# Lecture 6: Selection on Multiple Traits

Bruce Walsh lecture notes
Introduction to Quantitative Genetics
SISG, Seattle
19 – 21 July 2023

Genetic vs. Phenotypic correlations

- Within an individual, trait values can be positively or negatively correlated,
  - height and weight -- positively correlated
  - Weight and lifespan -- negatively correlated
- Such phenotypic correlations can be directly measured,
  - r_P denotes the phenotypic correlation
- Phenotypic correlations arise because genetic and/or environmental values within an individual are correlated.

The phenotypic values between traits x and y within an individual are correlated

r P

P_x P_y

E_x E_y

A_x A_y

r_E

r_A
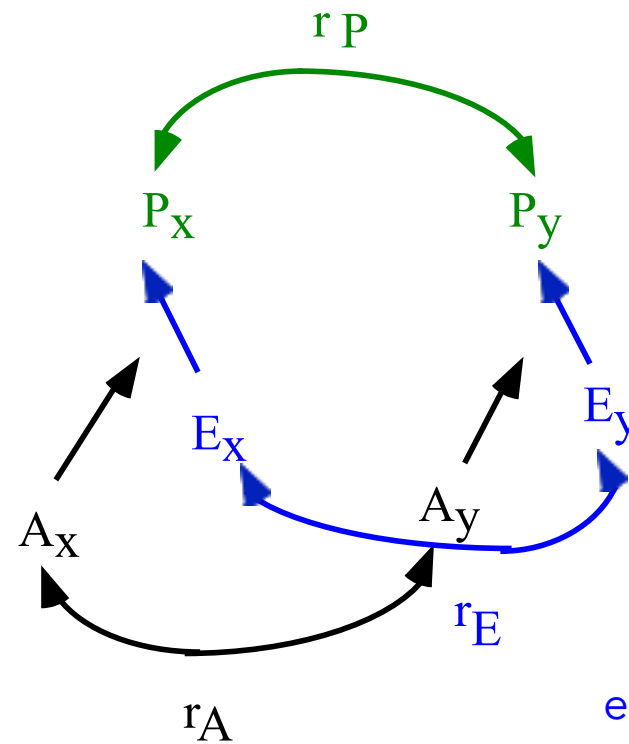
Correlations between the breeding values of x and y within the individual can generate a phenotypic correlation

Likewise, the environmental values for the two traits within the individual could also be correlated

Genetic & Environmental Correlations

- r_A = correlation in breeding values (the genetic correlation) can arise from
  - pleiotropic effects of loci on both traits
  - linkage disequilibrium, which decays over time
- r_E = correlation in environmental values
  - includes non-additive genetic effects (e.g., D, I)
  - arises from exposure of the two traits to the same individual environment

The relative contributions of genetic and environmental correlations to the phenotypic correlation

$$r_P = r_A h_X h_Y + r_E \sqrt{(1 - h_x^2)(1 - h_Y^2)}$$

If heritability values are high for both traits, then the correlation in breeding values dominates the phenotypic corrrelation

If heritability values in EITHER trait are low, then the correlation in environmental values dominates the phenotypic correlation

In practice, phenotypic and genetic correlations often have the same sign and are of similar magnitude, but this is not always the case

Estimating Genetic Correlations

Recall that we estimated V_A from the regression of trait x in the parent on trait x in the offspring,

Slope =
$(1/2)\ V_A(x)/V_P(x)$

Trait x in
offspring

$V_A(x) = 2\ *slope\ *\ V_P(x)$
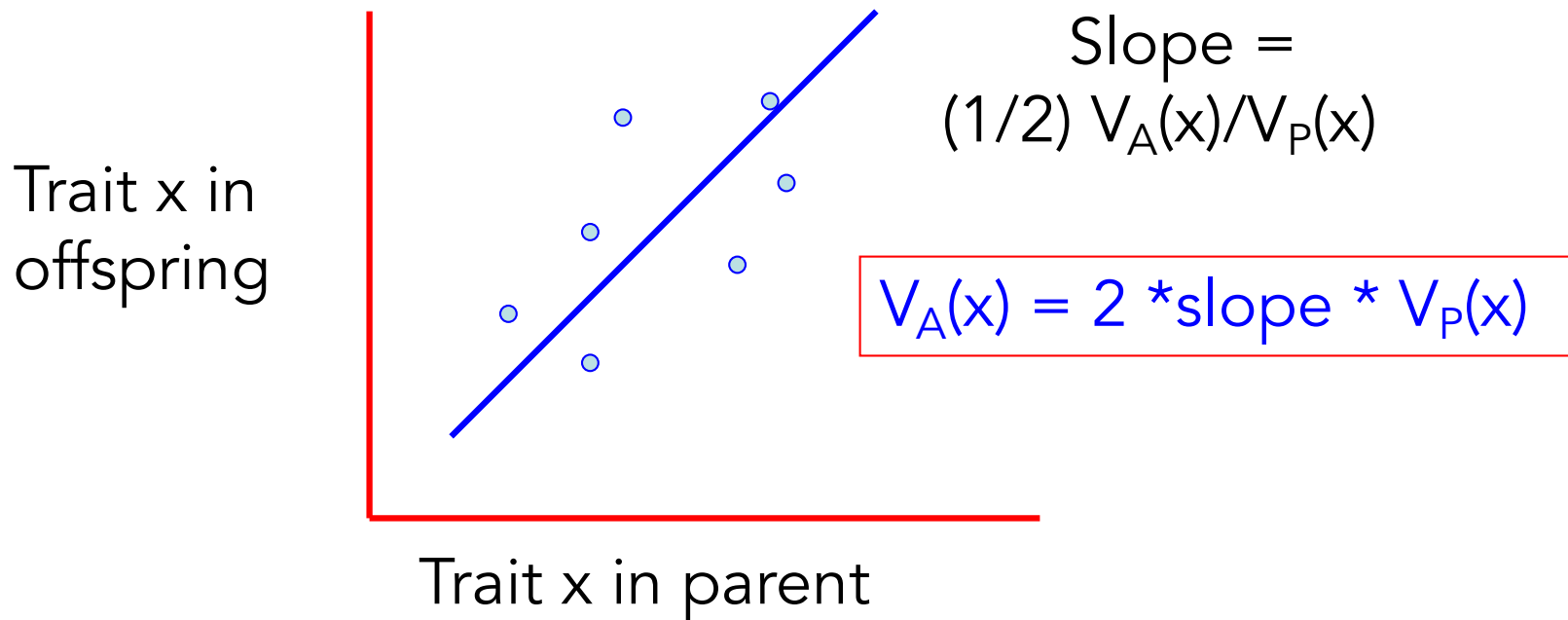
Trait x in parent

Estimating Genetic Correlations

Similarly, we can estimate $V_A(x,y)$, the covariance in the breeding values for traits x and y, by the regression of trait x in the parent and trait y in the offspring

Trait y in offspring

Slope = $(1/2) V_A(x,y)/V_P(x)$

$V_A(x,y) = 2 *slope * V_P(x)$

Trait x in parent
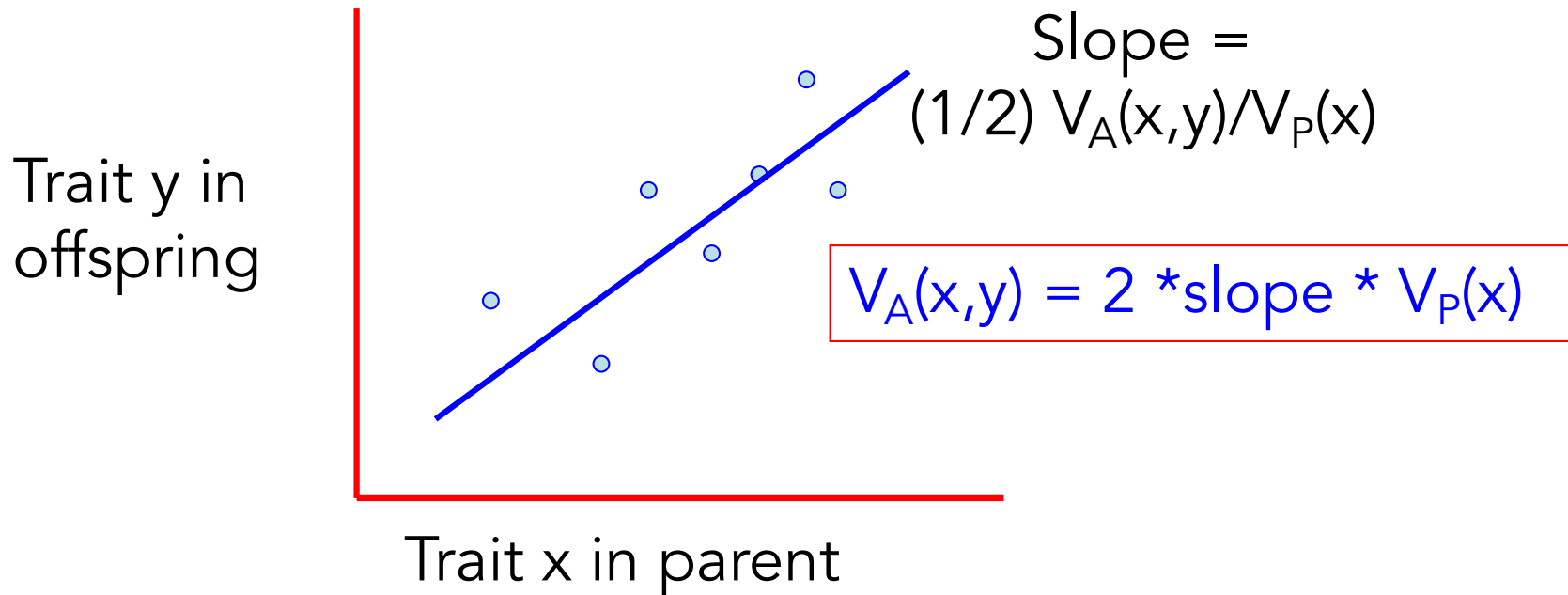
Thus, one estimator of $V_A(x,y)$ is

$$V_A(x,y) = \frac{2 * b_{y|x} * V_P(x) + 2 * b_{x|y} * V_P(y)}{2}$$

giving

$$V_A(x,y) = b_{y|x} V_P(x) + b_{x|y} V_P(y)$$

Put another way,

$$Cov(x_O,y_P) = Cov(y_O,x_P) = (1/2)Cov(A_x,A_y)$$
$$Cov(x_O,x_P) = (1/2) V_A(x) = (1/2)Cov(A_x, A_x)$$
$$Cov(y_O,y_P) = (1/2) V_A(y) = (1/2)Cov(A_y, A_y)$$

Likewise, for half-sibs,

$$Cov(x_{HS},y_{HS}) = (1/4) Cov(A_x,A_y)$$
$$Cov(x_{HS},x_{HS}) = (1/4) Cov(A_x,A_x) = (1/4) V_A(x)$$
$$Cov(y_{HS},y_{HS}) = (1/4) Cov(A_y,A_y) = (1/4) V_A(y)$$

General: $Cov(x_i,y_j) = 2\Theta_{ij}Cov(A_x,A_y)$

Key ideas so far …

– The genetic covariance between two traits is <u>the covariance in breeding values</u>, so that a positive value implies individuals with a high BV for one trait tend to have high BV for the other trait

– We can <u>estimate this genetic covariance from the phenotypic covariance</u> between trait A in one relative and trait B in another (known) relative.

– Knowledge of genetic correlations is required to <u>predict multiple trait response</u>.

Correlated Response to Selection

Direct selection of a character can cause a within-generation change in the mean of a phenotypically correlated character.

Select All

Direct selection on x also changes the mean of y
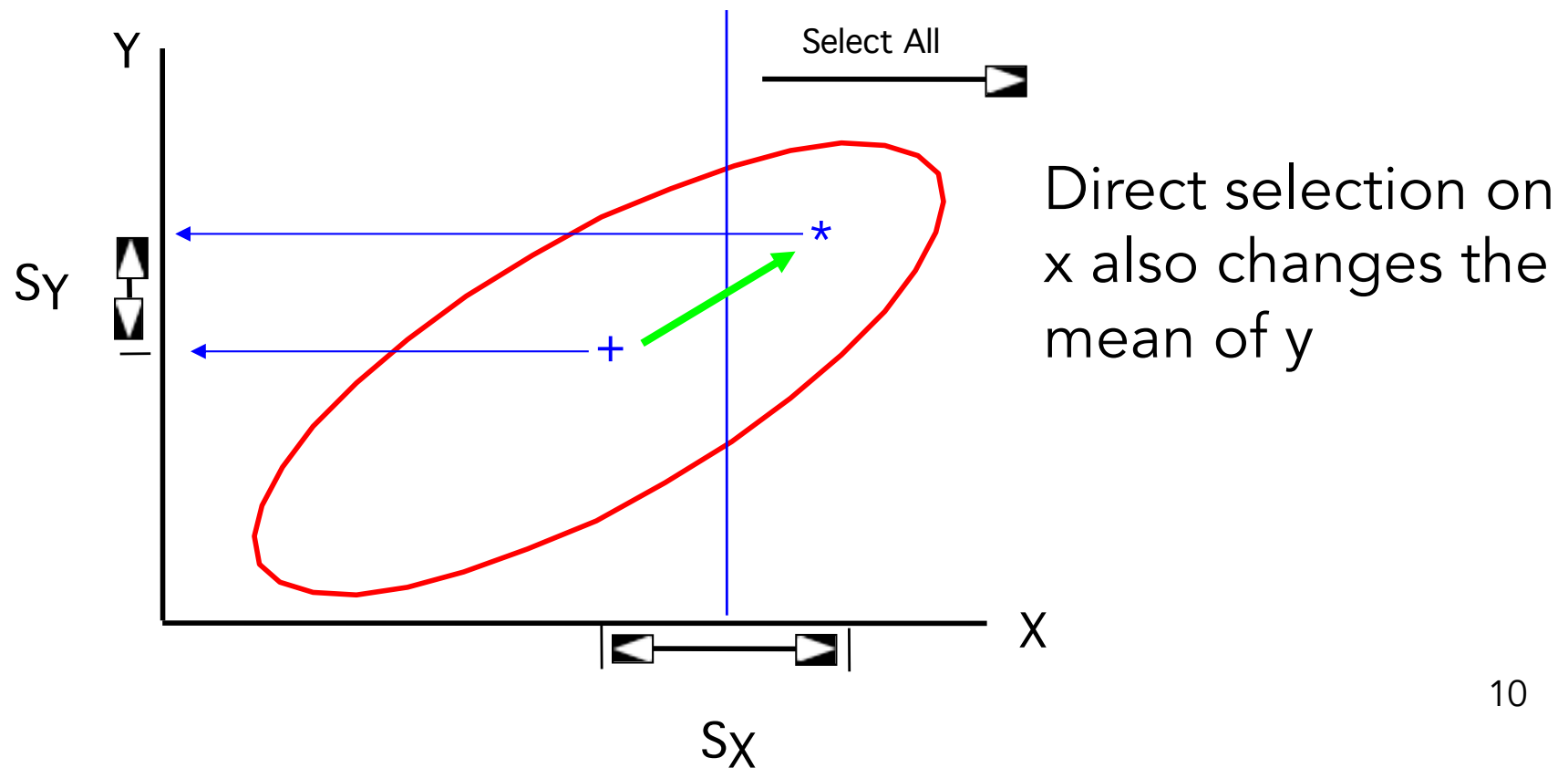
Phenotypic correlations induce within-generation changes
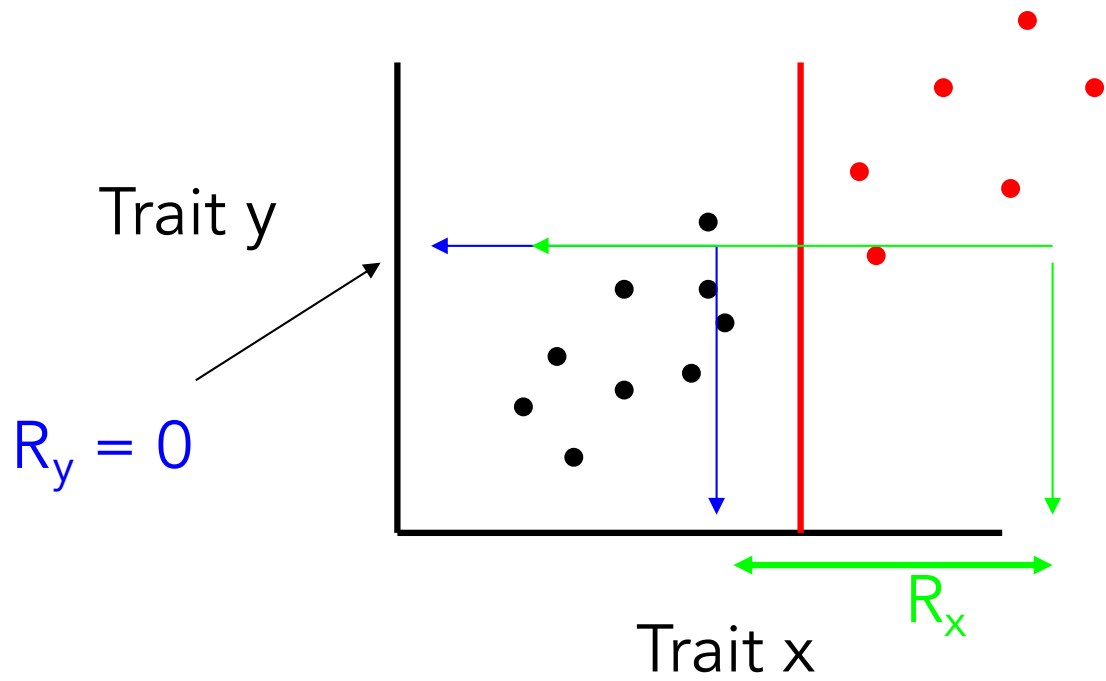
Phenotypic values

Trait y

S_y

S_x
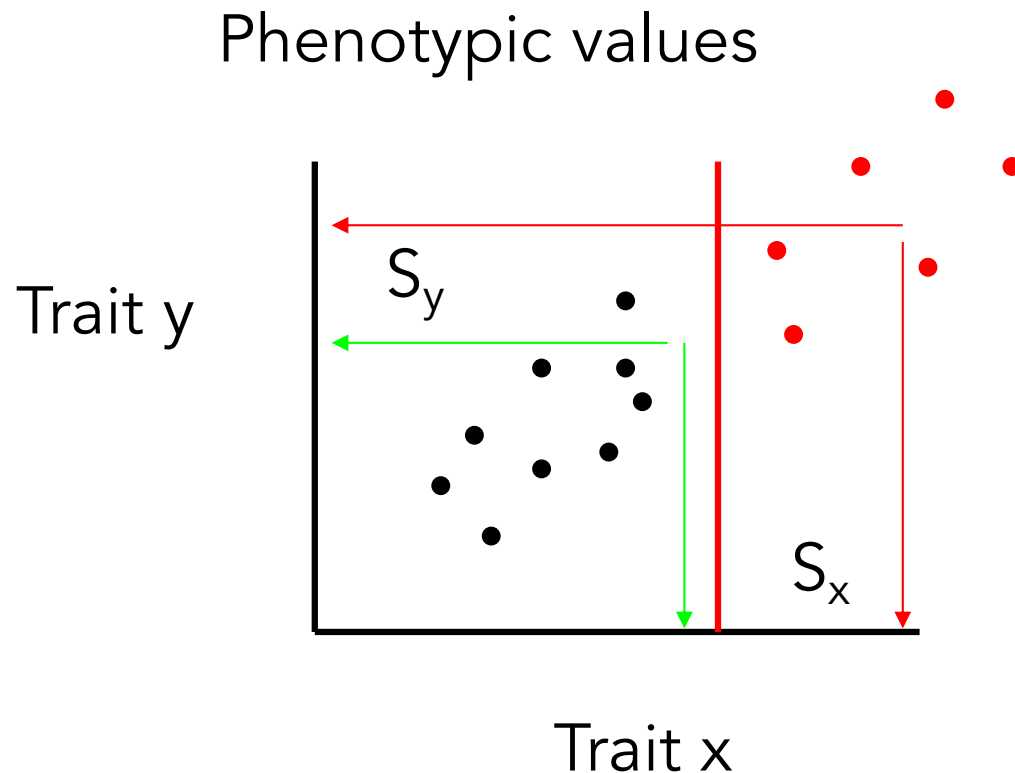
Trait x

For there to be a between-generation change, the breeding values must be correlated. Such a change is called a correlated response to selection

Example

- Suppose h^2 trait x = 0.5, h^2 trait y = 0.3
- Select on trait one to give S_x = 10
  - Expected response is R_x = 5
- Suppose $Cov(t_x, t_y)$ = 0.5, then S_y = 5
- What is the response in trait y?
  - is it CR_y = 0.3*5 = 1.5. NO!
  - Could be positive, negative, or zero
  - Depends on the Genetic correlation between traits x and y. Why??

Phenotypic values

Trait y

$R_y = 0$

R_x

Trait x
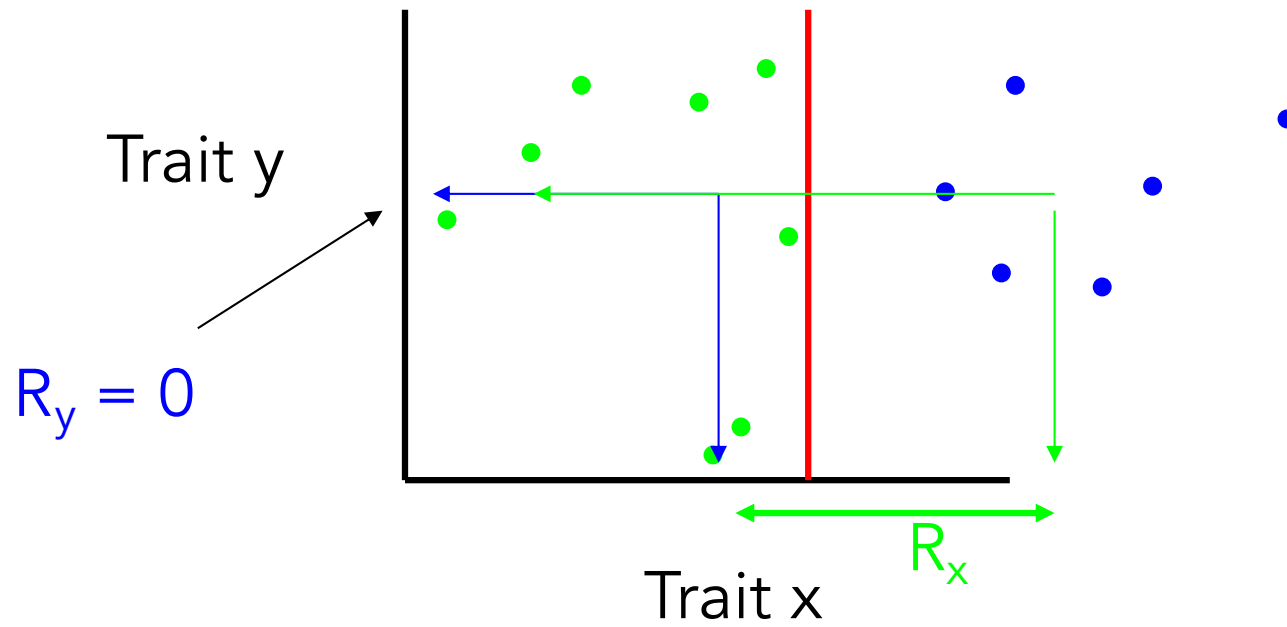
Phenotypic values are misleading, what we want are the breeding values for each of the selected individuals. Each arrow takes an individual's phenotypic value into its actual breeding value.

Breeding values

Trait y

$R_y = 0$

R_x

Trait x

Breeding values

Trait y

$R_y = 0$

R_x

Trait x
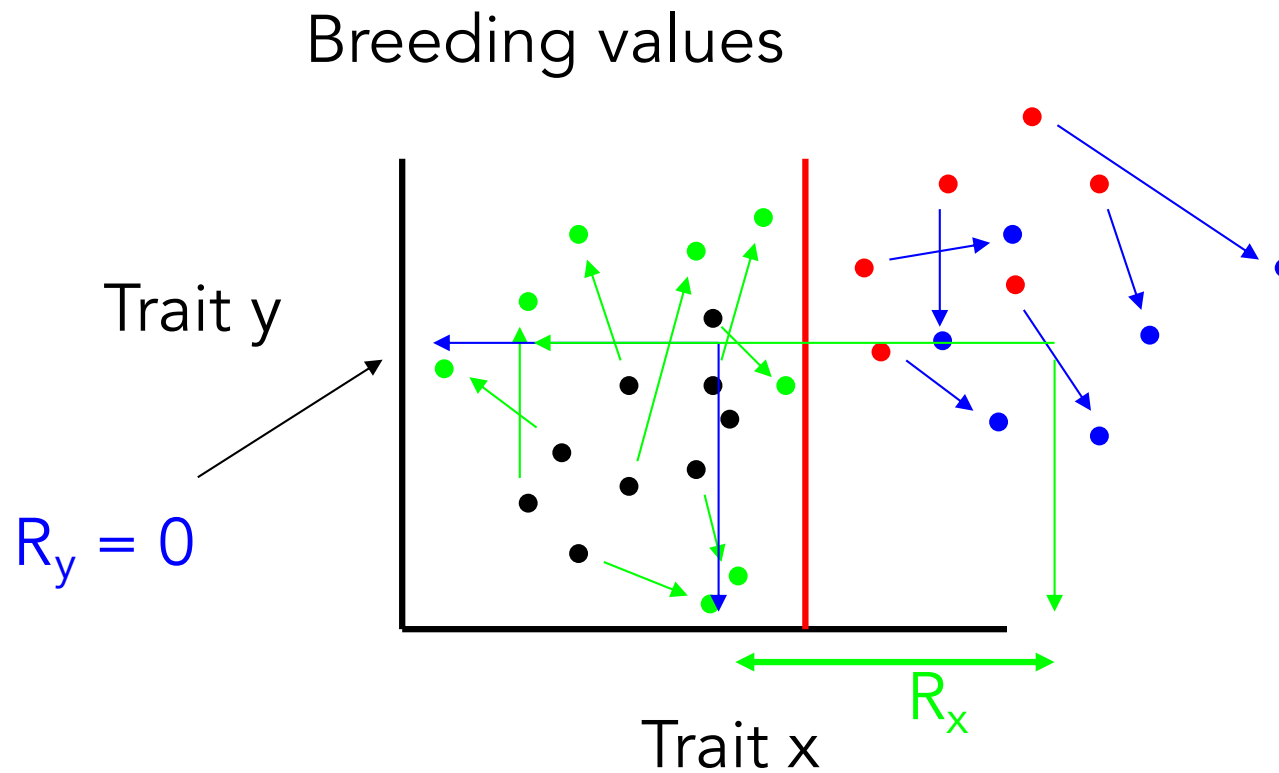
Predicting the correlated response

The change in character y in response to selection on x is the regression of the breeding value of y on the breeding value of x,

$$A_y = b_{Ay|Ax} A_x$$

where

$$b_{Ay|Ax} = \frac{Cov(A_x, A_y)}{Var(A_x)} = r_A \frac{\sigma(A_y)}{\sigma(A_x)}$$

If R_x denotes the direct response to selection on x, CR_y denotes the correlated response in y, with

$$CR_y = b_{Ay|Ax} R_x$$

We can rewrite $CR_y = b_{Ay|Ax} R_x$ as follows

First, note that $R_x = h^2_x S_x = i_x h_x \sigma_A (x)$

$\uparrow$

Recall that $i_x = S_x/\sigma_P$
(x) is the selection
intensity on x

Since $b_{Ay|Ax} = r_A \sigma_A(x) / \sigma_A(y)$,

We have $CR_y = b_{Ay|Ax} R_x = r_A \sigma_A (y) h_x i_x$

Substituting $\sigma_A (y) = h_y \sigma_P (y)$ gives our final result:

$$CR_y = i_x h_x h_y r_A \sigma_P (y)$$

$$CR_y = i_x h_x h_y r_A \sigma_P (y)$$

Noting that we can also express the direct response as

$$R_x = i_x h_x^2 \sigma_p (x)$$

shows that $h_x h_y r_A$ in the corrected response plays the same role as h_x^2 does in the direct response. As a result, $h_x h_y r_A$ is often called the co-heritability

Key ideas …

– Phenotypic correlations can result in a within-generation change

  • trait y has its trait mean changed by selection on a phenotypically-correlated trait x)

– Between-generation change (selection response) depends on genetic correlations.

  • For an indirectly selected trait, S has no bearing on R

  • We require a change in the mean **BV** in the selected parents, not simply a change in their mean **trait** value

Direct vs. Indirect Response

We can change the mean of x via a direct response R_x or an indirect response CR_x due to selection on y

$$\frac{CR_X}{R_X} = \frac{i_Y\, r_A\, \sigma_{AX}\, h_Y}{i_X\, h_X\, \sigma_{AX}} = \frac{i_Y\, r_A\, h_Y}{i_X\, h_X}$$

Hence, indirect selection gives a large response when

$$i_Y\, r_A\, h_Y > i_X\, h_X$$

• The selection intensity is much greater for y than x. This would be true if y were measurable in both sexes but x measurable in only one sex.

• Character y has a greater heritability than x, and the genetic correlation between x and y is high. This could occur if x is difficult to measure with precision but y is not.

20

G x E

The same trait measured over two (or more) environments can be considered as two (or more) correlated traits.

If the genetic correlation $|\rho| = 1$ across environments and the genetic variance of the trait is the same in both environments, then no G x E

However, if $|\rho| < 1$, and/or Var(A) of the trait varies over environments, then G x E present

Hence, dealing with G x E is a *multiple-trait problem*

Participatory breeding

The environment where a crop line is developed may be different from where it is grown

An especially important example of this is participatory breeding, wherein subsistence farmers are involved in the field traits.

Here, the correlated response is the yield in subsistence environment given selection at a regional center, while direct response is yield when selection occurred in subsistence environment. Regional center selection works when

$$i_Y \, r_A \, h_Y > i_X \, h_X$$

Matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad B = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \qquad C = \begin{pmatrix} i \\ j \end{pmatrix}$$

Dimensions given by rows x columns (r x c)

The identity matrix I, $\quad I_{2\times2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Matrix Multiplication

$$\mathbf{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix}$$

$$= \begin{pmatrix} ae + bg & af + bh \\ ce + dg & cf + dh \end{pmatrix}$$

In order to multiply two matrices, they must <u>conform</u>

$$A_{r \times c} \, B_{c \times k} = C_{r \times k}$$

Matrix Multiplication

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad B = \begin{pmatrix} e & f \\ g & h \end{pmatrix} \qquad C = \begin{pmatrix} i \\ j \end{pmatrix}$$

$$BA = \begin{pmatrix} ae + cf & eb + df \\ ga + ch & gd + dh \end{pmatrix} \qquad AC = \begin{pmatrix} ai + bj \\ ci + dj \end{pmatrix}$$

The identity matrix I serves the role of one in matrix multiplication: AI =A, IA = A

The Inverse Matrix, A^{-1}

For a square matrix A, define the Inverse of A, A^{-1}, as the matrix satisfying

$$\mathbf{A^{-1}A = AA^{-1} = I}$$

For $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ $\mathbf{A}^{-1} = \dfrac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

If this quantity (the determinant) is zero, the inverse does not exist.

The inverse serves the role of division in matrix multiplication

Suppose we are trying to solve the system $Ax = c$ for x.

$A^{-1} Ax = A^{-1} c$. Note that $A^{-1} Ax = Ix = x$, giving $x = A^{-1} c$

The Multivariate Breeders' Equation

Suppose we are interested in the vector R of responses when selection occurs on n correlated traits

Let S be the vector of selection differentials.

In the univariate case, the relationship between R and S was the Breeders' Equation, $R = h^2 S$

What is the multivariate version of this?

$$\mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix} \qquad \mathbf{R} = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{pmatrix}$$

$$\mathbf{P} = \begin{pmatrix} \sigma^2(z_1) & \sigma(z_1, z_2) \\ \sigma(z_1, z_2) & \sigma^2(z_2) \end{pmatrix}$$

$$\mathbf{G} = \begin{pmatrix} \sigma^2(A_1) & \sigma(A_1, A_2) \\ \sigma(A_1, A_2) & \sigma^1(A_2) \end{pmatrix}$$

The multivariate breeder's equation

$$R = G \, P^{-1} \, S$$

$$R = h^2 S = (V_A/V_P) \, S$$

Natural parallels with univariate breeder's equation

$P^{-1} S = \beta$ is called the selection gradient and measures the amount of direct selection on a character

The gradient version of the breeder's equation is given by $R = G\beta$. This is often called the Lande Equation (after Russ Lande)

Sources of within-generation change in the mean

Since β = P⁻¹ S, S = P β,
giving the j-th element as

Change in mean from phenotypically correlated characters under direct selection

Within-generation change in trait j

$$S_j = \sigma^2(P_j)\,\beta_j + \sum_{i \neq j} \sigma(P_j, P_i)\,\beta_i$$

Change in mean from direct selection on trait j

Within-generation change in the mean

$$S_j = \sigma^2(P_j)\,\beta_j + \sum_{i \neq j} \sigma(P_j, P_i)\,\beta_i$$

Response in the mean

Between-generation change (response) in trait j

Indirect response from genetically correlated characters under direct selection

$$R_j = \sigma^2(A_j)\,\beta_j + \sum_{i \neq j} \sigma(A_j, A_i)\,\beta_i$$

Response from direct selection on trait j

Correlated response

Direct response

32

Example in R

Consider three of these traits, $z_1 = $ oil content, $z_2 = $ protein content, and $z_3 = $ yield. For these characters, Brim et al. estimated the covariance matrices as

$$\mathbf{P} = \begin{pmatrix} 287.5 & 477.4 & 1266 \\ 477.4 & 935 & 2303 \\ 1266 & 2303 & 5951 \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} 128.7 & 160.6 & 492.5 \\ 160.6 & 254.6 & 707.7 \\ 492.5 & 707.7 & 2103 \end{pmatrix}$$

$h^2_{oil} = 128.7/287.5 = 0.45$, $h^2_{protein} = 254.6/935 = 0.27$, $h^2_{yield} = 2103/5951 = 0.35$.

Suppose you observed a within-generation change of -10 for oil, 10 for protein, and 100 for yield.

What is R? What is the nature of selection on each trait?

Enter G, P, and S

```
> P<-matrix(c(287.5,477.4,1266,477.4,935,2303,1266,2303,5951),nrow=3)
> P
        [,1]    [,2] [,3]
[1,]   287.5   477.4 1266
[2,]   477.4   935.0 2303
[3,] 1266.0 2303.0 5951
> G<-matrix(c(128.7,160.6,492.5,160.6,254.6,707.7,492.5,707.7,2103),nrow=3)
> G
        [,1]   [,2]    [,3]
[1,] 128.7 160.6   492.5
[2,] 160.6 254.6   707.7
[3,] 492.5 707.7 2103.0
> S<-matrix(c(-10,10,100),nrow=3)
> S
      [,1]
[1,]   -10
[2,]    10
[3,]   100
```

$$R = G\ P^{-1}S$$

```
> G %*% solve(P) %*% S
            [,1]
[1,] -13.57729
[2,]   12.28425
[3,]   65.14172
```

13.6 decrease in oil (breeder's -4.5)

12.3 increase in protein (breeder's 2.7)

65.1 increase in yield (breeder's 35.3)

S versus β : Observed change versus targets of Selection, β = P^{-1} S, S = P β,

$$S_j = \sigma^2(P_j)\,\beta_j + \sum_{i \neq j} \sigma(P_j, P_i)\,\beta_i$$

```
> solve(P) %*% S
            [,1]
[1,] -2.708160
[2,] -1.431750
[3,]  1.147009
```

```
> S
         [,1]
[1,]   -10
[2,]    10
[3,]   100
```

β: targets of selection

S: observed within-generation change

Observe a within-generation increase in protein, but the actual selection was to *decrease* it.

Constraints Imposed by
Genetic Correlations

While β is the directional optimally favored by selection, the actual response is dragged off this direction, with R = G β.

Example: Suppose

$$\mathbf{S} = \begin{pmatrix} 10 \\ -10 \end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix} 20 & -10 \\ -10 & 40 \end{pmatrix}, \qquad \mathbf{G} = \begin{pmatrix} 20 & 5 \\ 5 & 10 \end{pmatrix}$$

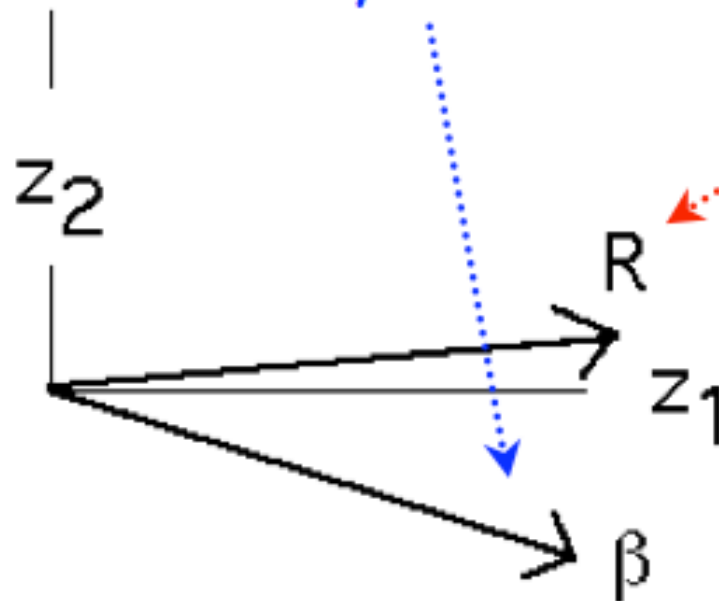What is the true nature of selection on the two traits?

$$\beta = \mathbf{P}^{-1}\mathbf{S} = \mathbf{P} = \begin{pmatrix} 20 & -10 \\ -10 & 40 \end{pmatrix}^{-1} \begin{pmatrix} 10 \\ -10 \end{pmatrix} = \begin{pmatrix} 0.43 \\ -0.14 \end{pmatrix}$$

What does the actual response look like?

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta} = \begin{pmatrix} 20 & 5 \\ 5 & 10 \end{pmatrix} \begin{pmatrix} 0.43 \\ -0.14 \end{pmatrix} = \begin{pmatrix} 7.86 \\ 0.71 \end{pmatrix}$$

Direction favored
by selection

Direction of
response

z_2

R

z_1

β

Multivariate Constraints to Response

Is there genetic variation in the direction of selection?

Consider the following G and β:

$$\mathbf{G} = \begin{pmatrix} 10 & 20 \\ 20 & 40 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Taken one trait at a time, we might expect $R_i = G_{ii}\beta_i$

Giving R_1 = 20, R_2 = -40.

What is the actual response?

$$\mathbf{R} = \mathbf{G}\beta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

38

Key ideas

– One cannot predict multiple-trait response simply by applying the breeders equation separately to each trait.

– The **G** and **P** matrices quantify the effects of genetic and phenotypic correlations.