Lecture 9
Short-Term Selection
Response: Breeder's equation

Bruce Walsh lecture notes
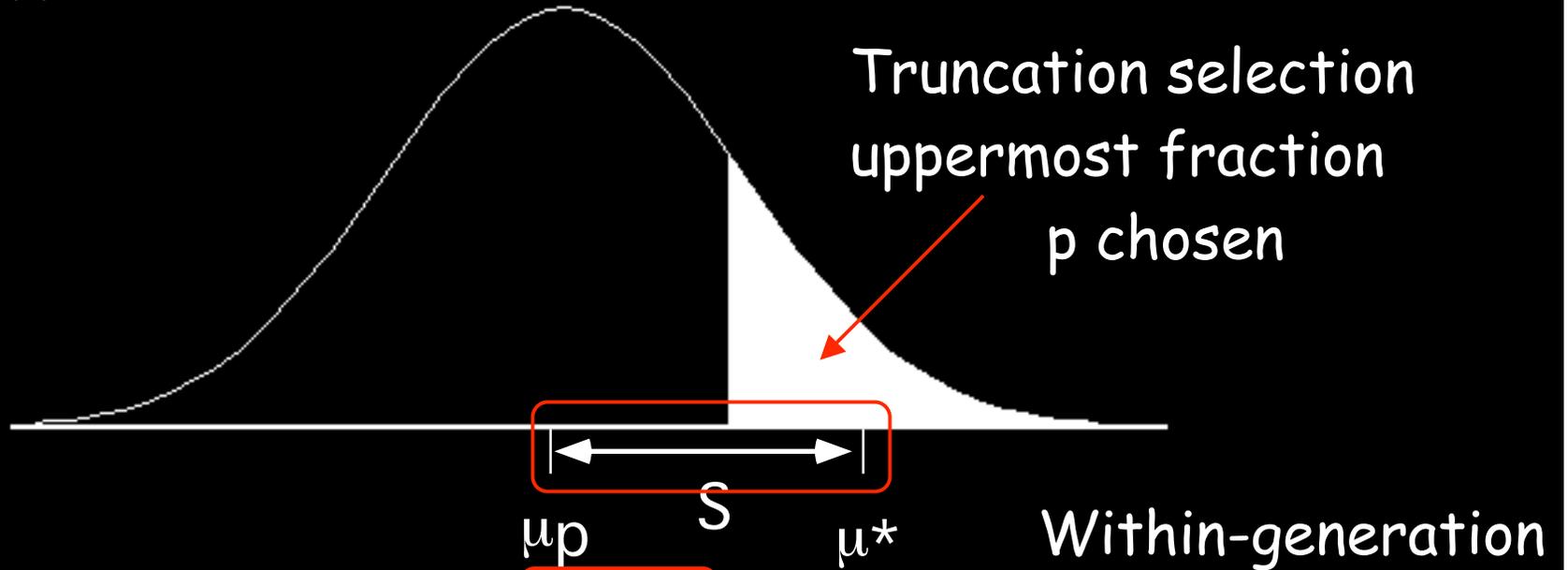Synbreed course
version 3 July 2013

Response to Selection

- Selection can change the distribution of phenotypes, and we typically measure this by changes in mean
  - This is a within-generation change
- Selection can also change the distribution of breeding values
  - This is the response to selection, the change in the trait in the next generation (the between-generation change)

The Selection Differential and the Response to Selection

- The selection differential S measures the within-generation change in the mean

  - $S = \mu^* - \mu$

- The response R is the between-generation change in the mean

  - $R(t) = \mu(t+1) - \mu(t)$

(A) Parental Generation

Truncation selection uppermost fraction p chosen

S

μp $\mu*$

Within-generation change

(B) Offspring Generation

R

Between-generation change

μO
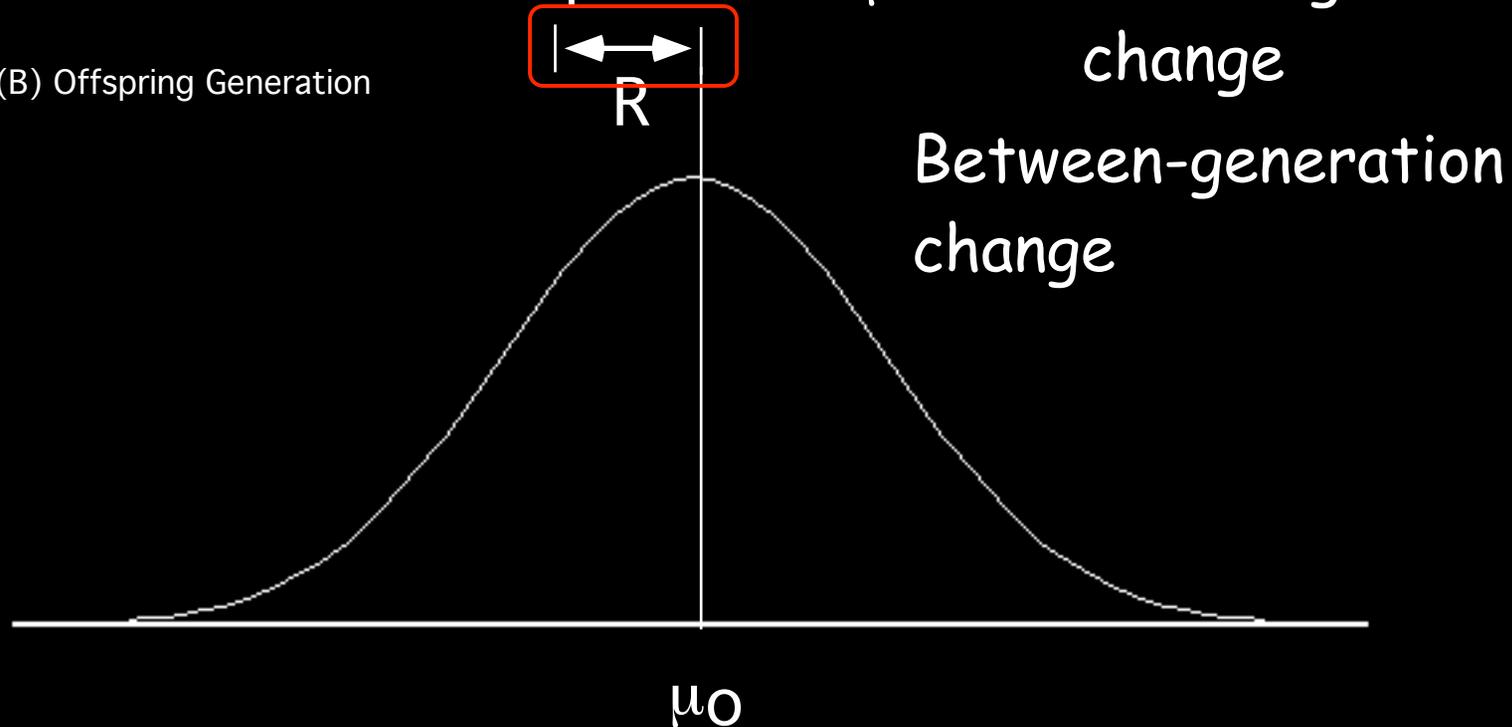
The Breeders' Equation: Translating S into R

Recall the regression of offspring value on midparent value

$$y_O = \mu_P + h^2 \left(\frac{P_f + P_m}{2} - \mu_P \right)$$

Averaging over the selected midparents,
$$E[\ (P_f + P_m)/2\ ] = \mu*,$$

Likewise, averaging over the regression gives
$$E[\ y_o - \mu\ ] = h^2\ (\ \mu* - \mu\ ) = h^2\ S$$

Since $E[\ y_o - \mu\ ]$ is the change in the offspring mean, it represents the response to selection, giving:

$$\boxed{R = h^2\ S}$$ The Breeders' Equation (Jay Lush)

- Note that no matter how strong S, if h^2 is small, the response is small

- S is a measure of selection, R the actual response. One can get lots of selection but no response

- If offspring are asexual clones of their parents, the breeders' equation becomes
  - $R = H^2 S$

- If males and females subjected to differing amounts of selection,
  - $S = (S_f + S_m)/2$
  - Example: Selection on seed number in plants -- pollination (males) is random, so that $S = S_f/2$

Pollen control

- Recall that $S = (S_f + S_m)/2$

- An issue that arises in plant breeding is pollen control --- is the pollen from plants that have also been selected?

- Not the case for traits (i.e., yield) scored after pollination. In this case, $S_m = 0$, so response only half that with pollen control

- Tradeoff: with an additional generation, a number of schemes can give pollen control, and hence twice the response

  - However, takes twice as many generations, so response per generation the same

Selection on clones

- Although we have framed response in an outcrossed population, we can also consider selecting the best individual clones from a large population of different clones (e.g., inbred lines)

- $R = H^2S$, now a function of the board sense heritability. Since $H^2 \geq h^2$, the single-generation response using clones exceeds that using outcrossed individuals

- However, the genetic variation in the next generation is significantly reduced, reducing response in subsequent generations
  - In contrast, expect an almost continual response for several generations in an outcrossed population.

8

Price-Robertson identity

- $S = \text{cov}(w, z)$

- The covariance between trait value z and relative fitness (w = W/Wbar, scaled to have mean fitness = 1)

- VERY! Useful result

- $R = \text{cov}(w, A_z)$, as response = within generation change in BV

  - This is called Robertson's secondary theorem of natural selection

Correcting for Reproductive Differences: Effective Selection Differentials

In artificial selection experiments, S is usually estimated as the difference between the mean of the selected adults and the sample mean of the population before selection. Selection need not stop at this stage. For example, strong artificial selection to increase a character might be countered by natural selection due to a decrease in the fertility of individuals with extreme character values. Biases introduced by such differential fertility can be removed by randomly choosing the same number of offspring from each selected parent, ensuring equal fertility.

Alternatively, biases introduced by differential fertility can be accounted for by using **effective selection differentials**, S_e,

$$S_e = \frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{n_i}{\overline{n}} \right) (z_i - \mu_z) \tag{10.8}$$

where z_i and n_i are the phenotypic value and total number of offspring of the ith parent, n_p the number of parents selected to reproduce, $\overline{n}$ the average number of offspring for selected parents, and μ_z is the mean before selection. If all selected parents have the same number of offspring ($n_i = \overline{n}$ for all i), then S_e reduces to S. However, if there is variation in the number of offspring n_i among selected parents, S_e can be considerably different from S. This corrected differential is also referred to as the **realized selection differential**.

Consider 5 individuals, z_i = trait value,
n_i = number of offspring

i	z_i	n_i	$n_i/\overline{n}$
1	45	1	0.3125
2	40	2	0.6250
3	35	3	0.9375
4	33	5	1.563
5	32	5	1.563

$$\frac{1}{n_p} \sum_{i=1}^{n_p} \left(\frac{n_i}{\overline{n}} \right) z_i = 34.69$$

Hence, $S_e = 4.69$, for an expected response of $R = 0.3 \cdot 4.69 = 1.4$. In this case, not using the effective differential results in an overestimation of the expected response.

Unweighted S = 7, offspring-weighted S = 4.69

Response over multiple generations

- Strictly speaking, the breeders' equation only holds for predicting a single generation of response from an unselected base population

- Practically speaking, the breeders' equation is usually pretty good for 5-10 generations

- The validity for an initial h^2 predicting response over several generations depends on:
  - The reliability of the initial h^2 estimate
  - Absence of environmental change between generations
  - The absence of genetic change between the generation in which h^2 was estimated and the generation in which selection is applied

The selection differential is a function of both the phenotypic variance and the fraction selected

50% selected
$V_p = 4$, $S = 1.6$

20% selected
$V_p = 4$, $S = 2.8$

20% selected
$V_p = 1$, $S = 1.4$

(A)

(B)

(C)

S

S

S
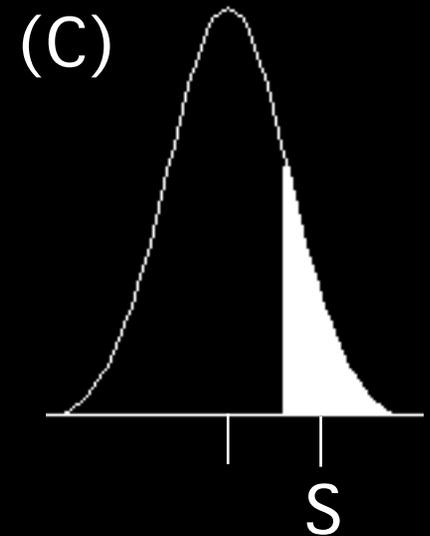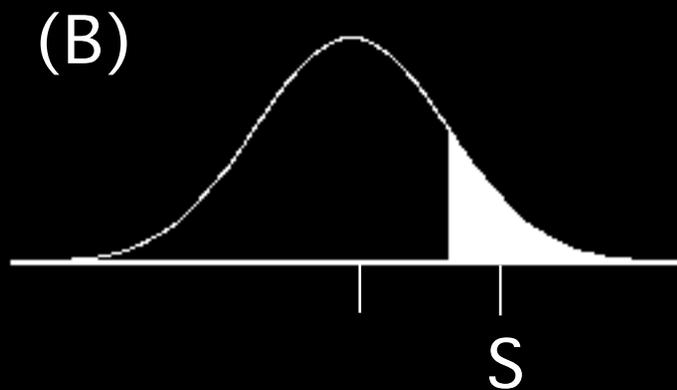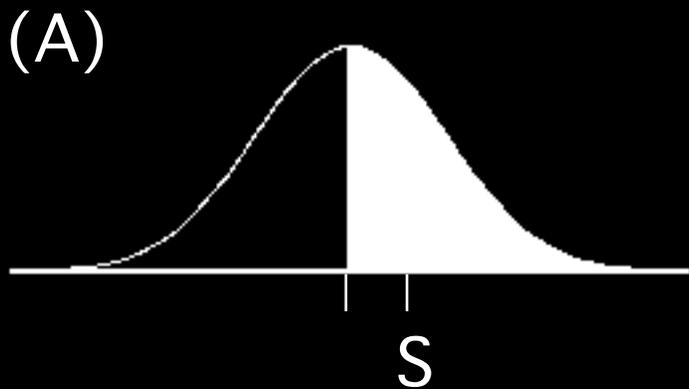
The Selection Intensity, i

As the previous example shows, populations with the same selection differential (S) may experience very different amounts of selection

The selection intensity i provides a suitable measure for comparisons between populations,
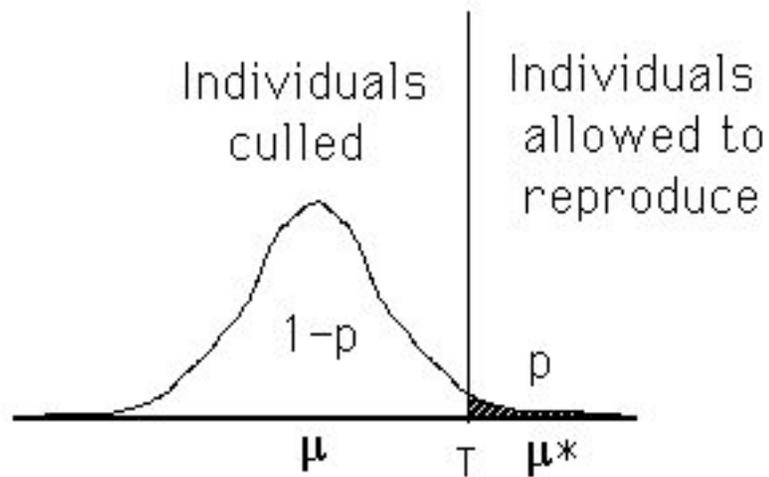
$$i = \frac{S}{\sqrt{V_P}} = \frac{S}{\sigma_p}$$

Truncation selection

- A common method of artificial selection is truncation selection --- all individuals whose trait value is above some threshold (T) are chosen.

- Equivalent to only choosing the uppermost fraction p of the population

Individuals culled | Individuals allowed to reproduce

$1-p$ p

μ T $\mu*$
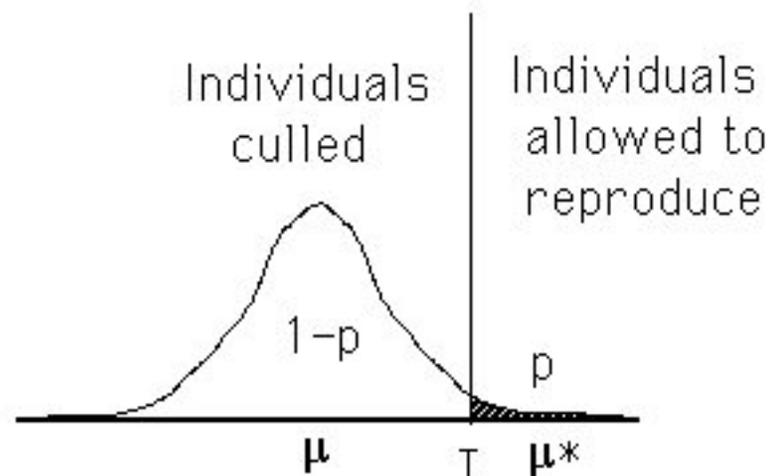
Selection Differential Under Truncation Selection

$$S = \mu^* - \mu$$

$$S = \varphi\left(\frac{T - \mu}{\sigma}\right)\frac{\sigma}{p}$$

Likewise,

$$\bar{i} = \frac{S}{\sigma} = \frac{\varphi(z_{[1-p]})}{p}$$

R code for i: `dnorm(qnorm(1-p))/p`

Truncation selection

- The fraction p saved can be translated into an expected selection intensity (assuming the trait is normally distributed),

  - allows a breeder (by setting p in advance) to chose an expected value of i before selection, and hence set the expected response

$$\bar{i} = \frac{S}{\sigma} = \frac{\varphi(z_{[1-p]})}{p}$$

  ⟵······ Height of a unit normal at the threshold value corresponding to p

  - R code for i: `dnorm(qnorm(1-p))/p`

p	0.5	0.2	0.1	0.05	0.01	0.005
i	0.798	1.400	1.755	2.063	2.665	2.892

Selection Intensity Version of the Breeders' Equation

$$R = h^2 S = h^2 \frac{S}{\sigma_p} \sigma_p = i\, h^2\, \sigma_p$$

Since $h^2\sigma_P = (\sigma^2_A/\sigma^2_P)\,\sigma_P = \sigma_A(\sigma_A/\sigma_P) = h\,\sigma_A$

$$R = i\, h\, \sigma_A$$

Since h = correlation between phenotypic and breeding values, $h = r_{PA}$

$$R = i\, r_{PA}\sigma_A$$

Response = Intensity * Accuracy * spread in Va

When we select an individual solely on their phenotype, the accuracy (correlation) between BV and phenotype is h

Accuracy of selection

More generally, we can express the breeders equation as

$$R = i \, r_{uA} \sigma_A$$

Where we select individuals based on the index u (for example, the mean of n of their sibs).

r_{uA} = the accuracy of using the measure u to predict an individual's breeding value = correlation between u and an individual's BV, A

Example 10.4. **Progeny testing**, using the mean of a parent's offspring to predict the parent's breeding value, is an alternative predictor of an individual's breeding value. In this case, the correlation between the mean x of n offspring and the breeding value A of the parent is

$$\rho(x, A) = \sqrt{\frac{n}{n + a}}, \quad \text{where} \quad a = \frac{4 - h^2}{h^2}$$

From Equation 10.11, the response to selection under progeny testing is

$$R = i\sigma_A \sqrt{\frac{n}{n + a}} = i\sigma_A \sqrt{\frac{h^2 n}{4 + h^2(n - 1)}}$$

Note that for very large n that the accuracy approaches one. Progeny testing gives a larger response than simple selection on the phenotypes of the parents (**mass selection**) when

$$\sqrt{\frac{n}{4 + h^2(n - 1)}} > 1, \quad \text{or} \quad n > \frac{4 - h^2}{1 - h^2}$$

In particular, $n > 4$, 5, and 7, for $h^2 = 0.1$, 0.25, and 0.5. Also note that the ratio of response for progeny testing (R_{pt}) to mass selection (R_{ms}) is just

$$\frac{R_{pt}}{R_{ms}} = \frac{1}{h}\sqrt{\frac{h^2 n}{4 + h^2(n - 1)}} = \sqrt{\frac{n}{4 + h^2(n - 1)}}$$

which approaches $1/h$ for large n.

Improving accuracy

- Predicting either the breeding or genotypic value from a single individual often has low accuracy --- h^2 and/or H^2 (based on a single individuals) is small

  - Especially true for many plant traits with high G x E

  - Need to replicate either clones or relatives (such as sibs) over regions and years to reduce the impact of G x E

  - Likewise, information from a set of relatives can give much higher accuracy than the measurement of a single individual

Stratified mass selection

- In order to accommodate the high environmental variance with individual plant values, Gardner (1961) proposed the method of stratified mass selection
    - Population stratified into a number of different blocks (i.e., sections within a field)
    - The best fraction p within each block are chosen
    - Idea is that environmental values are more similar among individuals within each block, increasing trait heritability.

Overlapping Generations

L_x = Generation interval for sex x

    = Average age of parents when progeny are born

The yearly rate of response is

$$R_y = \frac{i_m + i_f}{L_m + L_f} \; h^2\sigma_p$$

Trade-offs: Generation interval vs. selection intensity:
If younger animals are used (decreasing L), i is also lower,
as more of the newborn animals are needed as replacements

Computing generation intervals

OFFSPRING	Year 2	Year 3	Year 4	Year 5	total
Number (sires)	60	30	0	0	90
Number (dams)	400	600	100	40	1140

$$L_s = \frac{2 \cdot 60 + 3 \cdot 30}{60 + 30} = 2.33,$$

$$L_d = \frac{2 \cdot 400 + 3 \cdot 600 + 4 \cdot 100 + 5 \cdot 40}{400 + 600 + 100 + 40} = 2.81$$

Generalized Breeder's Equation

$$R_y = \frac{i_m + i_f}{L_m + L_f} \, r_{uA}\sigma_A$$

Tradeoff between generation length L and accuracy r

The longer we wait to replace an individual, the more accurate the selection (i.e., we have time for progeny testing and using the values of its relatives)

Example 10.8. As an example of the tradeoff between accuracy and generation intervals, consider a trait with $h^2 = 0.25$ and selection only on sires. One scheme is to simply select on the sire's phenotype, which results in a sire generation interval of 1.5 years. Alternatively, one might perform progeny testing to improve the accuracy of the selected sires. This results in an increase of the sire generation interval to (say) 2.5 years. Suppose in both cases, the dam interval is steady at 1.5 years.

Since the intensity of selection and additive genetic variation are the same in both schemes, the ratio of response under mass selection to response under progeny testing is just

$$\frac{R(\text{Sire phenotype})}{R(\text{progeny mean})} = \frac{\rho(A, \text{Sire phenotype})/(L_s + L_d)}{\rho(A, \text{progeny mean})/(L_s + L_d)}$$

Here, $\rho(A, \text{Sire phenotype}) = h = \sqrt{0.25} = 0.5$, with generation intervals $L_s + L_d = 1.5 + 1.5 = 3$. With progeny testing, (Example 10.4)

$$\rho(A, \text{progeny mean}) = \sqrt{\frac{n}{n+a}} = \sqrt{\frac{n}{n+15}}$$

as $a = (4 - h^2)/(h^2) = 15$, with a total generation interal of $L_s + L_d = 2.5 + 1.5 = 4$. Hence,

$$\frac{R(\text{Sire phenotype})}{R(\text{progeny mean})} = \frac{0.5/3.0}{\sqrt{\frac{n}{n+15}}/4} = \frac{2}{3} \cdot \sqrt{\frac{n+15}{n}}$$

If (say) $n = 2$ progeny are tested per sire, this ratio is 1.95, giving a much larger rate of response under sire-only selection. For $n = 12$, the ratio is exactly one, while for a very large number of offspring tested per sire, the ratio approaches $2/3$, or a 1.5-fold increase in the rate of response under progeny testing, despite the increase in sire generation interval.
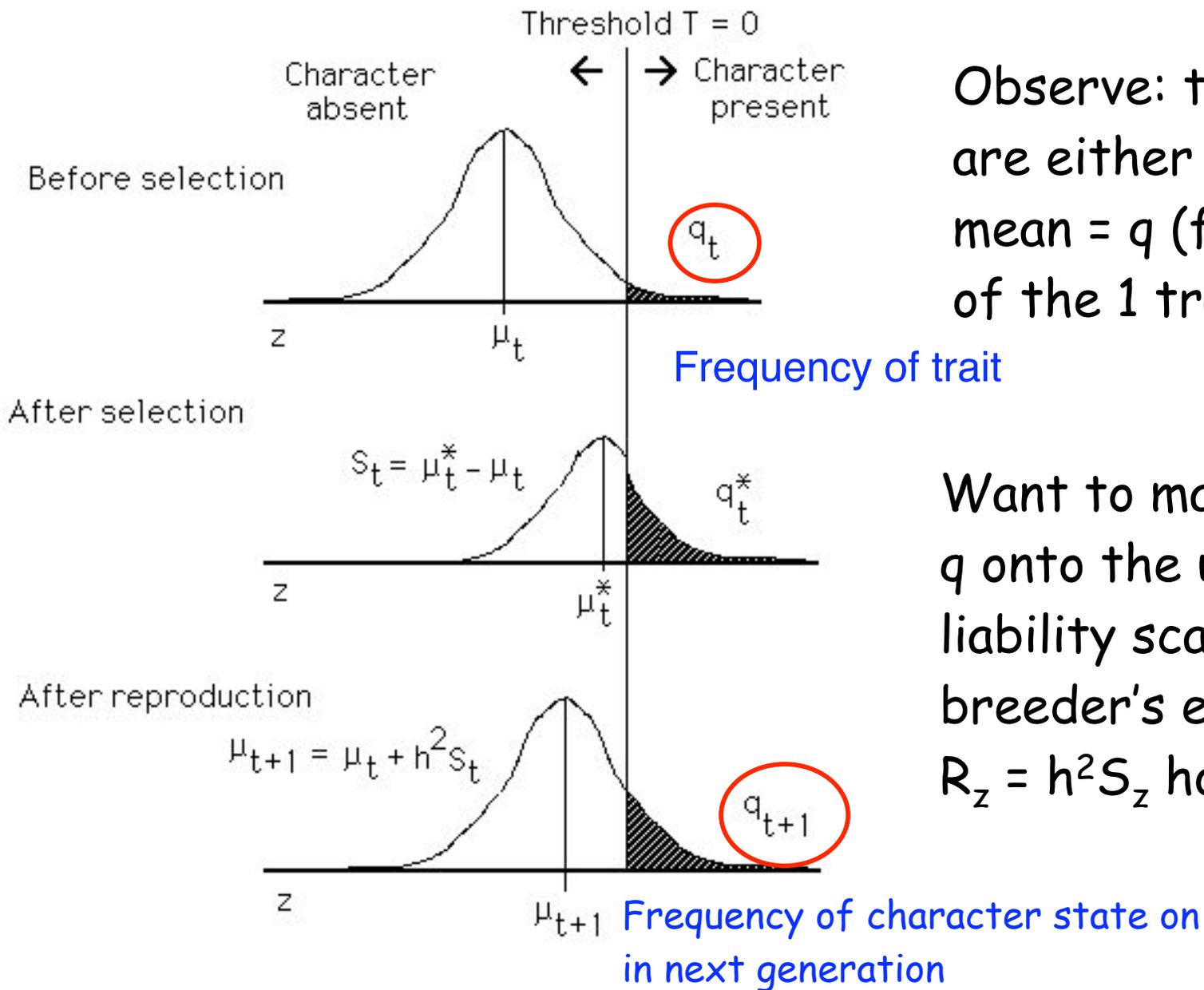
Selection on Threshold Traits

Response on a binary trait is a special case of response on a continuous trait

Assume some underlying continuous value z, the liability, maps to a discrete trait.
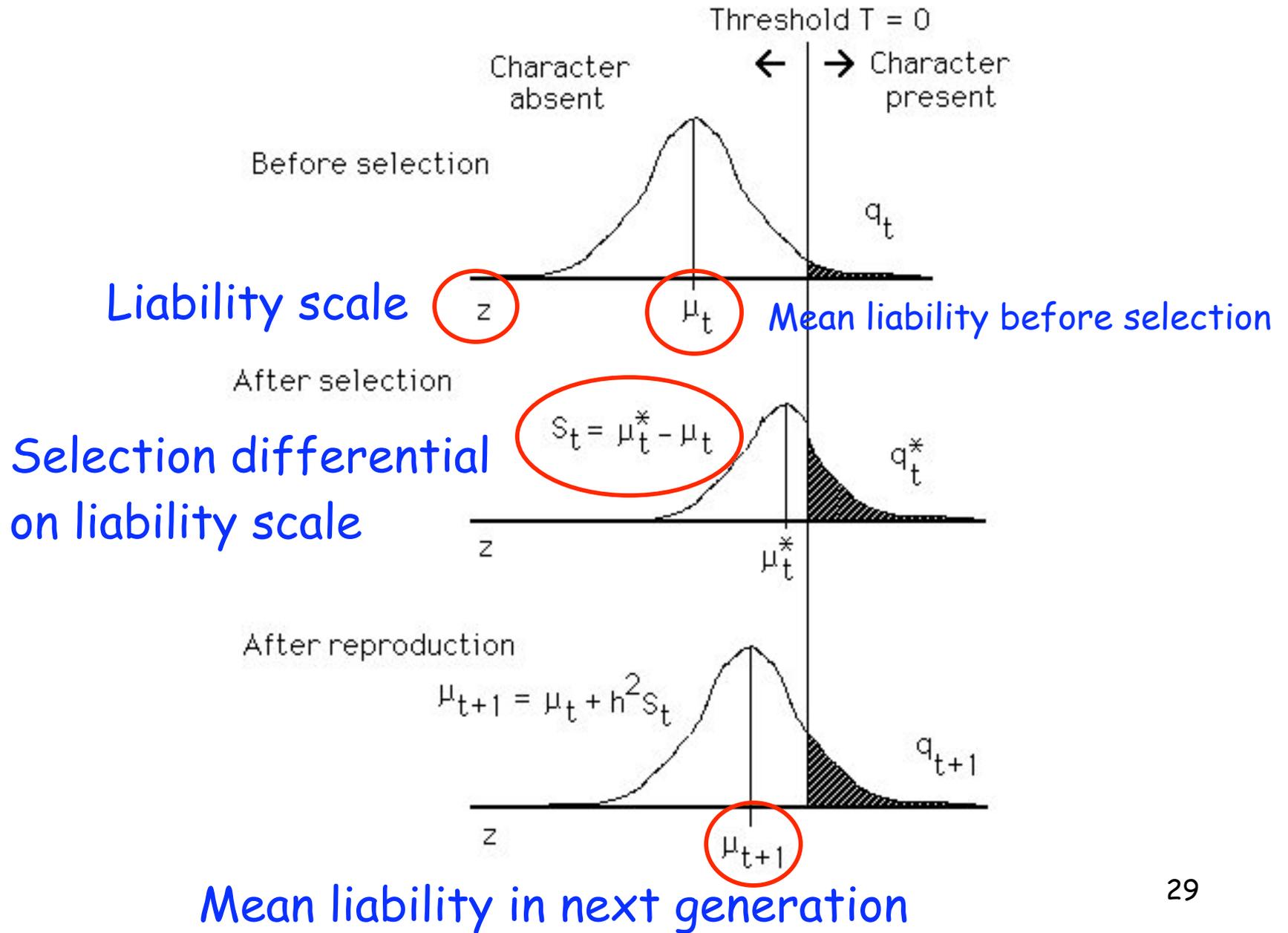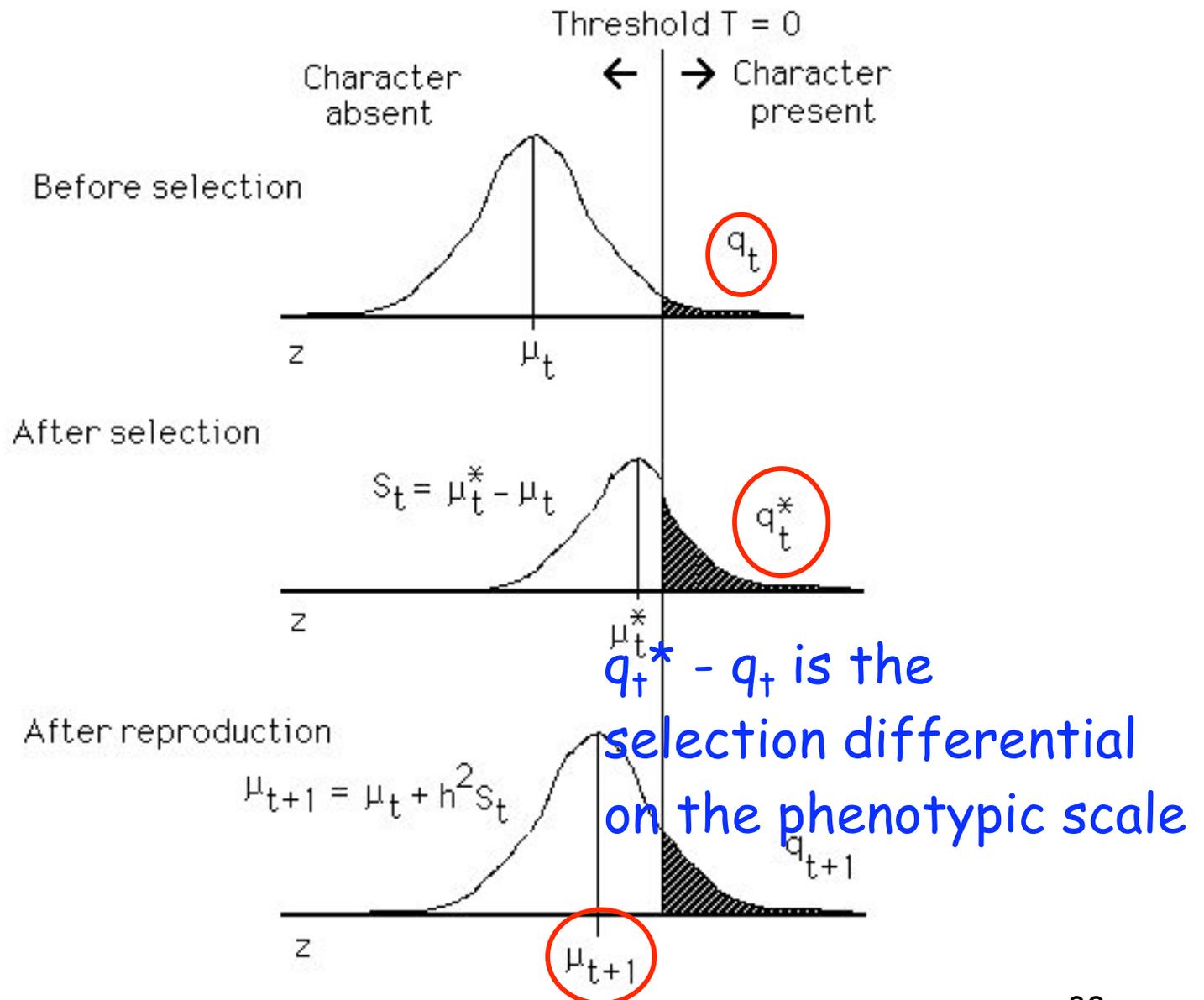
$z < T$ character state zero (i.e. no disease)

$z \geq T$ character state one (i.e. disease)

Alternative (but essentially equivalent model) is a probit (or logistic) model, when $p(z)$ = Prob(state one | z). Details in LW Chapter 14.

Threshold T = 0

Character absent ← | → Character present

Before selection

z μ_t q_t

Frequency of trait

After selection

$S_t = \mu_t^* - \mu_t$ q_t^*

z μ_t^*

After reproduction

$\mu_{t+1} = \mu_t + h^2 S_t$ q_{t+1}

z μ_{t+1}

Frequency of character state on in next generation

Observe: trait values are either 0,1. Pop mean = q (frequency of the 1 trait)

Want to map from q onto the underlying liability scale z, where breeder's equation $R_z = h^2 S_z$ holds

28

Threshold T = 0

Character absent ← | → Character present

Before selection

q_t

Liability scale z μ_t *Mean liability before selection*

After selection

$S_t = \mu_t^* - \mu_t$

q_t^*

Selection differential on liability scale

z μ_t^*

After reproduction

$\mu_{t+1} = \mu_t + h^2 S_t$

q_{t+1}

z μ_{t+1}

Mean liability in next generation

29

Threshold T = 0

Character absent ← | → Character present

Before selection

q_t

z μ_t

After selection

$S_t = \mu_t^* - \mu_t$

q_t^*

z μ_t^*

After reproduction

$\mu_{t+1} = \mu_t + h^2 S_t$

q_{t+1}

z μ_{t+1}

$q_t^* - q_t$ is the selection differential on the phenotypic scale

Mean liability in next generation

Steps in Predicting Response to Threshold Selection

i) Compute initial mean μ_0

$P(\text{trait}) = P(z \geq 0) = P(z - \mu \geq -\mu) = P(\textcolor{blue}{U} \geq -\mu)$

$\textcolor{blue}{U \text{ is a unit normal}}$

Hence, $z - \mu_0$ is a unit normal random variable

We can choose a scale where the liability z has variance of one and a threshold $T = 0$

Define $z_{[q]} = P(U < z_{[q]}) = q$. $P(U \geq z_{[1-q]}) = q$

General result: $\textcolor{red}{\mu = -z_{[1-q]}}$

For example, suppose 5% of the pop shows the trait. $P(U > 1.645)$ = 0.05, hence $\mu = -1.645$. Note: in R, $z_{[1-q]}$ = $\textcolor{orange}{\text{qnorm(1-q)}}$, with qnorm(0.95) returning 1.644854

Steps in Predicting Response to Threshold Selection

ii) The frequency q_{t+1} of the trait in the next generation is just

$$q_{t+1} = P(U > - \mu_{t+1}) = P(U > - [h^2 S + \mu_t])$$

$$= P(U > - h^2 S - z_{[1-q]})$$

iii) Hence, we need to compute S, the selection differential for the liability z

Let p_t = fraction of individuals chosen in generation t that display the trait

$$\mu_t^* = (1 - p_t) E(z \mid z < 0, \mu_t) + p_t E(z \mid z \geq 0, \mu_t)$$

$$\mu_t^* = (1 - p_t)E(z \mid z < 0, \mu_t) + p_t E(z \mid z \geq 0, \mu_t)$$

This fraction does not display
the trait, hence z < 0

This fraction displays
the trait, hence z ≥ 0

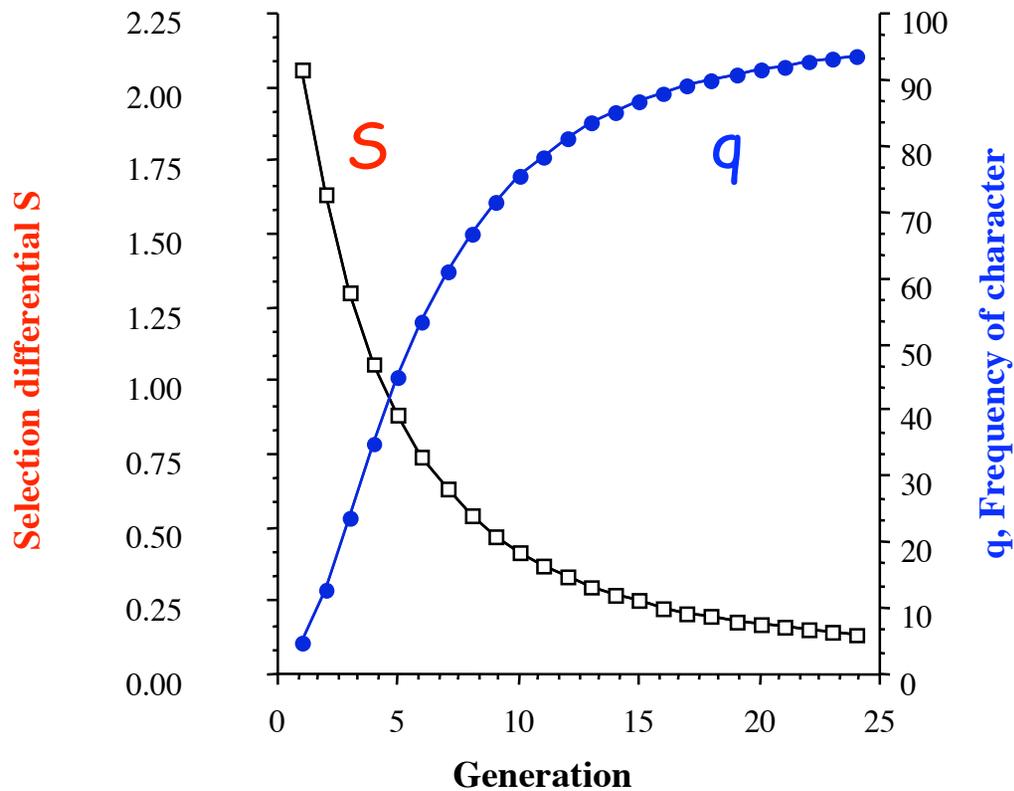## When z is normally distributed, this reduces to

$$S_t = \mu^* - \mu_t = \frac{\varphi(\mu_t)}{q_t} \frac{p_t - q_t}{1 - q_t}$$

Height of the unit normal density function
at the point μ_t

Hence, we start at some initial value given h^2 and
μ_0, and iterative to obtain selection response

33

Initial frequency of q = 0.05. Select only on adults showing the trait ($p_t = 1$)

Permanent Versus Transient Response

Considering epistasis and shared environmental values, the single-generation response follows from the midparent-offspring regression

$$R = h^2\,S + \frac{S}{\sigma_z^2}\left(\frac{\sigma_{AA}^2}{2} + \frac{\sigma_{AAA}^2}{4} + \cdots + \sigma(E_{sire}, E_o) + \sigma(E_{dam}, E_o)\right)$$

Breeder's Equation

Response from epistasis

Response from shared environmental effects

Permanent component of response

Transient component of response --- contributes to short-term response. Decays away to zero over the long-term

Permanent Versus Transient Response

The reason for the focus on h^2S is that this component is <u>permanent</u> in a random-mating population, while the other components are <u>transient</u>, initially contributing to response, but this contribution decays away under random mating

Why? Under HW, changes in allele frequencies are permanent (don't decay under random-mating), while LD (epistasis) does, and environmental values also become randomized

Response with Epistasis

The response after one generation of selection from an unselected base population with A x A epistasis is

$$R = S \left(h^2 + \frac{\sigma_{AA}^2}{2\,\sigma_z^2} \right)$$

The contribution to response from this single generation after τ generations of no selection is

$$R(1+\tau) = S \left(h^2 + (1-c)^\tau \frac{\sigma_{AA}^2}{2\,\sigma_z^2} \right)$$

c is the average (pairwise) recombination between loci involved in A x A

Response with Epistasis

$$R(1 + \tau) = S \left(h^2 + (1 - c)^\tau \frac{\sigma_{AA}^2}{2\sigma_z^2} \right)$$

Response from additive effects (h^2 S) is due to changes in allele frequencies and hence is permanent. Contribution from A x A due to linkage disequilibrium

Contribution to response from epistasis decays to zero as linkage disequilibrium decays to zero

Why breeder's equation assumption of an unselected base population?
If history of previous selection, linkage disequilibrium may be present
and the mean can change as the disequilibrium decays

For t generation of selection followed by
τ generations of no selection (but recombination)
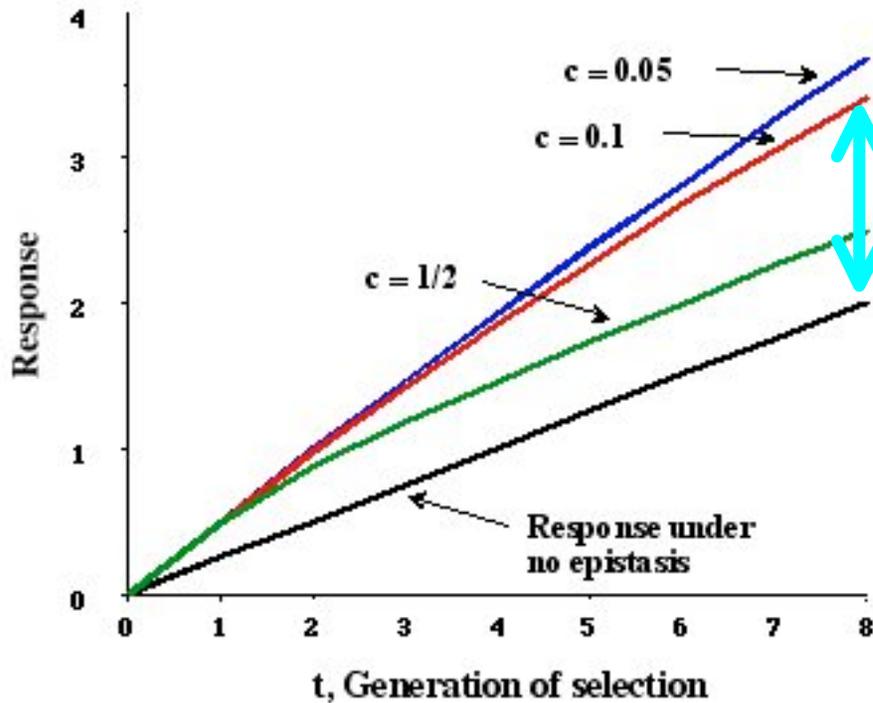
$$R(t+\tau) = t\,h^2\,S + (1-c)^\tau\,R_{AA}(t)$$

R_{AA} has a limiting value given by

$$\tilde{R}_{AA} = \lim_{t\to\infty} R_{AA}(t) = \frac{1}{c}\left(S\,\frac{\sigma_{AA}^2}{2\,\sigma_z^2}\right)$$

Time to equilibrium a function of c

$$t_{1/2} = \frac{-\ln(2)}{\ln(1-c)}$$

Decay half-life

$$= \frac{1}{c}\left(S\,\frac{\sigma_{AA}^2}{2\,\sigma_z^2}\right)$$

Fixed incremental difference that decays when selection stops

What about response with higher-order epistasis?

$S\sigma^2(A^i)/\sigma_z^2$,	AA	AAA	AAAA	AAAAA
$R(1)$	0.500	0.250	0.125	0.063
Limit	1.000	0.333	0.143	0.067
% $R(1)/$limit	50.0	75.0	87.5	93.8

Response in autotetraploids

- Autotraploids pass along two alleles at each locus to their offspring

- Hence, dominance variance is passed along

- However, as with A x A, this depends upon favorable combinations of alleles, and these are randomized over time by transmission, so D component of response is transient.

Autotetraploids

P-O covariance

Single-generation response

$$\sigma(z_p, z_o) = \frac{\sigma_A^2}{2} + \frac{\sigma_D^2}{6}, \qquad R = S\left(h^2 + \frac{\sigma_D^2}{3\sigma_z^2}\right)$$

$$R(t) = th^2 S + R_D(t)$$

Response to t generations of selection with constant selection differential S

$$R_D(t) = S\,\frac{3}{2}\left[1 - \left(\frac{1}{3}\right)^t\right]\frac{\sigma_D^2}{3\sigma_z^2}$$

Response remaining after t generations of selection followed by τ generations of random mating

$$t\,h^2\,S + (1/3)^\tau\,R_D(t)$$

Contribution from dominance quickly decays to zero

42

General responses

- For both individual and family selection, the response can be thought of as a regression of some phenotypic measurement (such as the individual itself or its corresponding selection unit value x) on either the offspring value (y) or the breeding value R_A of an individual who will be a parent of the next generation (the recombination group).

- The regression slope for predicting y from x is $\sigma(x,y)/\sigma^2(x)$ and $\sigma(x,R_A)/\sigma^2(x)$ for predicting the BV R_A from x

- With transient components of response, these covariances now also become functions of time --- e.g. the covariance between x in one generation and y several generations later

Ancestral Regressions

When regressions on relatives are linear, we can think of the response as the sum over all previous contributions

For example, consider the response after 3 gens:

$$R(3) = 8\,\beta_{3,0}\,S_0 + 4\,\beta_{3,1}\,S_1 + 2\,\beta_{3,2}\,S_2$$

2 parents

8 great-grand parents
S_0 is there selection differential
$\beta_{3,0}$ is the regression coefficient for an offspring at time 3 on a great-grandparent From time 0

4 grandparents
Selection diff S_1

$\beta_{3,1}$ is the regression of relative in generation 3 on their gen 1 relatives

Ancestral Regressions

More generally,

$$R(T) = \sum_{t=0}^{T-1} 2^{T-t} \beta_{T,t} S_t$$

$$\beta_{T,t} = \text{cov}(z_T, z_t)$$

The general expression $\text{cov}(z_T, z_t)$, where we keep track of the actual generation, as oppose to $\text{cov}(z, z_{T-t})$ -- how many generations separate the relatives, allows us to handle inbreeding, where the regression slope changes over generations of inbreeding.

Unless $2^t \bar{\beta}_{\tau+t,\tau}$ remains constant as t increases, the contribution to cumulative response from selection on adults in generation τ changes over time. For example, when loci are strictly additive (no dominance or epistasis), $\sigma_G(\tau + t, \tau) = 2^{-t} \sigma_A^2(\tau)$ and thus $2^t \beta_{\tau+t,\tau} = h_\tau^2$, the standard result from the breeders' equation. However, unless $2^t \sigma_G(\tau+t,\tau)$ remains constant, any response contributed decays. Hence any term of $\sigma_G(\tau + t, \tau)$ that decreases by more than $1/2$ each generation contributes only to the transient response.

Maternal Effects:

Falconer's dilution model

$$z = G + m \, z_{dam} + e$$

G = Direct genetic effect on character
G = A + D + I. $E[A] = (A_{sire} + A_{dam})/2$

maternal effect passed from dam to offspring $m \, z_{dam}$ is
just a fraction m of the dam's phenotypic value

The presence of the maternal effects means that response
is not necessarily linear and time lags can occur in response

m can be negative --- results in the potential for
a reversed response

Parent-offspring regression under the dilution model

In terms of parental breeding values,

$$E(z_o \mid A_{dam}, A_{sire}, z_{dam}) = \frac{A_{dam}}{2} + \frac{A_{sire}}{2} + m \, z_{dam}$$

Regression of BV on phenotype

$$A = \mu_A + b_{Az} \, (z - \mu_z) + e$$

The resulting slope becomes b_{Az} = h^2 2/(2-m)

With no maternal effects, b_{az} = h^2

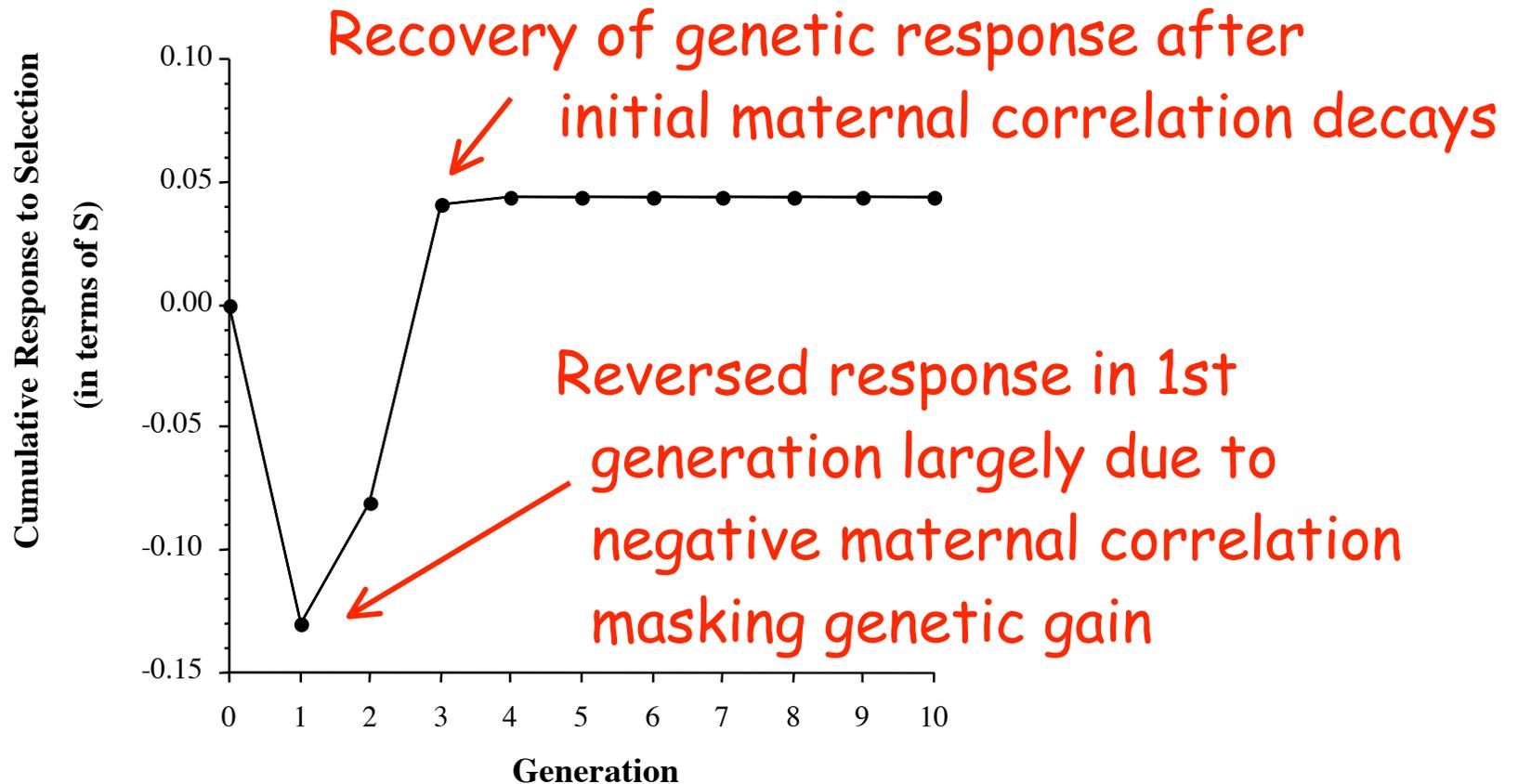# Parent-offspring regression under the dilution model

With maternal effects, a covariance between BV and maternal effect arises, with $\sigma_{A,M} = m\,\sigma_A^2\,/\,(2-m)$

The response thus becomes

$$\Delta\mu_z = S_{dam}\left(\frac{h^2}{2-m} + m\right) + S_{sire}\,\frac{h^2}{2-m}$$

Response to a single generation of selection

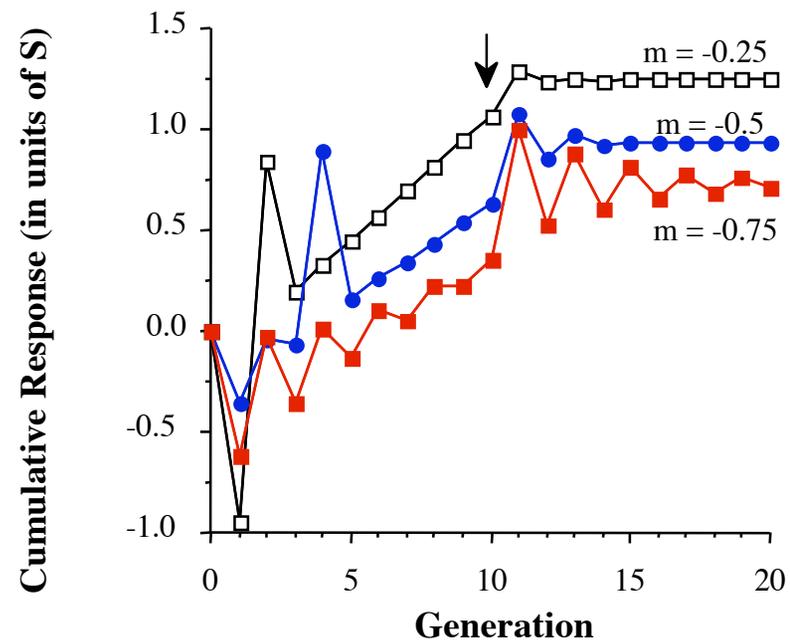## $h^2 = 0.11$, $m = -0.13$ (litter size in mice)

Recovery of genetic response after initial maternal correlation decays

Reversed response in 1st generation largely due to negative maternal correlation masking genetic gain

Maternal effects

- Other models, these are discussed in WL chapters 15 and 22.

- Ultimately, a multivariate problem
  - Direct trait (i.e., weight)
  - Contribution from the mother (beyond her genes

Selection occurs for 10 generations and then stops

$h^2 = 0.35$