

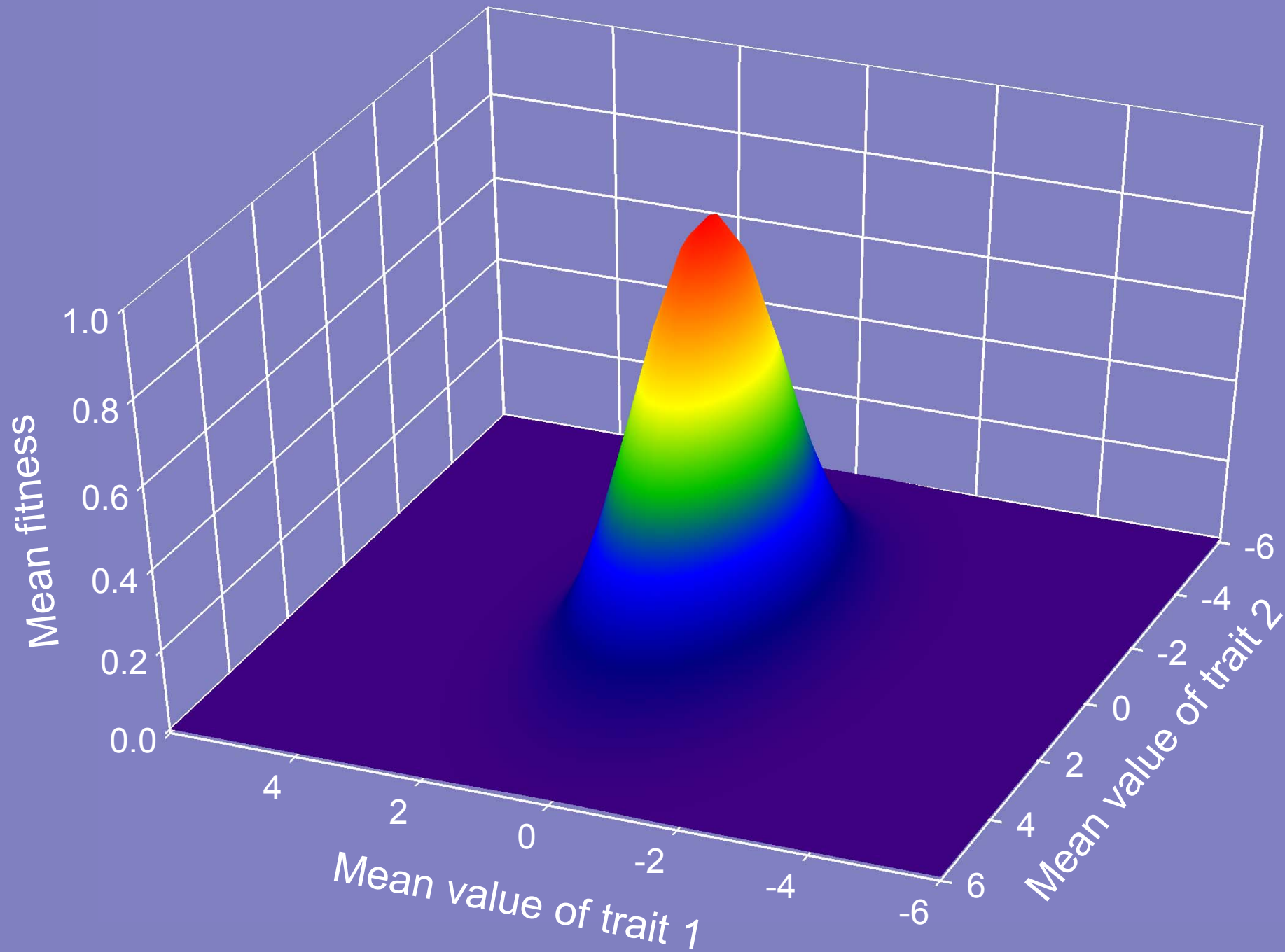
- A single (multivariate) peak
- A peak shift

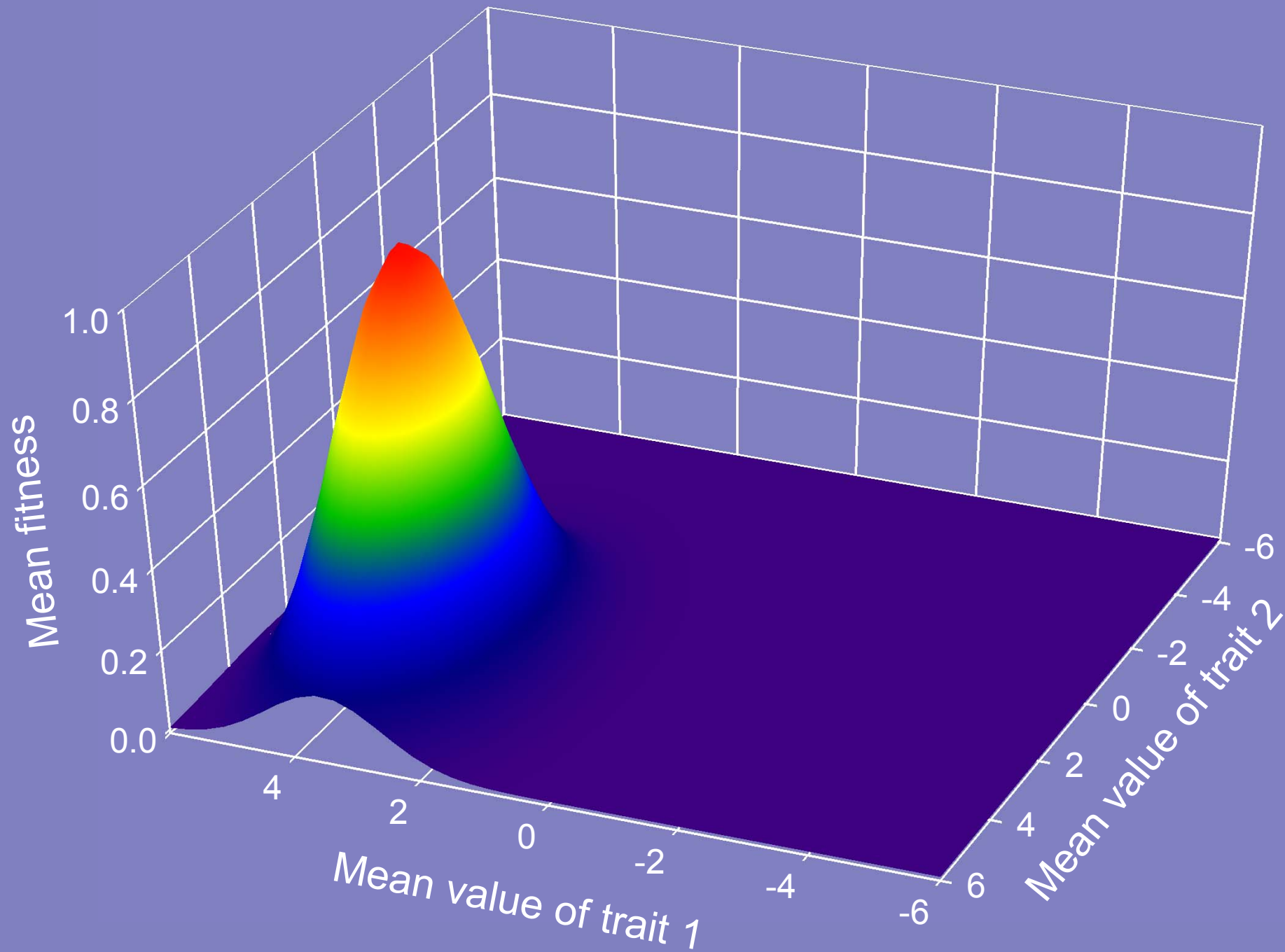


# Adaptive landscapes

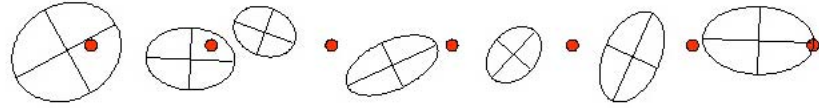
- A single (multivariate) peak
- A peak shift
- A peak with a moving optimum



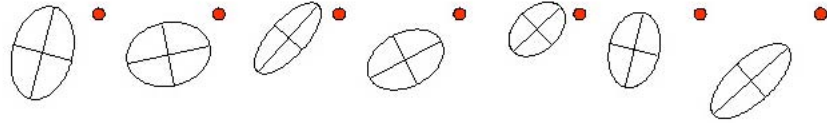




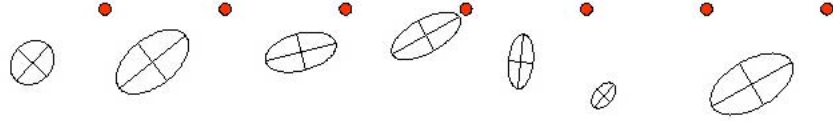
$r_{\mu} = 0$   
 $r_{\omega} = 0$



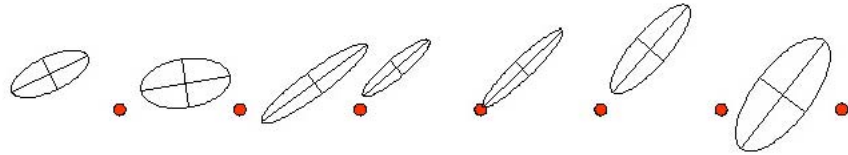
$r_{\mu} = 0$   
 $r_{\omega} = 0.75$



$r_{\mu} = 0$   
 $r_{\omega} = 0.9$



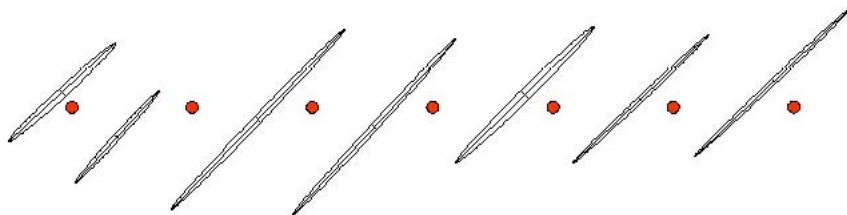
$r_{\mu} = 0.5$   
 $r_{\omega} = 0$



$r_{\mu} = 0.9$   
 $r_{\omega} = 0$



$r_{\mu} = 0.9$   
 $r_{\omega} = 0.9$



# Spatially structured populations

# Topics

- Migration between discrete demes (habitats) with different selection pressures
- Diffusion (of individuals) with spatially varying selection
- Clines
- Migration and multilocus selection

# Speciation

# Topics

- Mechanisms for post zygotic isolation, e.g.,
  - Muller-Dobzhansky incompatibilities
  - Inversions
- Parapatric speciation (isolation by distance, migration)
- Sympatric speciation