

- Artificial selection (AS) is the main agent for the genetic improvement of plants and animals
- Process: Use only the most suitable individuals as reproducers in relation to a specific objective
- Also a tool for experimental research: can alter a phenotypic trait permanently and this response to selection is from changes in allele frequencies at the loci affecting that trait
- The initial response to AS can be used to estimate the genetic variance of the selected trait in the population. It also produces changes in correlated traits, allowing the estimation of genetic covariances between different characters (or the same char in diff. environments)
- From long term response we can infer the number of genes involved and their initial frequencies (and possible interactions between natural and artificial selection)
- Selection in a population lacking genetic variation can enable the estimation of mutation rate, effects, dominance relationships, and pleiotropic effects
- Application: on phenotypes, but response can be improved when integrating information from relatives and molecular genetic markers
- Response to selection (and its prediction):
 - Consider the selection of a trait based on phenotype assuming the infinitesimal model: the distribution of phenotypic values is normal with mean \bar{X} and variance σ_p^2
 - Measure the response to selection as the change in population mean after one generation
 - regress average values of a character trait for pairs of parents onto those of their offspring
 - Slope of this regression line is the estimate of heritability
 - truncation selection: just select the parents with the highest phenotypic values, and the mean of the trait value of these parents deviated from the overall mean is the selection differential (S)
 - The mean of the offspring from selected parents (deviated from the population mean) is the response to selection R
 - We get the expression $R = h^2 S$ known as the breeder's equation
 - Regression to mediocrity: the average value for a trait of selected offspring is lower than that of their selected parents, and this is because heritability is always less than 1, with some of the superiority of the parents being attributable to environmental influences
 - The smaller number of parents used, the greater the selection differential
 - The response to selection can be predicted without doing the selection if estimates on heritability and phenotypic variance are available (and assuming normality)
 - Standardize $R/\sigma_P = h^2(S/\sigma_P)$ leads to the expression for response to selection $R = i h \sigma_A$ where i is the standardized selection differential and σ_A^2 is the additive variance for the trait
 - Because of the normal distribution, the average value of selected individuals can be expressed as $i = z/\phi$, where z is the value of the ordinate and ϕ is the truncation point for a selected proportion
 - But in a population with finite census size, selection intensity is lower than that from the expression above (caveat: inbreeding generated by selection implies a reduction in long-term response)
 - Sex considerations: must be averaged across sex, or if only one sex is considered then $i = i_f/2$

- From the standardized selection differential, we see it is key to select a low proportion of individuals as parents (applied: few or 1 sire and many dams), or alternatively we can reduce the phenotypic variance of the population (by either reducing environmental variation or experimental error)
- Overlapping generations: present the response per year $R_a = R/I_g$ where I_g is the generation interval
- Correlated responses
 - We can see correlated responses when the trait that we select for is correlated with other traits (r_A), resulting in an expression for coheritability which is $R_{CY} = i_X r_A h_X h_Y \sigma_{PY}$
- Measure of the response
 - Extending the response to selection over a few generations, the change in mean follows a linear trajectory
 - Means between generations are subject to change from chance, measurement errors, environmental factors or the genetic composition of the population (but assuming randomness these should not produce a bias)
 - A bias could be produced from sustained environmental changes or adaptation by natural selection. To avoid: maintain control lines without selection
 - If the response is linear, we can quantify it as the regression of the mean of each generation on the ordinate, creating a descriptive statistic for the summary of response but not the magnitude of applied selection
 - To incorporate the magnitude of selection, we calculate the total response as the regression coefficient of the cumulative response on the selection differential accumulated over generations (termed realized heritability h_r^2)
 - If several lines are available to apply selection, then we can obtain estimates for the empirical variance of selection response. Recall that the variance of line means for an additive trait (neutral model) is approximated by tV_A/N_e when $t \ll 2N_e$, but the variance of measurement errors is V_P/n , so we can compute the expected value for the variance of the phenotypic response $E[V(R)] = V_P(\frac{th^2}{N_e} + \frac{1}{n})$, useful for getting the sampling variance of the realized heritability
 - Standard errors of estimated realized heritability are obtainable from the standard error of the regression but this is biased downwards since the means of different generations are correlated
 - But estimations of realized heritability from REML approaches accounts for this correlation, accounts for fixed effects, genetic changes and dominance effects
 - Epistatic variation: realized heritability and selection response are functions of additive x additive components, so that the realized heritability becomes $h_r^2 = (V_a + 0.5V_{AA})/V_P$ ignoring higher order interaction terms
- Asymmetry of response
 - Responses to selection can be asymmetric due to genetic drift (especially with low N_e), genetic basis of the character (whether it is due to few loci of large effect, if allele frequencies are not intermediate, or dominance), due to inbreeding depression (reducing the response effect, ID can be partially eliminated using an unselected line of the same census size as the selected line as a control), maternal effects, changes in phenotypic variance, and opposition of natural selection to change in the mean of one direction of selection but not the other
 - Selection that lowers fecundity or increases mortality rates of offspring will negatively bias selection differentials, detectable by calculating selection differentials weighing value of parents by contribution to offspring
 - asymmetry most often occurs when a selected trait is component of fitness since the response to decrease the mean is accompanied by inbreeding depression

- Furthermore, deleterious alleles that are at low frequencies in the initial population will show larger responses to selection
- Change of allele frequency due to selection
 - Changes in population mean of a trait is due to the change in allele frequencies of the loci that affect that character
 - But often loci cannot be identified and changes in frequencies cannot be tracked
 - Can relate the effect of an allele on a quantitative trait using the expected change in allele frequency by selection
 - Allele frequencies change by natural selection according to the selective advantage factor s as average fitness approaches, or $\Delta q = spq/2$
 - Selection coefficient against a genotype is $s = \frac{\alpha_i}{\sigma_P}$
 - The above expression allows us to calculate the probability of fixating an allele with artificial selection using Kimura equation which is a function of s , N_e and N
 - Given that the contribution to the mean of a locus is $M = \alpha Q$
- Effect of selection on genetic variance
 - the Bulmer effect: a transient decrease in genetic variance in the first few generations of selection. This reduction arises from the fact that selected individuals correspond to extremes, meaning reduced variance.
 - The reduced additive variance $V_A^* = V_A(1 - kh^2)$
 - The interpretation of the reduced additive variance lies in negative LD, since selected individuals will be more extreme and similar. By virtue of carrying alleles that increase a trait, individuals also carry more alleles that reduce the trait, generating a negative LD with magnitude $-kh^2$.
 - The disequilibrium is lost by half every generation, because every generation offspring are produced in full-sib families that have genetic correlations of $1/2$. Half of genetic variance is within families and half is between families, and the total variance in the offspring $V_{AW} + V_{AB} = V_A(1 - 0.5kh^2)$
- N_e with artificial selection
 - In selection scenarios, we can present N_e as $\frac{N}{1+Q^2C}$, where C^2 is the variance of the selective advantage and Q is the drag of the selective advantage
 - G is the proportion of variance remaining after decay by genetic drift and selection, and ignoring drift, $G = 1 - kh^2$
 - The asymptotic effect size with selection is half the number of breeders
- Combined effect of genetic drift and selection on genetic variance
 - The total genetic variance in generation t is equal to $V_{A,t} = V_{A,t-1}(1 - \frac{1}{2N_e})$. This is called genic variance to distinguish from additive genetic variance in 9.11
 - Adding the reduction in variance by selection, the additive variance with selection per generation is obtained as $V_{A,t}^* = 0.5V_{A,t-1}^*(1 - 1/N_e)(1 - kh_{t-1}^2) + 0.5V_{A,t-1}$
 - Must be noted that N_e calculated by these equations strictly corresponds to the generation for which it is calculated (since N_e is reduced over generations)
- Long term response of selection
 - Short term response to selection only depends on intensity of selection, heritability and phenotypic variance
 - But in the long term, the response to selection is dependent on initial allele frequencies, effects of genes, effective population size and mutation. Response to selection is also dependent and greater with the number of loci that affect the trait segregating in the population
 - It is actually possible to make an estimate of the number of loci affecting a trait using the divergence between two lines from the same base population selected in opposite directions

- If these lines were initially inbred, and each line is different for the trait, they are likely heterozygous. In the selection scenario, assuming high N_e and no genetic drift, and that for all loci the difference between the two homozygotes is the same (a), then the total divergence reached at the selection limit is $R_T = na$.
- Initial additive variance of the cross: $\sigma_A^2 = \sum_n (a^2/2)(0.5)(0.5) = na^2/8$, yielding that the divergence is equal to n (in other words, if we know the magnitude of the response and the initial additive variance we can estimate the number of loci, termed the effective number of loci)
- The effective number of loci is typically an underestimation if effects are not equal and alleles are not fixed (or an overestimate if the initial frequencies are not intermediate)
- This entire procedure of estimating the effective number of loci is the Castle-Wright method, and though generally the assumptions do not hold, it has some use in counting the number of loci with large effects in a cross
- Predicting long term response
 - Using the infinitesimal model, we can obtain an estimate for the theoretical maximum response of selection over the long term
 - Ignoring mutation and the bulmer effect, any reductions in additive variance are exclusively from genetic drift
 - The long term response to selection can then be approximated by $R_t = 2N_e R_1 F_t$, and that over many generations the cumulative response in the limit is $2N_e$ times the initial response (implying additivity)
 - Antagonism between short and long term response: in the first generation, response depends on selection intensity and can be increased by reducing the number of parents and effective size (restricting long-term response)
 - The maximum final response over the long term will occur when a proportion of $\phi = 0.5$ is selected
 - half-life of a response: the number of generations elapsed for the cumulative response to be half the response in the limit can be approximated by $R_t \approx 2N_e R_1 [1 - e^{-t/2N_e}]$
 - Impact of mutation: for a given trait, population and selection intensity, the response to selection is proportional to the additive variance in the mutation-drift equilibrium, given by $R_{M1,inf} = i\sigma_{A,inf}^2/\sigma_P$
 - The above is also the asymptotic response due to new mutations
 - The cumulative response to selection hits a plateau once the initial additive variance is exhausted, and further responses are only due to new mutations (becoming linear)
- Infinitesimal model versus empirical data
 - The greater the leptokurtosis of the distribution, the lower the response will be at the limit than that predicted by the infinitesimal model
 - The response can also be irregular with high variation between replicates of the selected line
 - Utility of the infinitesimal model: does well for theoretical predictions about the impact of an initial bottleneck
 - Reduction of selection response with linkage: pronounced only if genome size is very small, in which case the infinitesimal model struggles to hold up
 - More problems: phenotypic variance can increase as generations go on, and this can be somewhat corrected by calculating the coefficient of phenotypic variation (std. dev divided by the mean) but even so, the coefficient increases drastically. A possible explanation is that homozygous genotypes have higher environmental sensitivity and/or genotype-environment interactions
 - Selection can increase environmental sensitivity since the phenotypic of individuals with greater sensitivity will increase in better environments, and those are the individuals that are selected, increasing the sensitivity of the population as a whole

- Illinois experiment: sought to increase and decrease the percentage of protein and fat in corn grain for more than 100 generations, showed no limit to selection response
- Reasons for limits to selection response: depletion of genetic variation by fixation of selected alleles, or without exhausting genetic variation: favourable genes might be dominant in which case selection leads to high frequencies but will not be fixed, or overdominant genes can contribute null additive variance (and again no selection response)
- Most common reason for selection limit: natural selection is opposed to artificial selection. It is likely that lethal or deleterious genes with pleiotropic effects increase in frequency with selection
- Ways to quantify and work around the above problem: compare the selection differential weighted by the number of offspring compared to the unweighted value, or we can relax selection and see if this change is lost by natural selection
- Family, within-family and BLUP selection
 - Sometimes it is more prudent to carry out selection on entire families rather than individuals, or within families (when heritability is low, for milk production or fertility)
 - Selection criterion now is not the phenotype but the family mean, the deviation of the individual from the family mean, or a combined phenotypic value of individual and relatives, which generalizes to the equation $R = ir_{CA}\sigma_A$ where r_{CA} is the accuracy of selection criterion (calculated as the correlation between applied criterion C and the additive value A of an individual). For individual selection, this value is simply the square root of heritability
 - Conducting within-family selection: take the best male and female of each family, then compute C as the phenotypic value in relation to other individuals of the same sex
 - Within-family deviations: another method where each individual is selected based on the deviation of phenotypic value from average of family (here more than one individual from each sex can be selected)
 - Within family versus within family deviations: get the expression $r_{P_W,A} = \frac{h(1-r)}{\sqrt{1-t}}$. For full-sib families, $r = 0.5$, so the accuracy of within-family selection is half that of individual selection
 - If t is high the accuracy of within-family selection improves, and this is the advised method when there is high environmental variation common to families
 - Selecting complete families: the accuracy of family selection is half that of individual selection but increases when n is high and t is low, or when heritability and c -squared is low (meaning that the noise from low heritability is compensated with the average values of individuals in the family). In short, family selection is good when heritability and common environmental variation are low.
 - Selection indices: increase selection response by using a combination of within- and between- family information. Example index: $I = b_w P_W + b_F P_F$ tracking deviations with a family and between families to the general mean
 - Getting coefficients to optimize response: maximize selection accuracy by minimizing sum of squares of linear regression of index I on additive values
 - In doing the above, family selection is good when $t \ll r$, within-family selection is good when $t \gg r$ and individual selection is good when $t \approx r$
 - Optimal index will always be achieved using BLUP which uses information in the matrix of additive relationships of individuals from all generations (good because we account for genetic changes across generations and fixed effects)
 - Comparisons of all methods thus far: optimized selection yields maximal response for short term, family response is better than individual initially if t is low but

then drops sharply, and within-family selection yields a better response than family selection

- Long term response depends on accuracy, additive variance and N_e . So family selection implies a big decrease in N_e because entire families are lost, within-family equalizes family contributions, so effective size is equal to $2n$ breeding individuals. Individual, index and BLUP selection yields effective sizes intermediate to the previous 2.
- Long term response of within-family selection is greater than others: family selection is much lower
- Selection of multiple traits: use selection indices, but again more precise using BLUP
- Use of molecular markers
 - Marker assisted selection: additive effects of markers associated with a trait are condensed into a molecular value for each individual, treat this as a character correlated with a trait of interest
 - MAS can be applied to BLUP yielding marker-assisted BLUP: good when phenotypic information is not available
 - Success of marker-assisted selection is modest: association between marker and loci with an actual effect can be spurious and is lost by recombination. Extra response is reduced with generations of selection because variance of marker-assisted loci decrease as allele frequency decreases. Long-term response can also be reduced if selection decreases N_e .

Papers

- Vilas et al 2015: Allelic diversity for neutral markers retains a high adaptive potential for quantitative traits than expected heterozygosity
- Dudley and Lambert 2004: 100 generations of selection for oil and protein in corn
- Yoo 1980: Long-term selection for a quantitative trait in large replicate populations of *Drosophila melanogaster*
- Lander and Botstein 1989: deduce the presence or absence of loci with large effects
- Lopez-Fanjul 1977: optimal selected proportion to produce a maximum response after a certain number of generations
- Hill and Robertson 1966: selection with infinitesimal model under linked loci
- Long-term selection: Goodale 1938 until Wilson et al. 1971
- Wilson 1971: increase in phenotypic variation over generations of selection
- Mulder 2008, Jill and Mulder 2011: reduce phenotypic variation between individuals by selecting for homogeneity given genetic variation for residual variation
- Hill 1996: compare within-family selection and within-family deviations

Problems

- (1) The number of eggs laid in one hour by 85 females of *Drosophila melanogaster* has been evaluated, a trait with an approximate heritability of $h^2 = 0.2$. If the offspring of 20 percent of females with the greatest egg laying were selected to form the next generation, what would be the expected mean of the population in that generation?
See ch9.r for the answer to this question.
- (2) In a selection experiment, 60 families of 10 sibs (5 of each sex) are evaluated in each generation for a trait with phenotypic variance 24 and heritability 0.3. Of the 300 individuals evaluated for each sex, the 60 with the highest phenotypic value are selected as parents for the next generation. What is the expected selection response in the first generation and the accumulated one in the initial 5 generations?

$$\phi = 60/300 = 0.2$$

$$i = z/\phi = 0.2881/0.2 = 1.44$$

$$R = ih^2\sigma_P = 1.4(0.3)(\sqrt{24}) = 2.06$$

$$t(R) = 5(R) = 10.3$$

- (3) In the selection experiment described in the previous problem, what would be the effective size of the selected population and the standard error of the accumulated response?

$$x \approx 0.82$$

$$k = i(i - x) = 1.4(1.4 - 0.82) \approx 0.781$$

$$C^2 = i^2(h^2/2) = 1.4^2(0.3/2) = 0.294$$

$$G = 1 - kh^2 = 1 - 0.781(0.3) = 0.766$$

$$Q = 2/(2 - 0.766) = 1.62$$

$$N_e \approx \frac{N}{1 + Q^2C^2} = \frac{300}{1 + 1.62^2(0.294)} = 67.72$$

$$E[V(R)] = V_P[(th^2/N_e) + (1/n)] = (24)[(5 * 0.3/67.72) + 0.10] = 2.93$$

$$SD = \sqrt{E[V(R)]} = \sqrt{2.93}$$

- (4) Continuing with 9.2 and 9.3, what would be the cumulative response up to generation 100, considering or not the contribution of mutation, if the mutational variance for the trait is $V_M = 0.002$?

$$F_{100} = 1 - (1 - 1/(2N_e))^{100} = 1 - [1 - 1/(2 * 67.72)]^{100} = 0.523$$

$$\sigma_{A,0}^2 = V_P h^2 = 24(0.3) = 7.2$$

$$R_T = (2N_e i \sigma_p)[\sigma_{A,0}^2 F_t + \sigma_M^2(t - 2N_e F_t)] =$$

$$2(67.72)(1.4)/\sqrt{24} * [(7.2 * 0.523) + 0.002(100 - 2 * 67.62 * 0.523)] = 148.12$$

$$R_t = 2(67.72)(1.4)/\sqrt{24} * [(7.2 * 0.523)] = 145.86$$

- (5) In the selection experiment in 9.2, individual selection was carried out. What would be the expected response to one generation of selection of within-family selection or family selection is carried out?

$$r_{CA} = h = \sqrt{0.3}$$

$$t = h^2/2 = 0.15$$

$$r = 0.5$$

$$n = 5$$

$$r_{PWA} = \frac{h(1-r)}{\sqrt{1-t}} = \sqrt{0.3} * \frac{1-0.5}{\sqrt{1-.015}} = 0.297$$

$$r_{FA} = \frac{h[1 + (n-1)r]}{\sqrt{n[1 + (n-1)t]}} = \frac{\sqrt{0.3} * [1 + (5-1) * 0.5]}{\sqrt{5[1 + (5-1)0.15]}} = 0.581$$

$$R = ir_{CA}\sigma_A = 1.4(0.297)\sqrt{7.2} = 1.11$$

$$R = ir_{FA}\sigma_A = 1.4(0.581)\sqrt{7.2} = 2.18$$

Self Assessment

- (1) For a given heritability and phenotypic variance, the response to artificial selection will be greater the greater the dominance variance.
False, the Response to selection is calculated by $R = ih^2\sigma_P$, which accounts for heritability and phenotypic variance but not dominance variance.
- (2) To estimate realized heritability, it is necessary to carry out two or more generations of selection.
False, realized heritability is an estimate of the heritability of a trait in the starting population, and this can be computed using values from 1 generation of parents and offspring.
- (3) The scale effects on the phenotypic variance may be responsible for the asymmetric responses.
True, phenotypic variance may change according to the direction of selection, which will also change the selection differential through changes in the estimate of realized heritability.
- (4) Since artificially selected parents belong to one extreme of the population phenotypic distribution, a negative LD is generated that reduces genetic variance of the offspring.
True, this is known as the Bulmer effect (transient reduction in additive genetic variance, which can be interpreted by negative LD for traits in selected individuals).
- (5) The Bulmer effect only affects the between-family component of the variance.
True. For Bulmer's effect, half of the genetic variance in offspring (assuming full-sub families) will be within families and this is not affected by phenotypic values in the absence of linkage. The other half of genetic variance is between families which is affected by the reduction in additive genetic variance.
- (6) If, given a fixed number of individuals evaluated, the selection intensity is increased, the response to long-term selection will increase.
- (7) If artificial selection is applied to a population without genetic variability, the cumulative response will be due only to mutation and it will have an initially convex and then linear form.
True, this is demonstrated on page 224 and figure 9.9.
- (8) The deleterious pleiotropic effects of selected loci may be one of the reasons for the existence of limits in the response to selection.
True, this was mentioned on page 227 where 1/3 of spontaneous mutations in a selection experiment for bristle number in *Drosophila* were lethals with pleiotropic effects.
- (9) Selection within families is recommended when heritability is low and the environmental variation common to members of the families is scarce.
False, we can see from figure 9.13 that response for selection within families is maximized when heritability is high and common environmental variance is also high.
- (10) Marker-assisted selection increases the short-term response but may reduce the long-term response.
True, the long-term response can be reduced because this response is reduced with generations of selection since variation of marker-assisted loci decreases as their allele frequency increases until fixation. Or, long-term response can decrease if MAS reduces N_e and thus long-term response on polygenic variation.