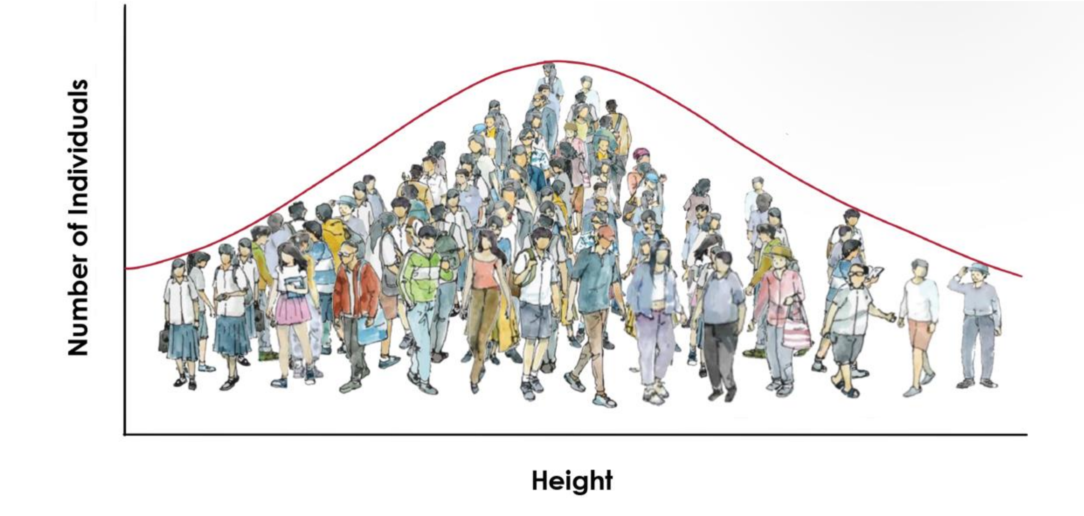
Introduction to Multivariate Quantitative Genetics



EQGW 2025– Mountain Lake

Jacqueline Sztepanacz

University of Toronto

Topics we will cover:

- Multivariate quantitative genetics

1. Pleiotropy & Genetic correlations
2. The G matrix
3. Genetic constraints

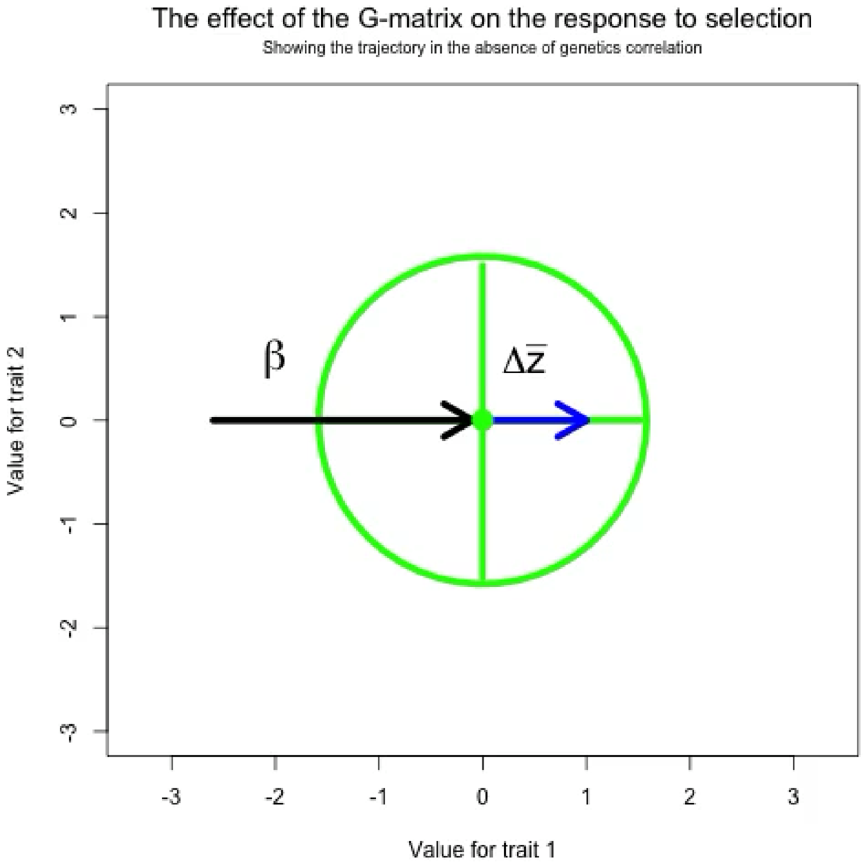
- **Selection**

1. Empirical methods to estimate selection
2. Empirical results

Selection

- genetic constraints and evolutionary lines of least resistance are only relevant in the context of selection

- how can we estimate selection?



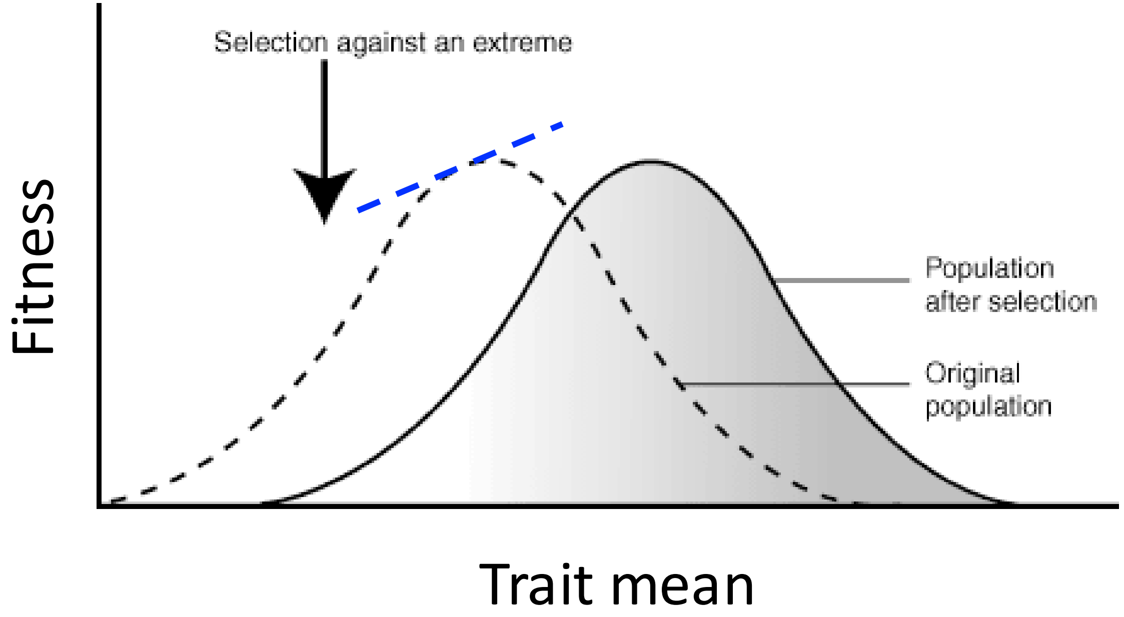
https://phenotypicevolution.com/?page_id=113

Breeder’s equation, selection differential

Lande-Arnold Regression

- Lande and Arnold (1983) proposed a method to estimate selection on phenotypes using regression
- operationally we can obtain selection gradients by estimating the partial regression coefficients of trait values on relative fitness (multiple regression)
- easy to implement, and can be used to disentangle selection acting directly and indirectly on a trait

Lande-Arnold Regression



$$\bar{w} \sim \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n$$

Lande and Arnold 1983

Partitioning selection among traits

Selection Gradient Analysis: Total selection on a trait z is the sum of its direct effect on fitness plus the indirect effects of correlated traits

$$S_z = cov(z, w) = \beta_z + r_{zy}\beta_y$$

$$S_y = cov(y, w) = r_{zy}\beta_z + \beta_y$$

Partitioning selection among traits

direct

$$S_z = cov(z, w) = \boxed{\beta_z} + r_{zy}\beta_y$$

$$S_y = cov(y, w) = r_{zy}\beta_z + \boxed{\beta_y}$$

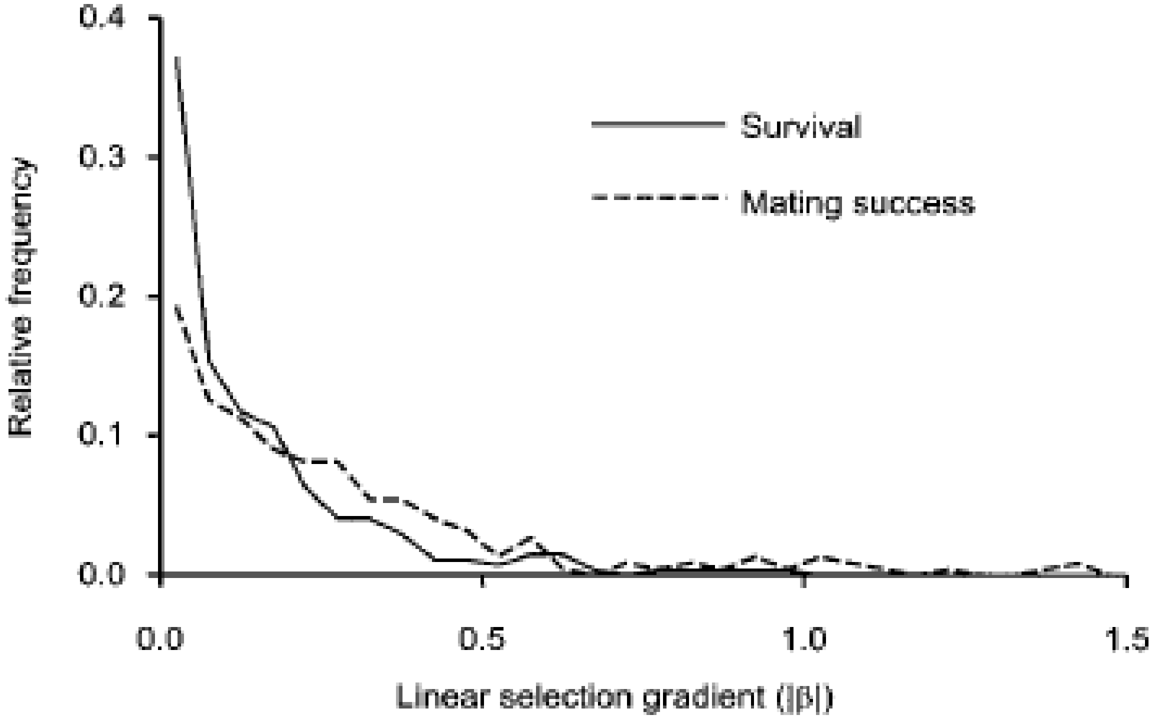
direct

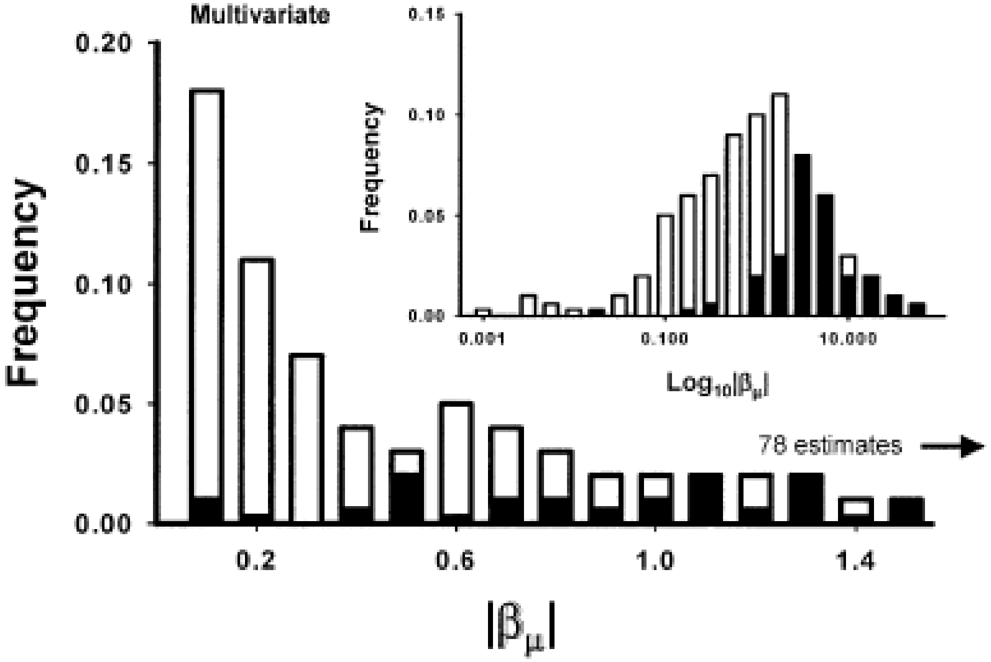
Partitioning selection among traits

$$S_z = cov(z, w) = \underbrace{\overset{\text{direct}}{\boxed{\beta_z}}} + \overset{\text{indirect through } y}{\boxed{r_{zy}\beta_y}}$$

$$S_y = cov(y, w) = \underbrace{\boxed{r_{zy}\beta_z}}_{\text{indirect through } z} + \underbrace{\boxed{\beta_y}}_{\text{direct}}$$

Directional selection is strong





- median variance standardized directional selection gradient of 0.153 for natural selection 0.250 for sexual selection
- Median mean standardized multivariate directional selection gradient was 0.54 (mean 0.28)

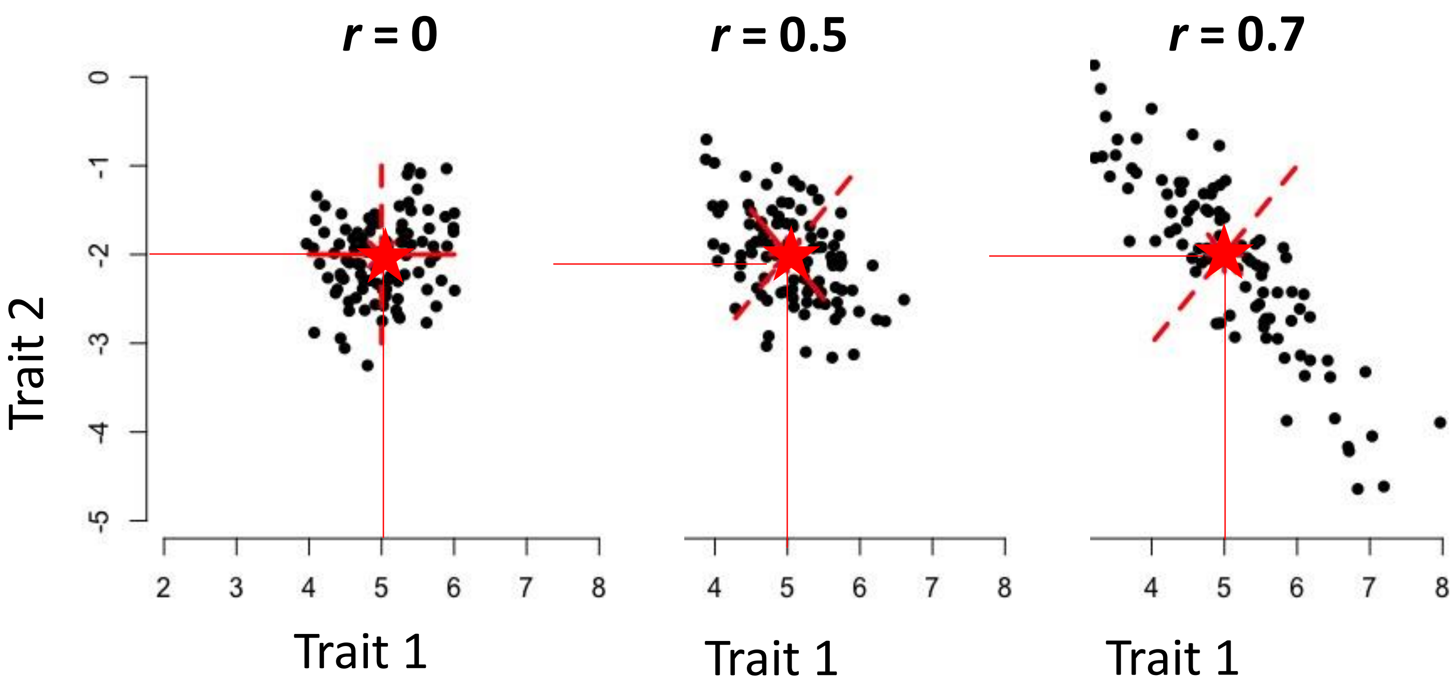
Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hoang, A., Hill, C.E., Beerli, P. and Kingsolver, J.G., 2001. Strength and tempo of directional selection in the wild. *PNAS*, *98*(16), pp.9157-9160.

Hereford, J., Hansen, T.F. and Houle, D., 2004. Comparing strengths of directional selection: how strong is strong?. *Evolution*, *58*(10), pp.2133-2143

Problems with correlated traits

- performing multiple regression on correlated traits can be a challenge because it is hard to estimate their independent effects on fitness
- this can lead to large standard errors of estimates and inaccurate estimates of selection
- over the years solutions like dropping traits, estimating selection on PCs, etc have been proposed

Problems with correlated traits

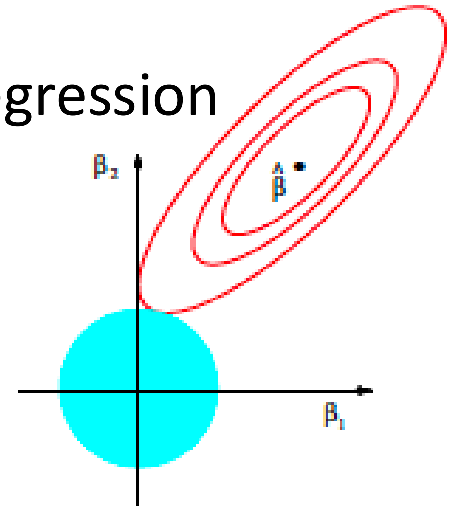


- when traits are correlated it can be hard to estimate their independent effects on fitness
- leads to large standard errors and inaccurate estimates of selection
- dropping correlated traits from analysis is a common solution

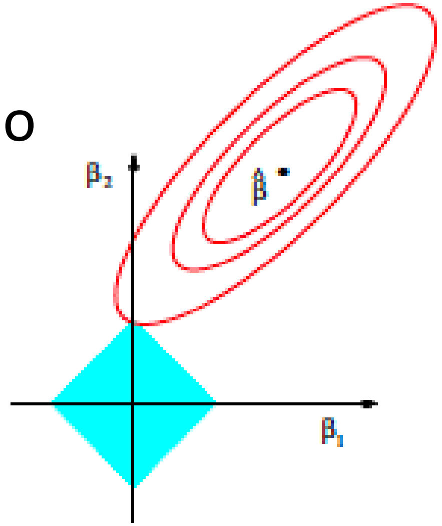
Regularized regression

- Linear regression with a penalty added
- Coefficients are constrained to be within a certain space

Ridge regression



Lasso



$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right\}.$$

$$\hat{\beta}^{\text{lasso}} = \underset{\beta}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^{N} \Big(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \Big)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}.$$

Regularized regression for estimating selection

- simulation study to compare the accuracy of Lande-Arnold regression to regularized regression for estimates selection

- re-analysed published studies of selection using regularized regression

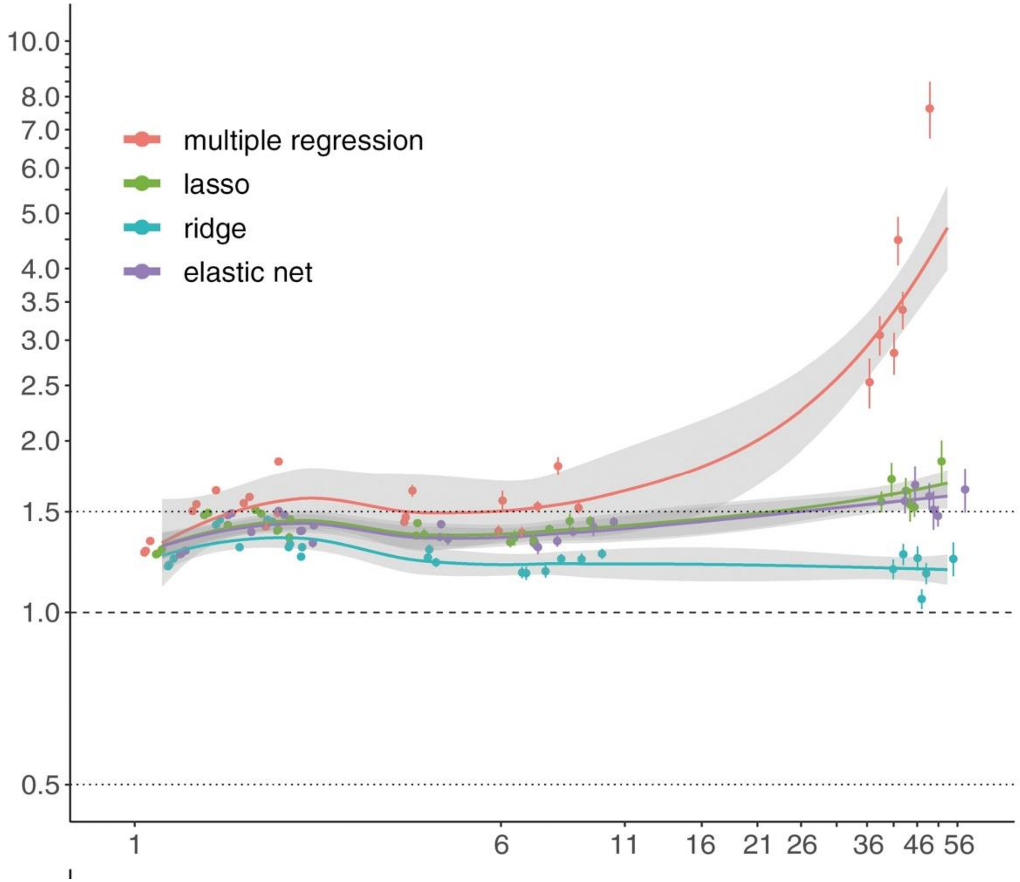
Traits	Sample Size	Variance Inflation Factor	Traits under selection	Fitness distribution
4	100	low	all	binomial
7	400	mid/high	up to half	poisson
12	1000			
17				

Sztepanacz and Houle (2024) *Evolution Letters* 8: 361-373

More accurate estimates of the total strength of selection

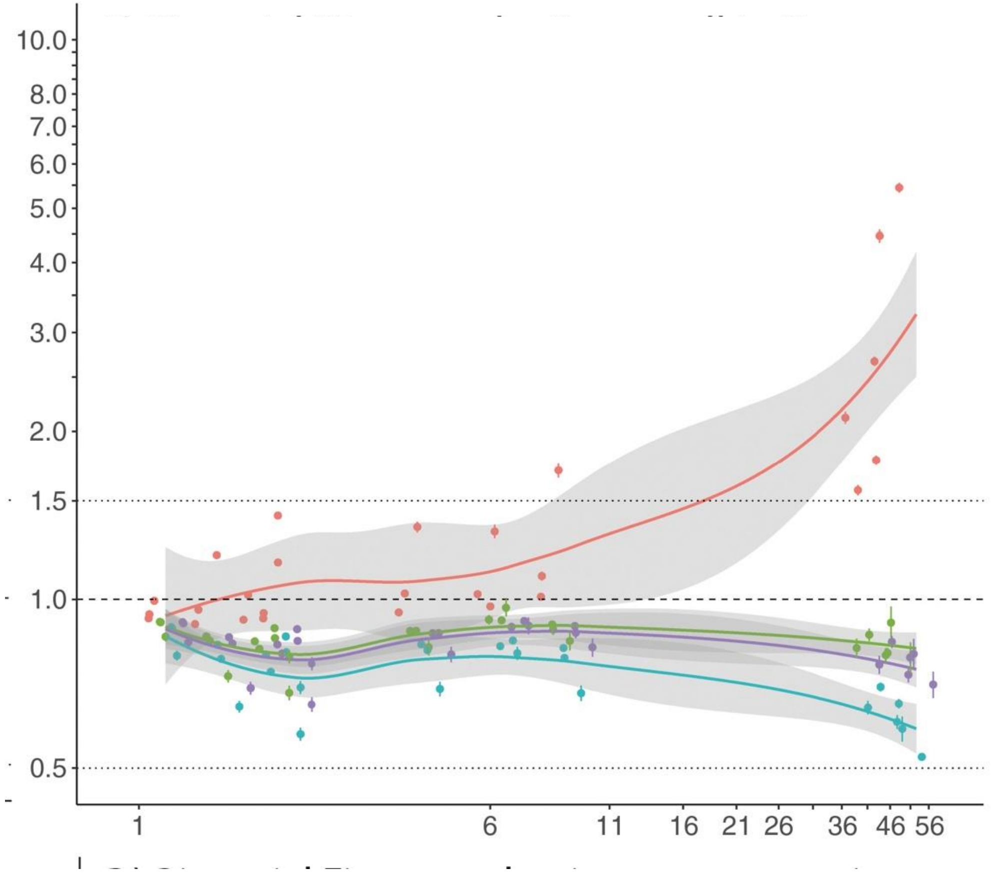
$$\frac{\|\hat{\beta}\|}{\|\beta\|}$$

Poisson fitness (seed set)



Variance inflation factor

Binomial fitness (mating success)

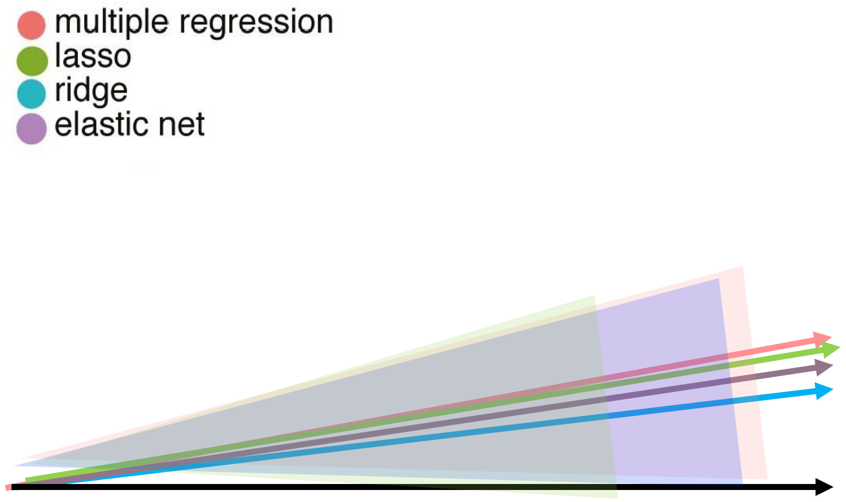


Variance inflation factor

Sztepanacz and Houle (2024) *Evolution Letters* 8: 361-373

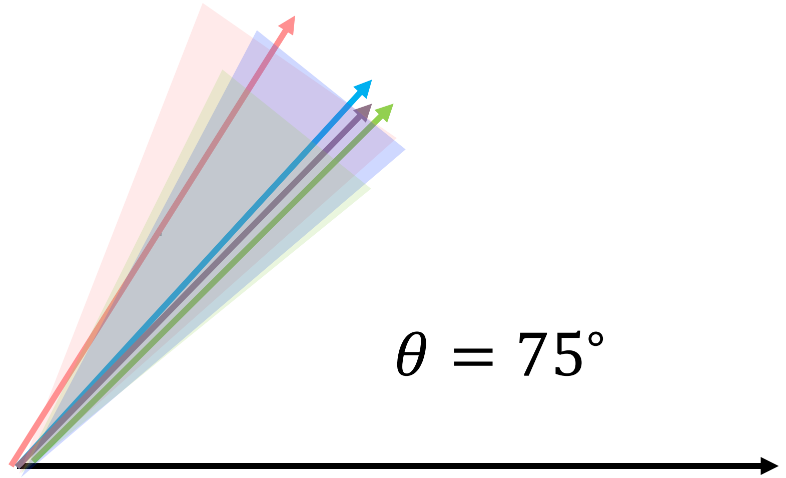
Does not improve estimates of the direction of selection

4 traits low multicollinearity (VIF ~1)



$\theta = 12°$

17 traits high multicollinearity (VIF ~56)



$\theta = 75°$

Sztepanacz and Houle (2024) *Evolution Letters* 8: 361-373

Arabidopsis phenology

Estimated selection on *Arabidopsis* phenology from Chong et al., (2018)

Method	OLS regression	PC regression (4 PCs)	Lasso	Ridge
Coefficients (+/− SE)				
Flowering time	−0.299 (0.191)	−0.181	−0.371	−0.143
Flowering duration	0.050 (0.182)	0.186	–	0.109
Branch number	0.080 (0.085)	0.034	–	0.083
Rosette diameter	0.059 (0.068)	0.072	–	0.072
Rosette leaf number	0.061 (0.097)	0.033	–	−0.006
R^2 –	0.55 [0.007, 0.939]	–	0.680 [0.081, 0.997]	0.964 [0.838, 0.997]

Sztepanacz and Houle (2024) *Evolution Letters* 8: 361-373

Stabilizing selection

- even though we often focus on directional selection, stabilizing selection is the most common form of selection operating in populations

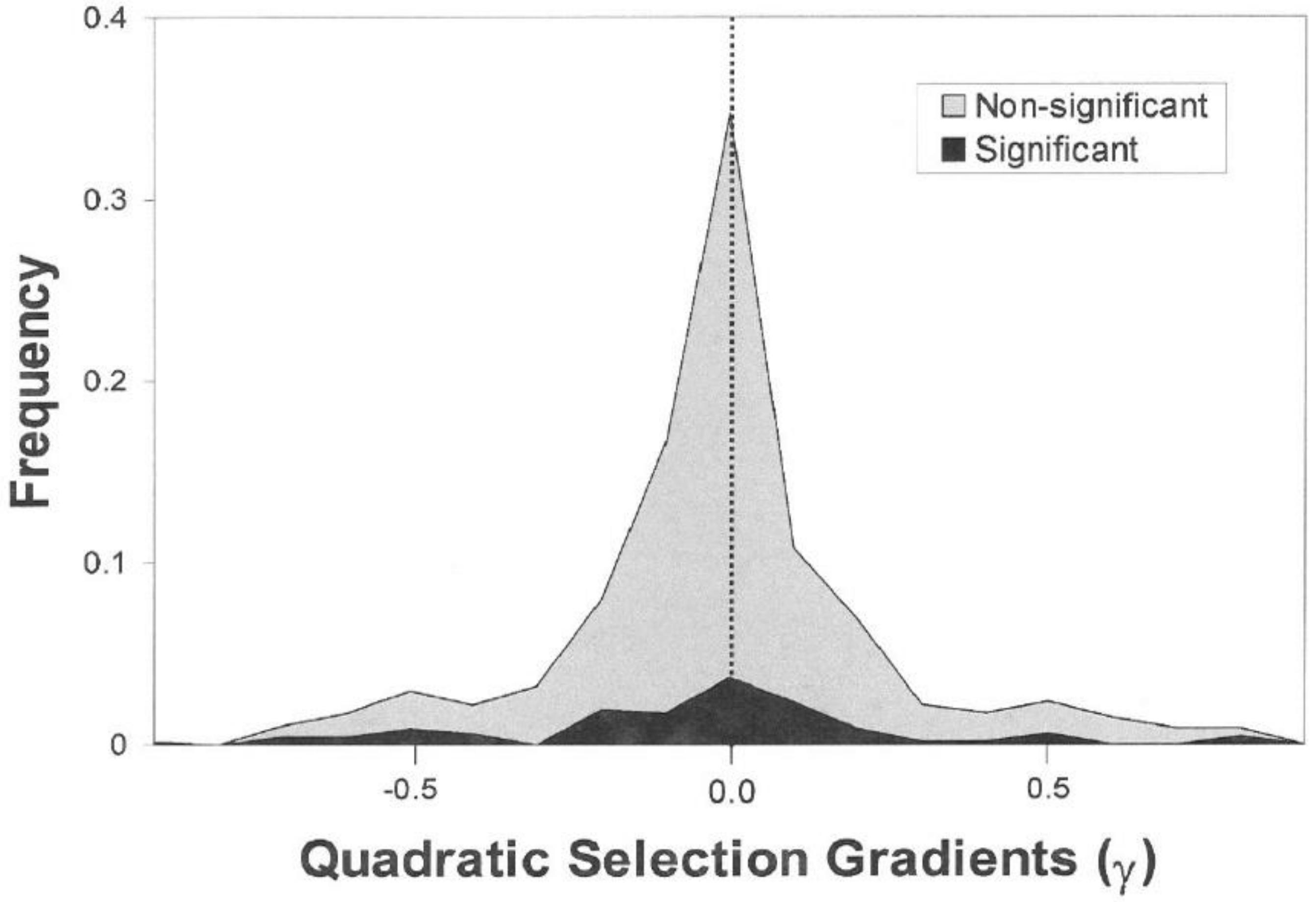
- stabilizing selection reduces the variance in the population

- disruptive selection increases the variance (eg. selection for extremes, negative frequency dependent selection)

Quadratic regression to estimate correlational and quadratic selection

$$\bar{w} \sim \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n + \frac{1}{2}\gamma_1 x_1^2 + \frac{1}{2}\gamma_2 x_2^2 + \gamma_{12} x_1\, x_2 \ldots.$$

Quadratic regression to estimate correlational and quadratic selection



- disruptive selection appears to be as common as stabilizing selection

Quadratic regression to estimate correlational and quadratic selection

Table 9.1 The vector of standardized linear gradients (β) and the matrix of standardized quadratic and correlational selection gradients (γ) for male call structure components in *T. commodus*. Note that the quadratic gradients were not doubled in the original study but have been done so here. The significance of selection gradients were tested using permutation test with 9999 iterations (see text): *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

		γ				
	β	CPN	CIPD	TN	ICD	DF
CPN	0.007	0.012				
CIPD	–0.003	0.017	–0.012			
TN	0.015	0.019	0.039	–0.080		
ICD	–0.214***	–0.022	–0.036	0.086*	–0.160**	
DF	0.059	0.024	–0.031	0.041	–0.013	–0.094*

Brooks R, Hunt J, Blows MW, Smith MJ, Bussière LF, Jennions MD. Experimental evidence for multivariate stabilizing sexual selection. Evolution. 2005 Apr;59(4):871-80. PMID: 15926696.

Canonical rotation of correlational and quadratic selection

Table 9.2 The **M** matrix of eigenvectors derived from the canonical analysis of γ. The linear (θ_i) and quadratic (λ_i) gradients of selection acting along these eigenvectors are provided in the last two columns. We tested the significance of selection along these eigenvectors using a permutation test based on the double regression method of (Bisgaard and Ankenman 1996) but note that the results were qualitatively the same when the permutation test of Reynolds et al. (2010) was used. Randomization test: *P <0.05, **P <0.01, ***P <0.001

	M					Selection	
	CPN	CIPD	TN	ICD	DF	θ_i	λ_i
m_1	0.800	0.497	0.305	−0.057	0.130	0.028	0.035
m_2	−0.446	0.806	−0.001	−0.091	−0.377	−0.017	−0.003
m_3	−0.302	0.003	0.776	0.500	0.240	−0.082	−0.019
m_4	−0.257	0.208	−0.102	−0.405	0.846	0.144***	−0.108**
m_5	0.068	0.244	−0.543	0.758	0.258	−0.160***	−0.240***

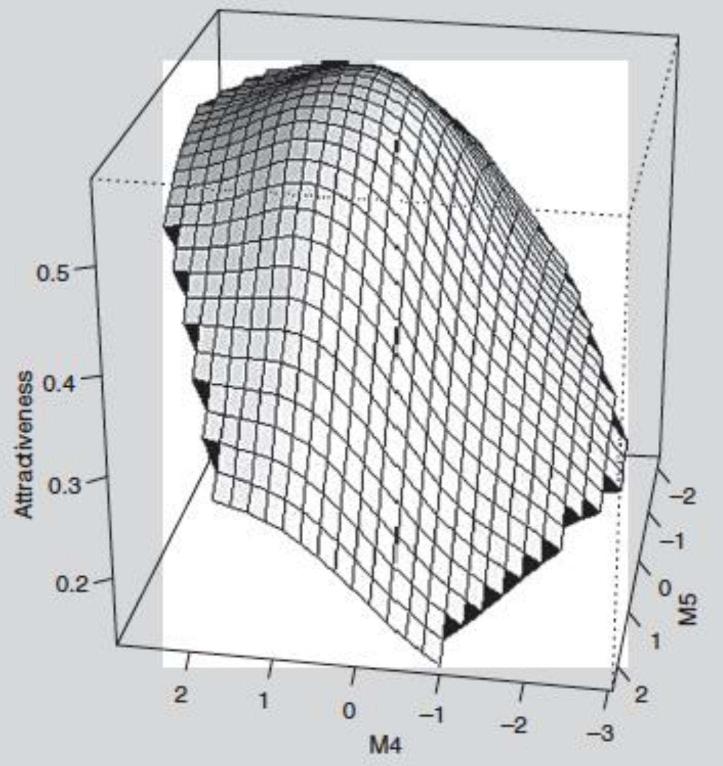
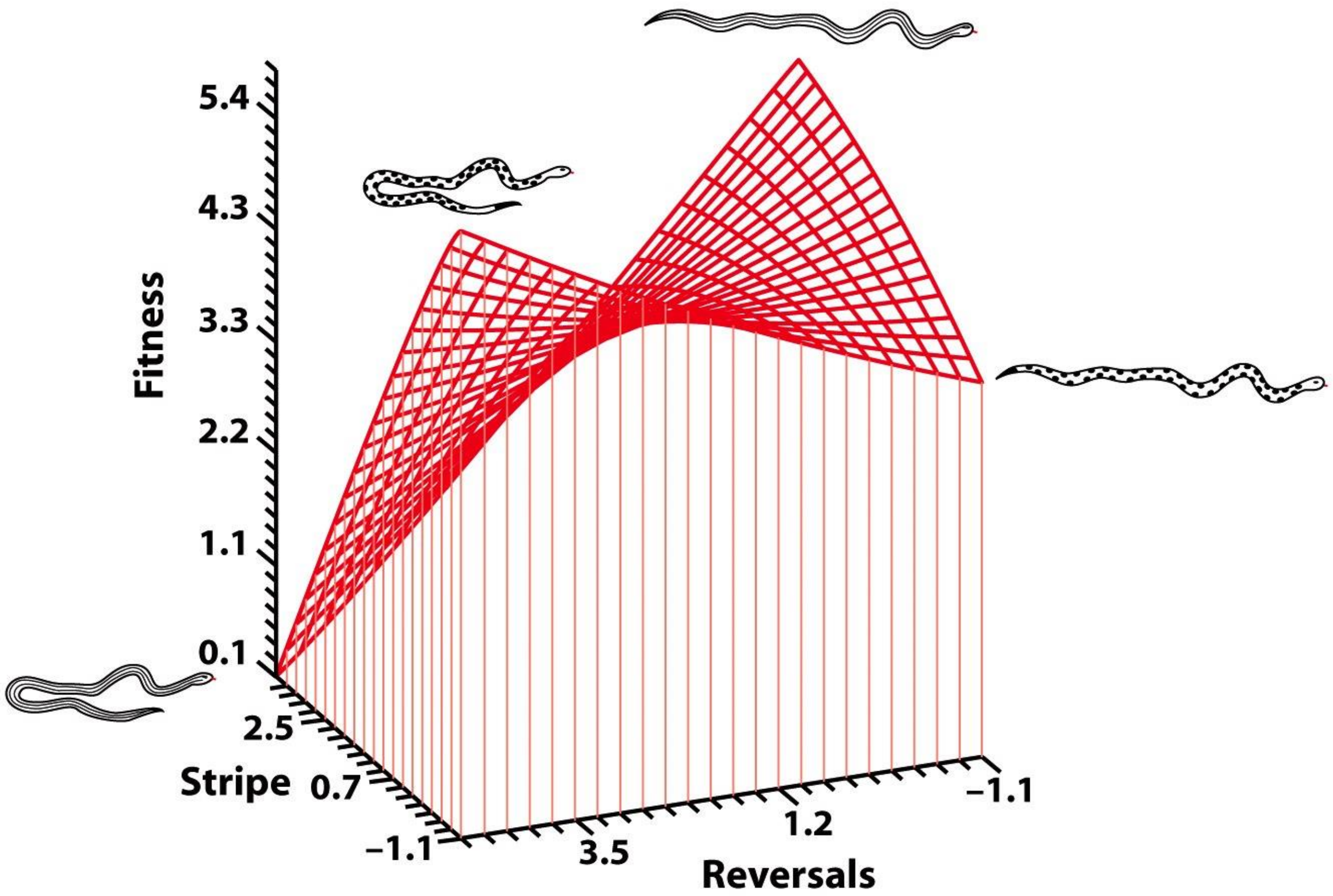


Figure 9.3 Thin-plate spline perspective view visualization of the fitness surface on the two major axes of non-linear selection, m_4 and m_5. The thin-plate spline was estimated using the *Tps* function in R (version 9.2.1), using the value of the smoothing parameter, λ, that minimized the GCV score. Redrawn from Brooks et al. (2005), with permission from John Wiley and Sons.

Brooks R, Hunt J, Blows MW, Smith MJ, Bussière LF, Jennions MD. Experimental evidence for multivariate stabilizing sexual selection. Evolution. 2005 Apr;59(4):871-80. PMID: 15926696.

Multivariate saddles may be common



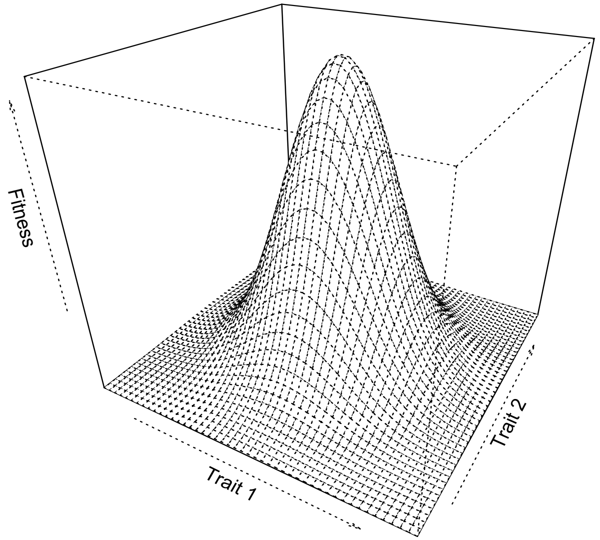
Brodie III, E.D., 1992. Correlational selection for color pattern and antipredator behavior in the garter snake Thamnophis ordinoides. *Evolution*, *46*(5), pp.1284-1298.

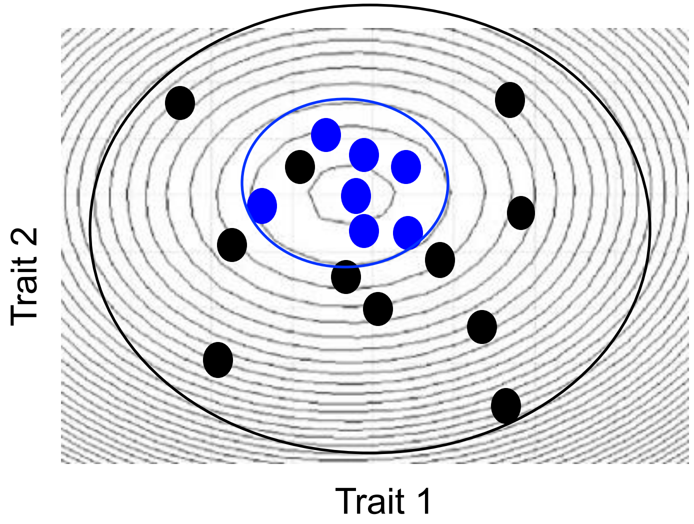
Pleiotropic model of MSB

- alleles affect a focal trait and their pleiotropic effects on other traits are condensed into an effect on net fitness

- mutations may increase or decrease the value of a focal trait, but their effects on fitness are almost certainly deleterious

- individuals with more extreme values of a focal trait will tend to carry alleles that have deleterious effects with respect to net fitness

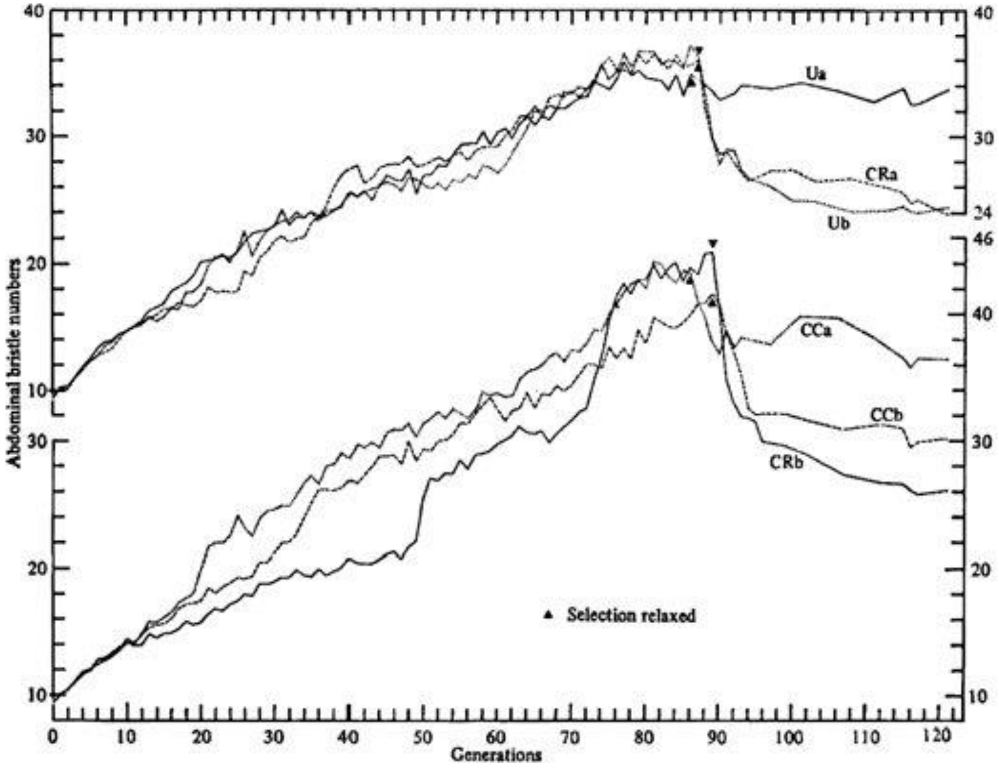


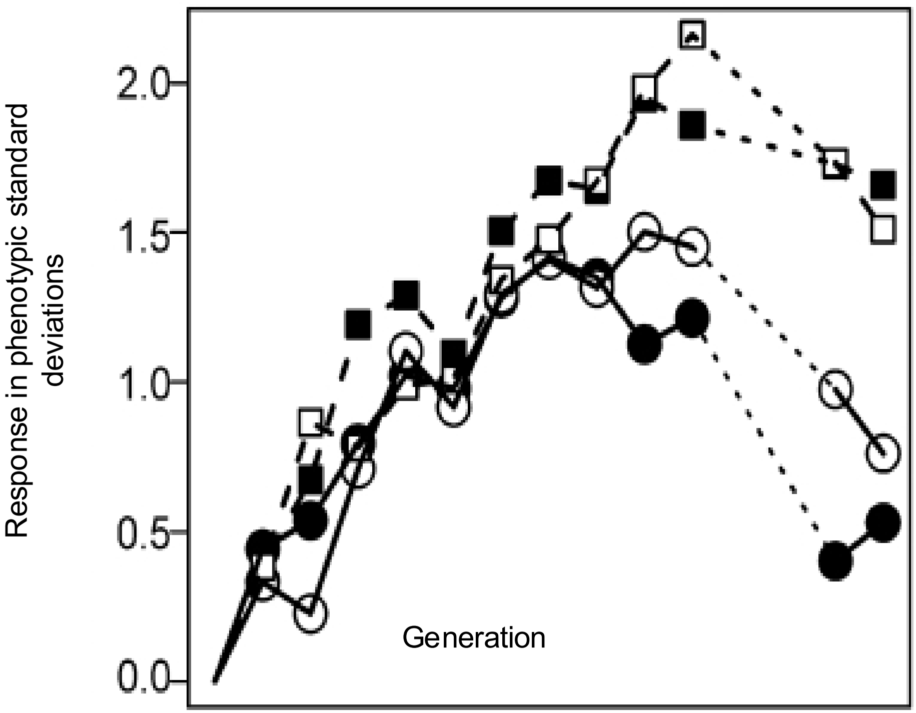


Hill & Keightley, 1990

Artificial selection experiments

- Evolutionary limits are also indicative of pleiotropic effects on fitness

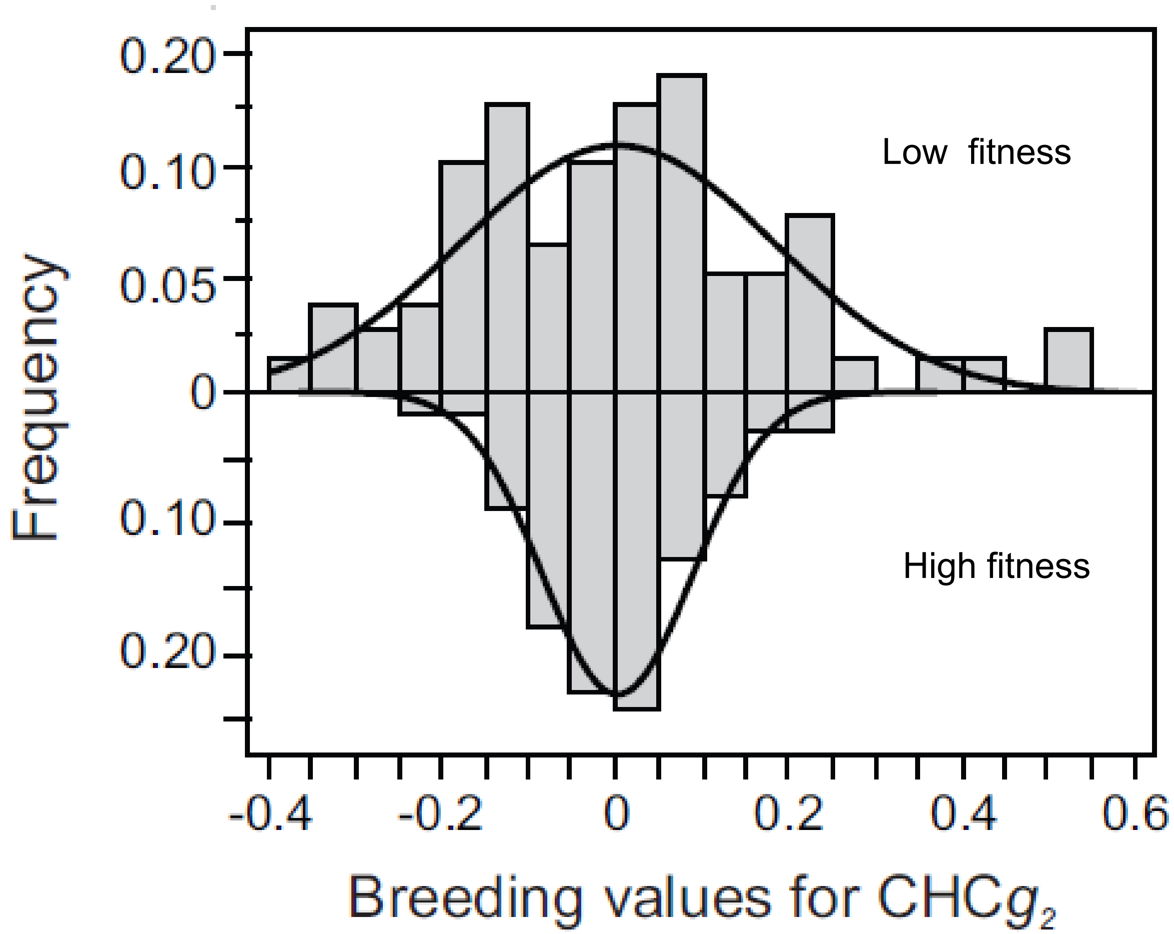




Yoo, 1980

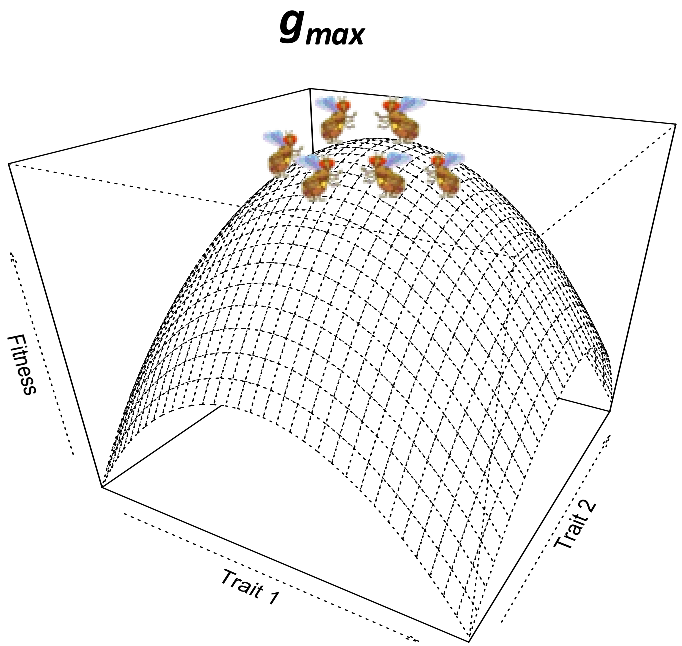
Apparent stabilising selection

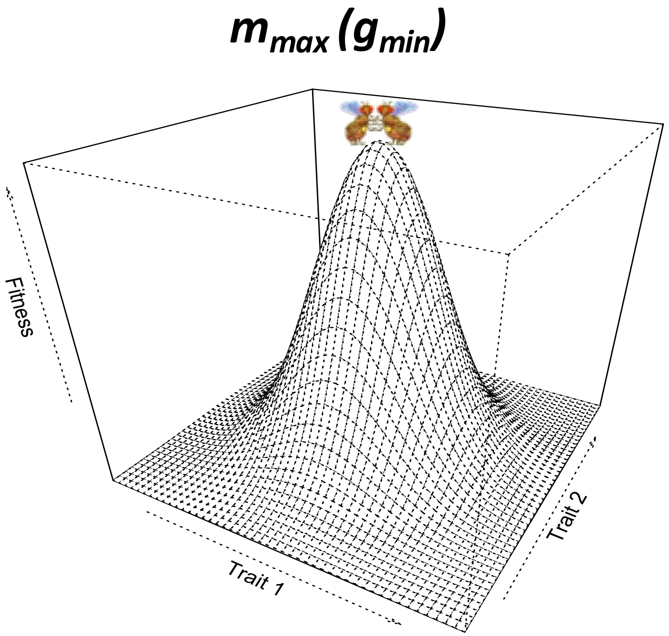
One example



Sztepanacz J.L., and Rundle, H.D. (2012). Reduced genetic variance among high fitness individuals: inferring stabilizing selection on male sexual displays in *Drosophila serrata*. ***Evolution.*** 66(10): 3101-3110

Apparent stabilising selection

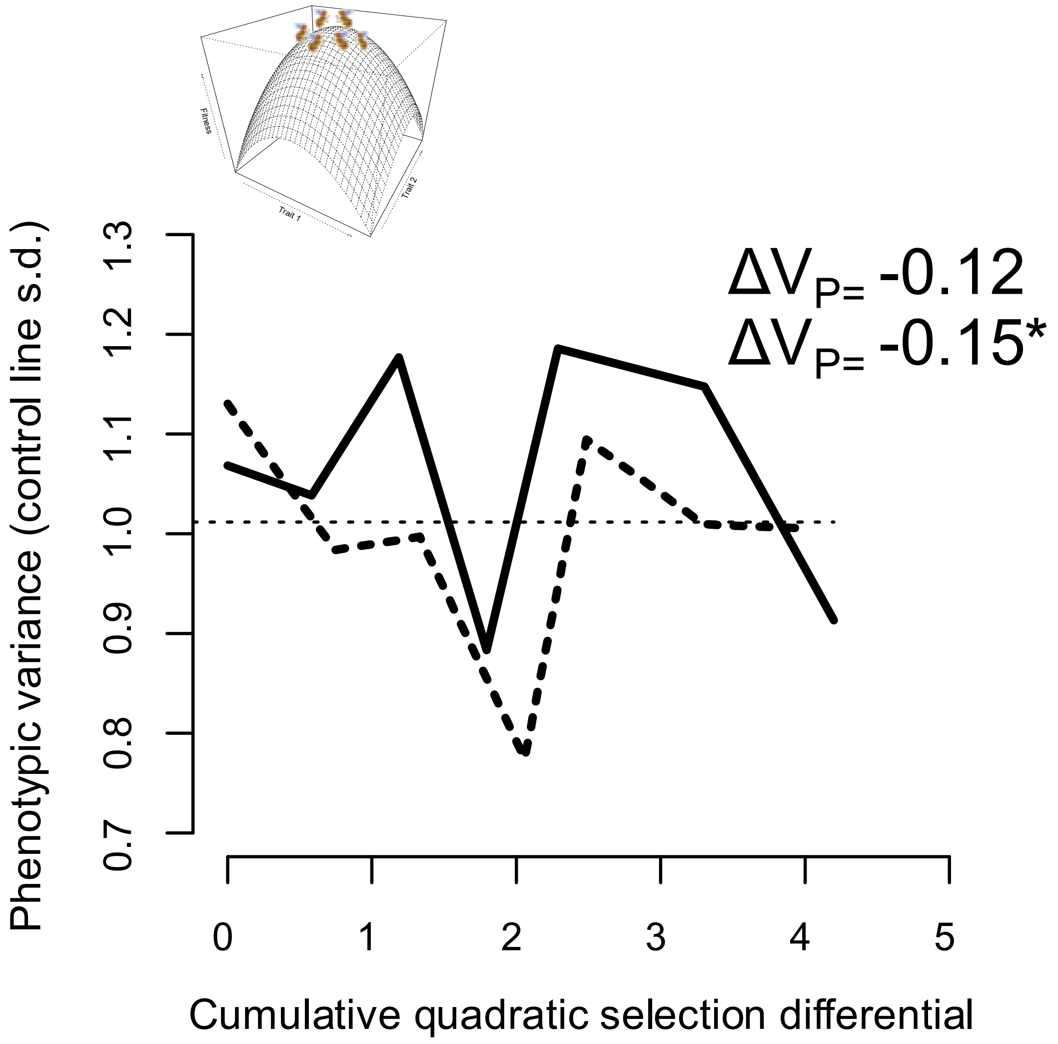


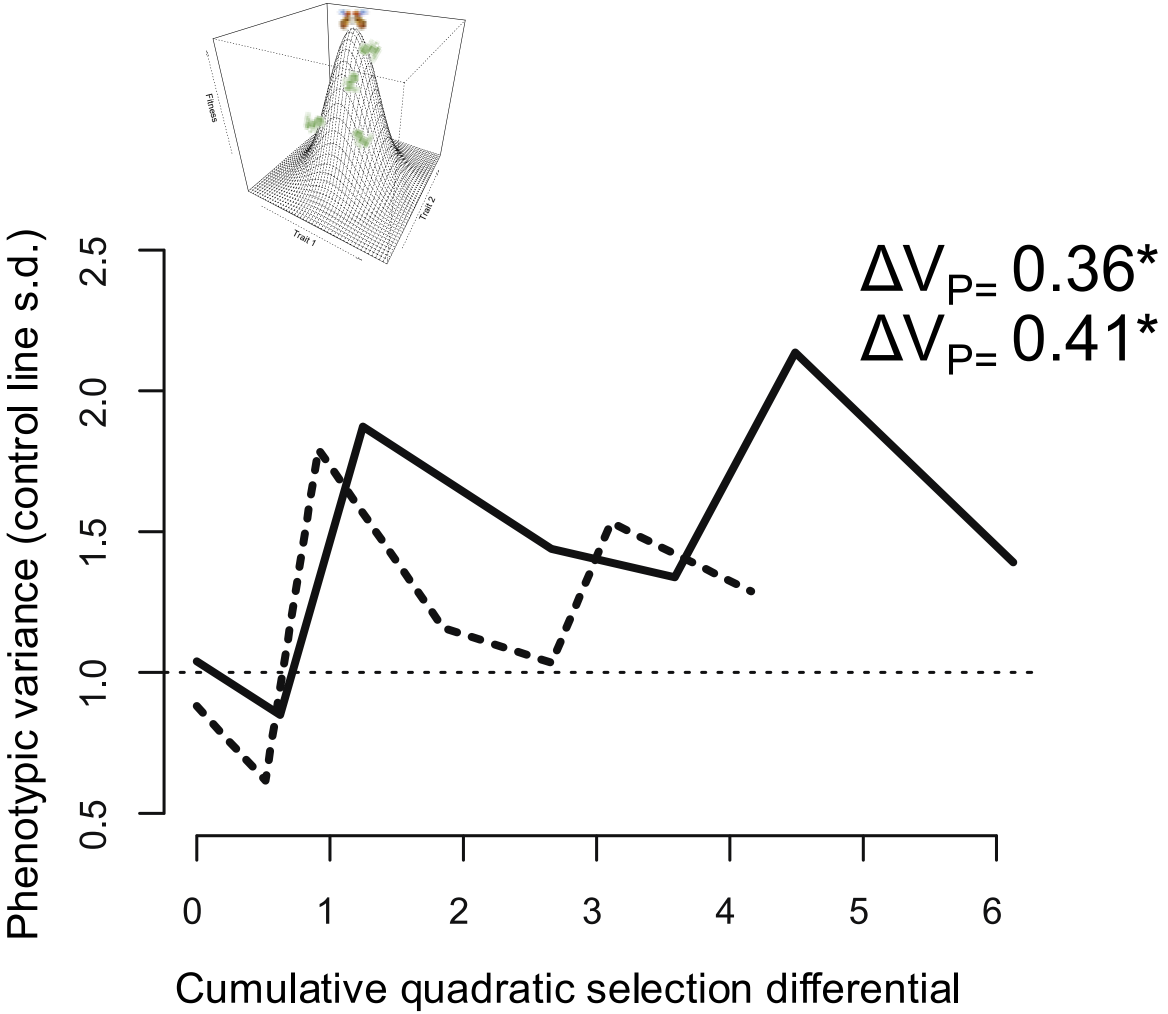


- persistence time of new mutations: 6753 generations
- persistence time of new mutations: 562 generations

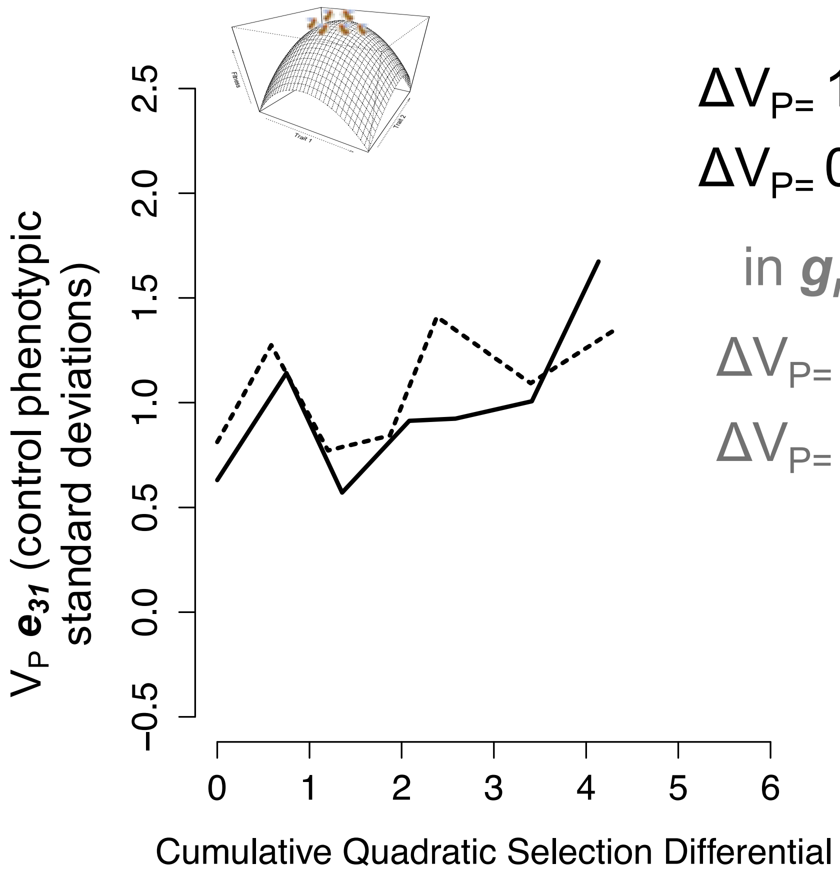
Sztepanacz, J.L. and Mark. W. Blows. (2017). [Artificial selection to increase the phenotypic variance in g_{max} fails.] ***The American Naturalist.*** 190(5): 707-723

Apparent stabilising selection





Correlated responses to selection were larger than responses on targeted traits



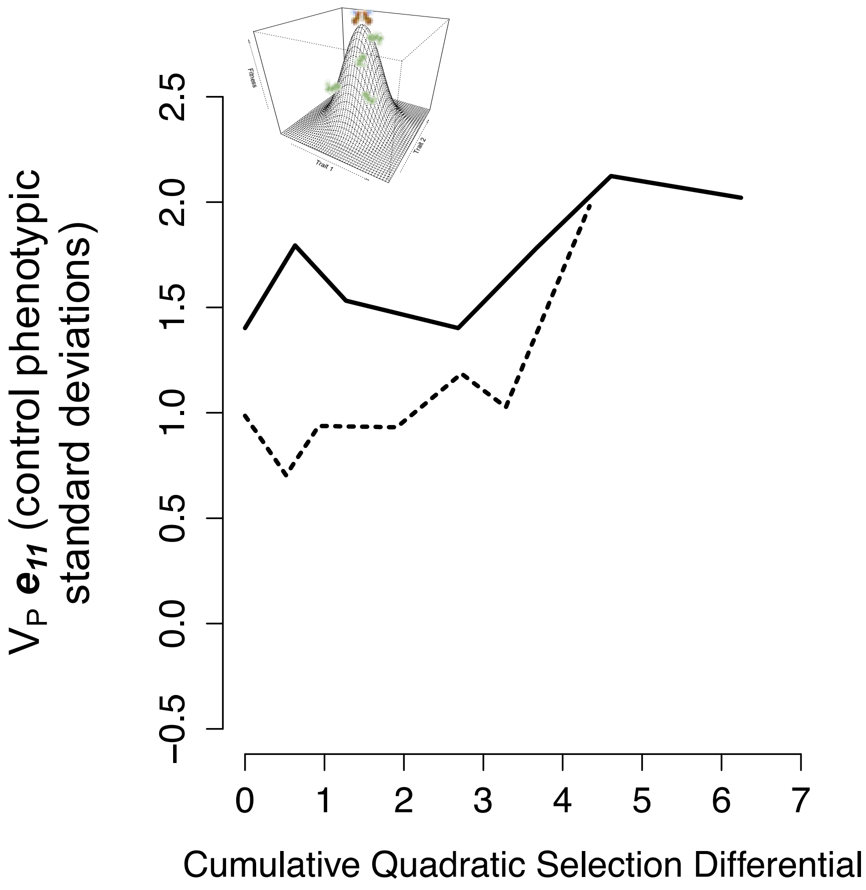
$\Delta V_{P=}$ 1.044*

$\Delta V_{P=}$ 0.533*

in $\boldsymbol{g_{max}}$

$\Delta V_{P=}$ -0.12

$\Delta V_{P=}$ -0.15*



$\Delta V_{P=}$ 0.618

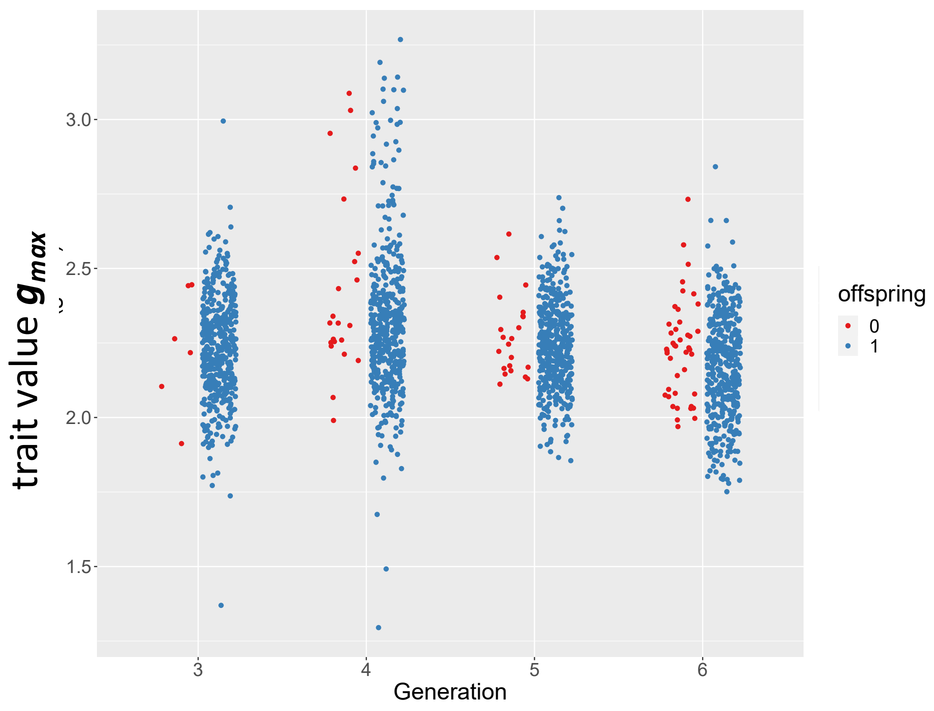
$\Delta V_{P=}$ 0.994*

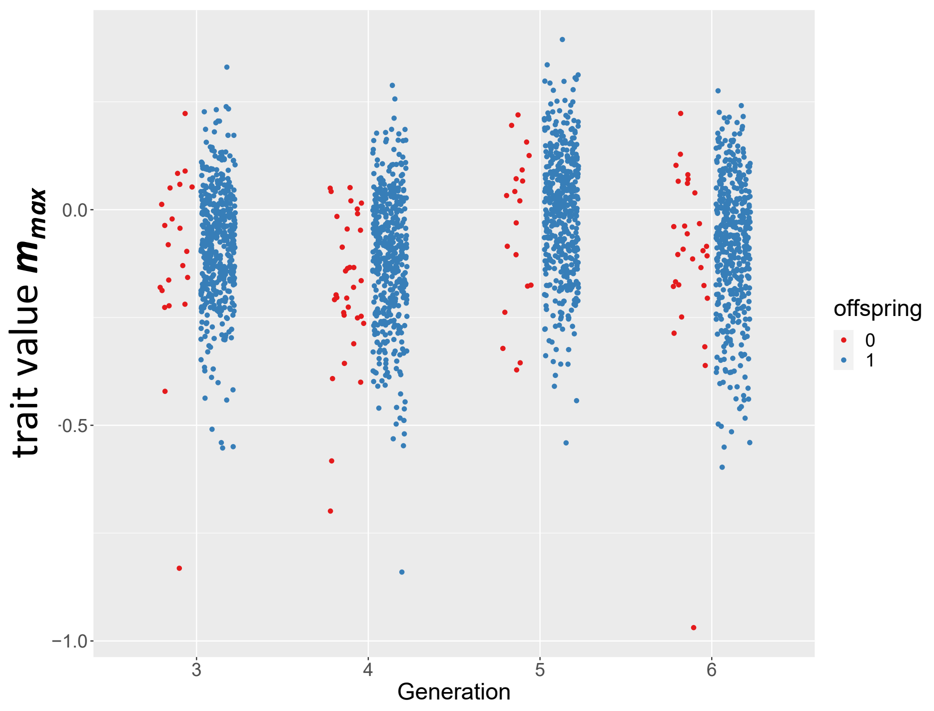
in $\boldsymbol{m_{max}}$

$\Delta V_{P=}$ 0.36*

$\Delta V_{P=}$ 0.41*

Correlated selection against extreme phenotypes





- significant quadratic (stabilizing) selection via male reproductive failure in both g_{max} and m_{max} populations
- the multivariate trait combination under significant stabilizing selection via male reproductive failure was not g_{max} or m_{max}

Key take-aways:

- genetic variation is unevenly distributed across multivariate trait combinations because of pleiotropy
- the uneven distribution of genetic variance can lead to evolutionary constraints
- we can estimate selection on multiple traits using linear or quadratic regression approaches
- Most stabilizing selection may arise through the pleiotropic effects of alleles on multiple traits

Questions?