



Statistical
GenOmics Lab



Institute of Cytology
and Genetics

Блок 3 (или 4?): ОСНОВЫ КОЛИЧЕСТВЕННОЙ И ЭВОЛЮЦИОННОЙ ГЕНЕТИКИ

Яков Цепилов Александрович
кбн, нс лаб. Теоретической и Прикладной
Функциональной Геномики ИЦиГ СО РАН

Новосибирск - 2017

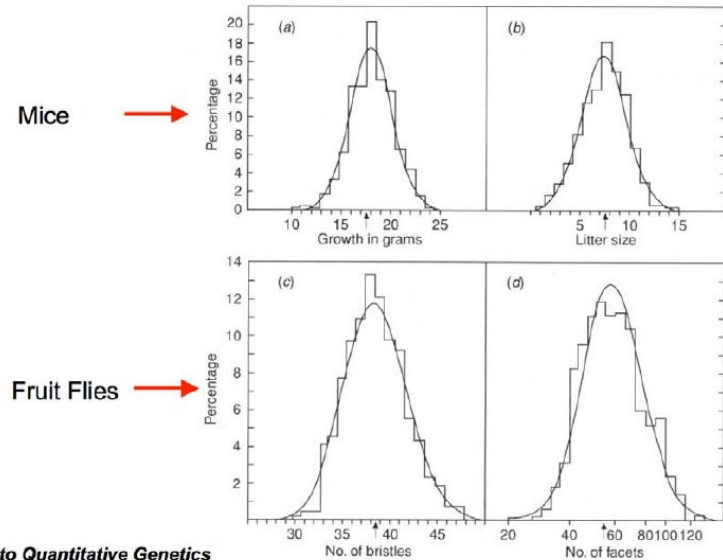
Примерный план на 1 часть (очень примерный)

- Признаки – количественные и качественные
 - Полигенная модель
- Наследуемость
 - Как измерять?
- Отбор
- Много признаков (чуть-чуть)
- Что там у человека?
- Геномная селекция

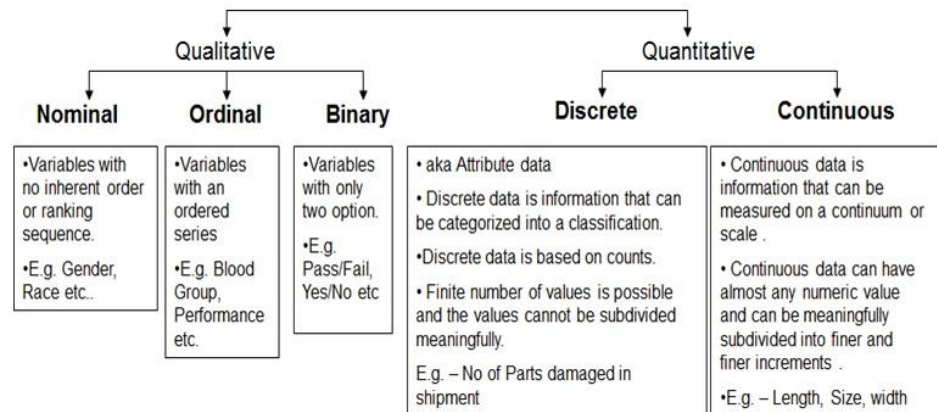
Признаки

- 1) Количественные/Качественные
- 2) Сильно или слабовариативные
- 3) Наследуются или не наследуются

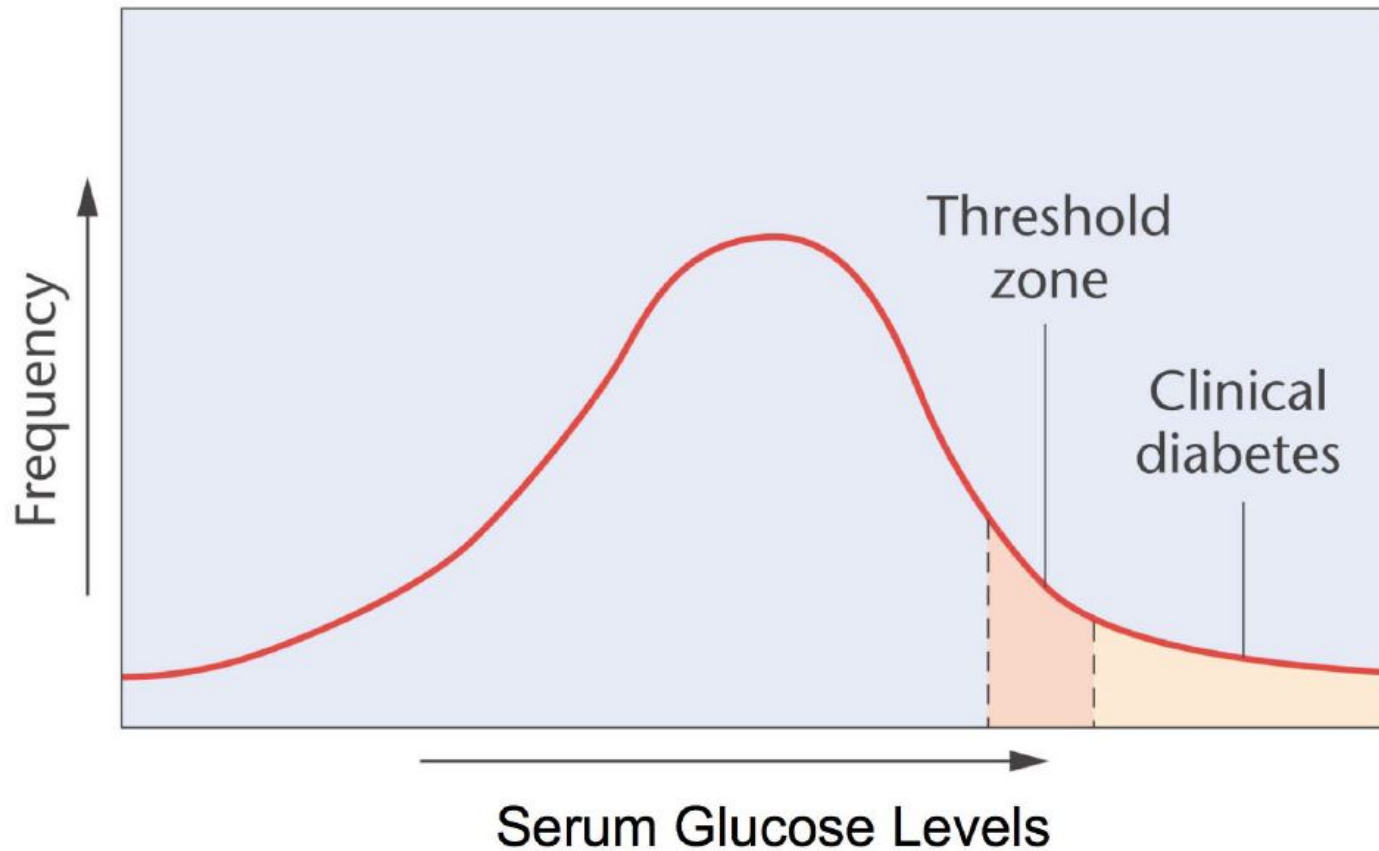
...



In: *Introduction to Quantitative Genetics*
Falconer & Mackay 1996



Качественные признаки глубоко в душе количественные



Придадим немного формализма

Самый простой случай: биаллельный полиморфизм

Familiar model

one allele is dominant (uppercase)

other allele is recessive (lowercase)

Additive model

Active allele (uppercase)

Inactive allele (lower case)

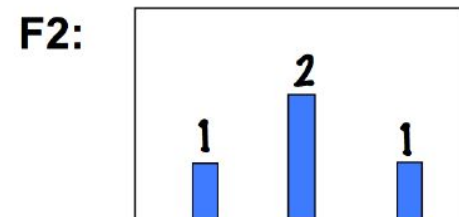
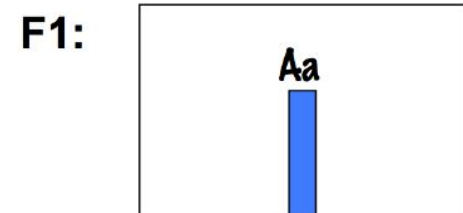
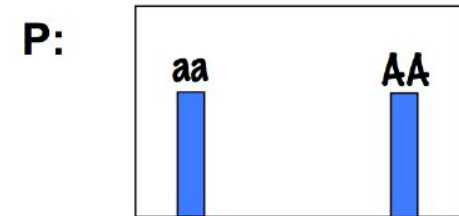


Расщепление по одному гену

If we do a $Aa \times Aa$ cross we would expect the following genotypes and phenotypes in the progeny:

Genotype	Number of active alleles	Frequency	Phenotypic value
AA	2	$1/4$	$\mu + 2x$
Aa	1	$1/2$	$\mu + x$
aa	0	$1/4$	μ

μ = average phenotype

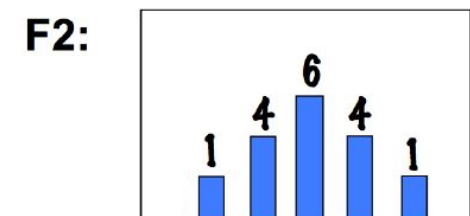
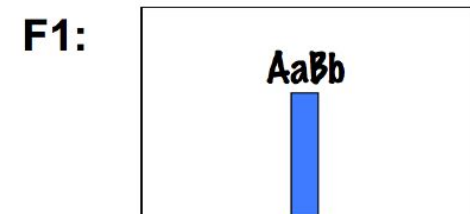
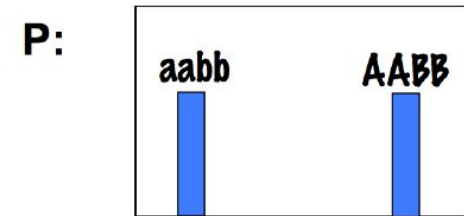


Phenotypic value

Расщепление в F2

If we do a $AaBb \times AaBb$ cross we would expect the following genotypes and phenotypes in the progeny:

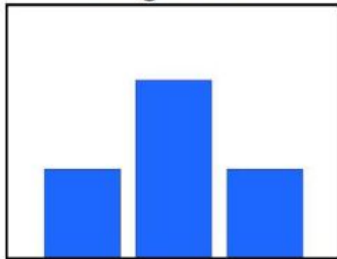
Genotype	Number of active alleles	Frequency	Phenotypic value
(1) AABB	4	$1/16$	$\mu + 4x$
(2) AABb, (2) AaBB	3	$4/16$	$\mu + 3x$
(4) AaBb + (1) Aabb + (1) aaBB	2	$6/16$	$\mu + 2x$
(2) Aabb + (2) aaBb	1	$4/16$	$\mu + x$
(1) aabb	0	$1/16$	μ



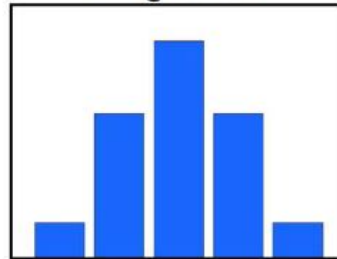
Phenotypic value

УХОДИМ ОТ ДИСКРЕТНОСТИ

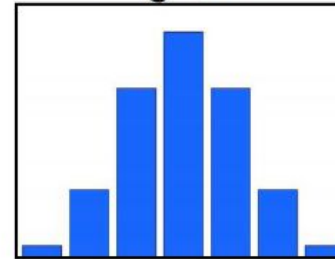
1 gene



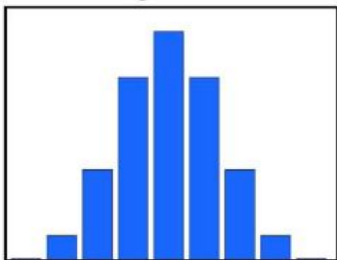
2 genes



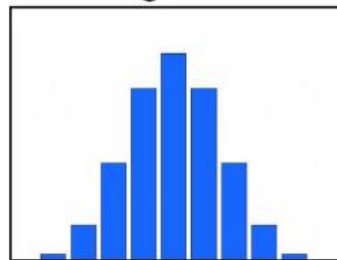
3 genes



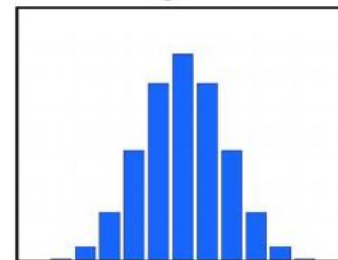
4 genes



5 genes



6 genes



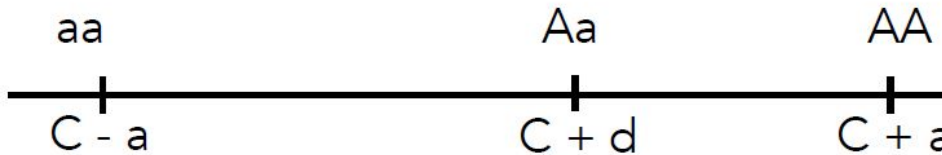
$2n+1$ phenotypic classes for n genes

Формализация модели 0

$$Y = G + E$$

Формализация модели 1

It will prove very useful to decompose the genotypic value into the difference between homozygotes ($2a$) and a measure of dominance (d or $k = d/a$)



Let $p = \text{freq}(A)$, $q = 1 - p = \text{freq}(a)$. Assuming random-mating (Hardy-Weinberg frequencies),

Genotype	aa	Aa	AA
Value	$C - a$	$C + d$	$C + a$
Frequency	q^2	$2pq$	p^2

$$\text{Mean} = q^2(C - a) + 2pq(C + d) + p^2(C + a)$$

$$\mu_{\text{RM}} = C + a(q-p) + d(2pq)$$

Contribution from homozygotes

Contribution from heterozygotes

Suppose two inbred lines are crossed. If A is fixed in one population and a in the other, then $p = q = 1/2$

Genotype	aa	Aa	AA
Value	$C - a$	$C + d$	$C + a$
Frequency	$1/4$	$1/2$	$1/4$

$$\text{Mean} = (1/4)(C - a) + (1/2)(C + d) + (1/4)(C + a)$$

$$\mu_{\text{RM}} = C + d/2$$

Note that C is the average of the two parental lines, so when $d > 0$, F_2 exceeds this. Note also that the F_1 exceeds this average by d , so only half of this passed onto F_2 .

Формализация модели 2

- Fisher (1918) decomposed the contribution to the genotypic value from a single locus as $G_{ij} = \mu + \alpha_i + \alpha_j + \delta_{ij}$
 - Here, μ is the mean (a function of p)
 - α_i are the average effects
 - Hence, $\mu + \alpha_i + \alpha_j$ is the **predicted genotypic value** given the average effect (over all genotypes) of alleles i and j .
 - The **dominance deviation** associated with genotype G_{ij} is the difference between its true value and its value predicted from the sum of average effects (essentially a residual)

И ВОТ ОНА

A key concept is the **Additive Genetic Value (A)** of an individual

$$A(G_{ij}) = \alpha_i + \alpha_j$$

$$A = \sum_{k=1}^n \left(\alpha_i^{(k)} + \alpha_k^{(k)} \right)$$

$\alpha_i^{(k)}$ = effect of allele i at locus k

A is called the **Breeding value** or the **Additive genetic value**

И дисперсии

$$\sigma^2(G) = \sum_{k=1}^n \sigma^2(\alpha_i^{(k)} + \alpha_j^{(k)}) + \sum_{k=1}^n \sigma^2(\delta_{ij}^{(k)})$$

Additive Genetic Variance
(or simply Additive Variance)

Dominance Genetic Variance
(or simply dominance variance)

Hence, total genetic variance = additive + dominance variances,

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2$$

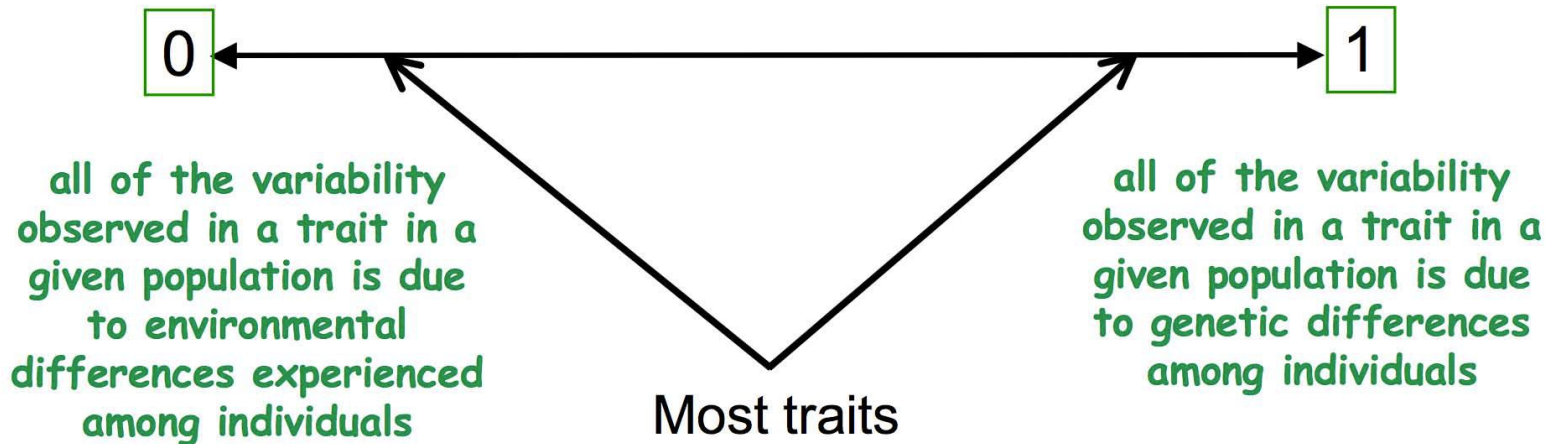
Ключевые концепции

- α_i = average effect of allele i
 - Property of a single allele in a particular population (depends on genetic background)
- A = Additive Genetic Value (A)
 - A = sum (over all loci) of average effects
 - Fraction of G that parents pass along to their offspring
 - Property of an Individual in a particular population
- $\text{Var}(A)$ = additive genetic variance
 - Variance in additive genetic values
 - Property of a population

Дальше проще - наследуемость

Heritability: the proportion of phenotypic variation among individuals in a population that is due to genetic variation

$$H^2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$



Некоторые важные МИСКОНЦЕПЦИИ

Да/Нет

- 1) У конкретного индивида очень высокое значение признака, значит его наследуемость очень высокая!
- 2) Наследуемость признака равна 0, значит генетика не важна для этого признака.
- 3) Наследуемость этой болезни равна 1, значит попытки как-то повлиять на болезнь через среду априори безрезультатны.

Некоторые важные МИСКОНЦЕПЦИИ

Да/Нет

- 1) У конкретного индивида очень высокое значение признака, значит его наследуемость очень высокая!
Нет, наследуемость это характеристика популяционная, не индивида!
- 2) Наследуемость признака равна 0, значит генетика не важна для этого признака.
- 3) Наследуемость этой болезни равна 1, значит попытки как-то повлиять на болезнь через среду априори безрезультатны.

Некоторые важные МИСКОНЦЕПЦИИ

Да/Нет

- 1) У конкретного индивида очень высокое значение признака, значит его наследуемость очень высокая!
Нет, наследуемость это характеристика популяционная, не индивида!
- 2) Наследуемость признака равна 0, значит генетика не важна для этого признака.
Нет, наследуемость ничего не говорит о генетической архитектуре, только о том, какая часть генетической вариативности влияет на признак
- 3) Наследуемость этой болезни равна 1, значит попытки как-то повлиять на болезнь через среду априори безрезультатны.

Некоторые важные МИСКОНЦЕПЦИИ

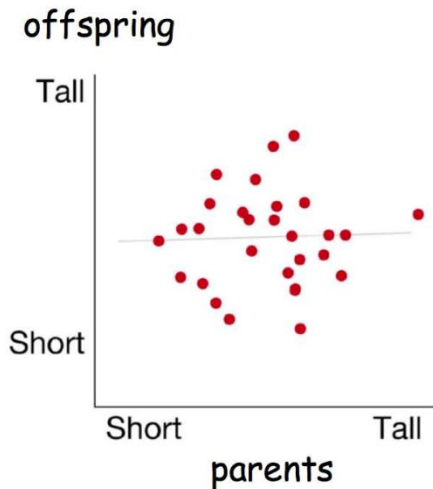
Да/Нет

- 1) У конкретного индивида очень высокое значение признака, значит его наследуемость очень высокая!
Нет, наследуемость это характеристика популяционная, не индивида!
- 2) Наследуемость признака равна 0, значит генетика не важна для этого признака.
Нет, наследуемость ничего не говорит о генетической архитектуре, только о том, какая часть генетической вариативности влияет на признак
- 3) Наследуемость этой болезни равна 1, значит попытки как-то повлиять на болезнь через среду априори безрезультатны.
Нет, оценки наследуемости в другой среде могут быть совершенно другими

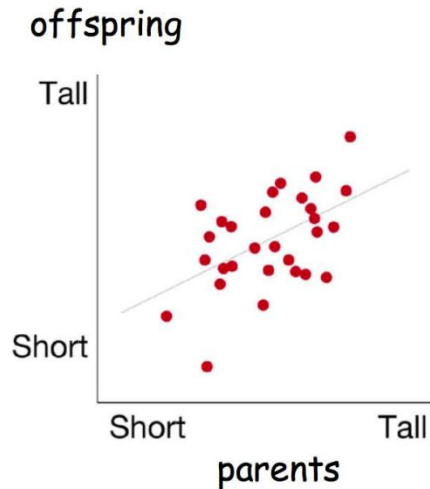
Как оценивать?

h^2 is the **regression** (slope) of offspring on parents

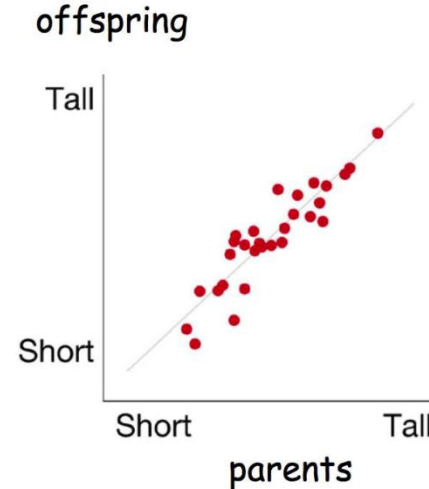
$$h^2 \approx 0$$



$$h^2 \approx \frac{1}{2}$$



$$h^2 \approx 1$$



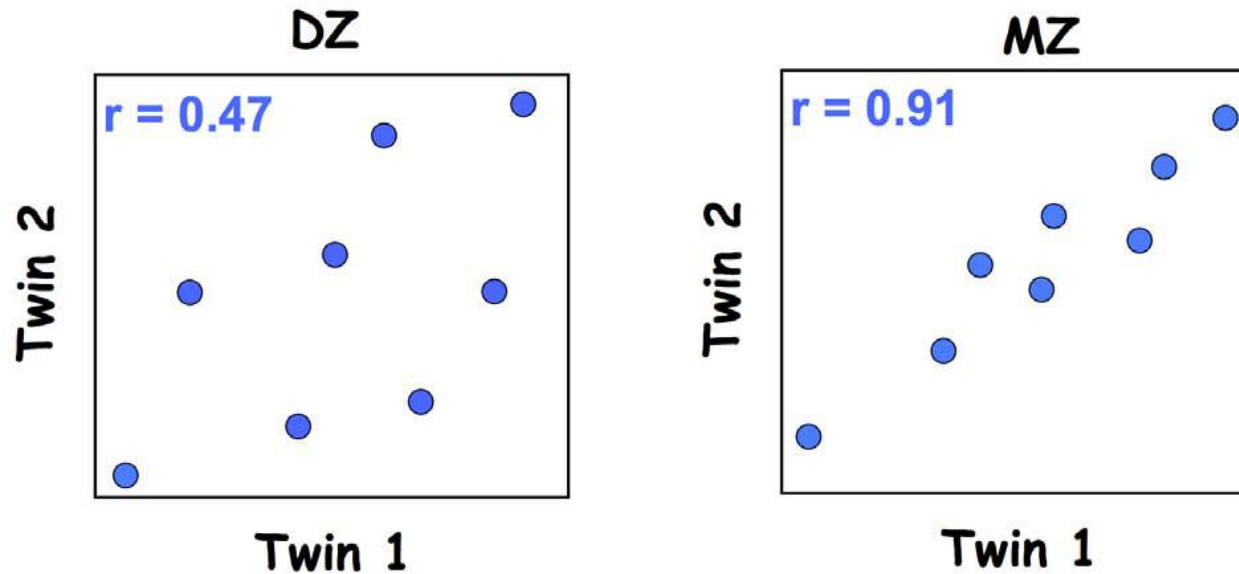
В общем случае: родственная структура – чем больше родственники похожи по определённому фенотипу – тем больше его наследуемость

Близнецовые исследования – «лучшие»:

МБ (MZ) – 100% генов одинаковые, дисперсия только средовая

ДБ (DZ) – 50% генов общие, вклад в дисперсию признака как от генов, так и от среды

Близнецовые исследования



$$H^2 = 2(r_{MZ} - r_{DZ})$$

$$H^2 = 2(0.91 - 0.47)$$

$$H^2 = 0.88$$

АДДИТИВНОСТЬ И НЕАДДИТИВНОСТЬ

$$V_G = V_A + V_D + V_I$$



Parent-offspring resemblance

Broad-sense heritability: $H^2 = V_G/V_P$

Narrow-sense heritability: $h^2 = V_A/V_P$

Отбор

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

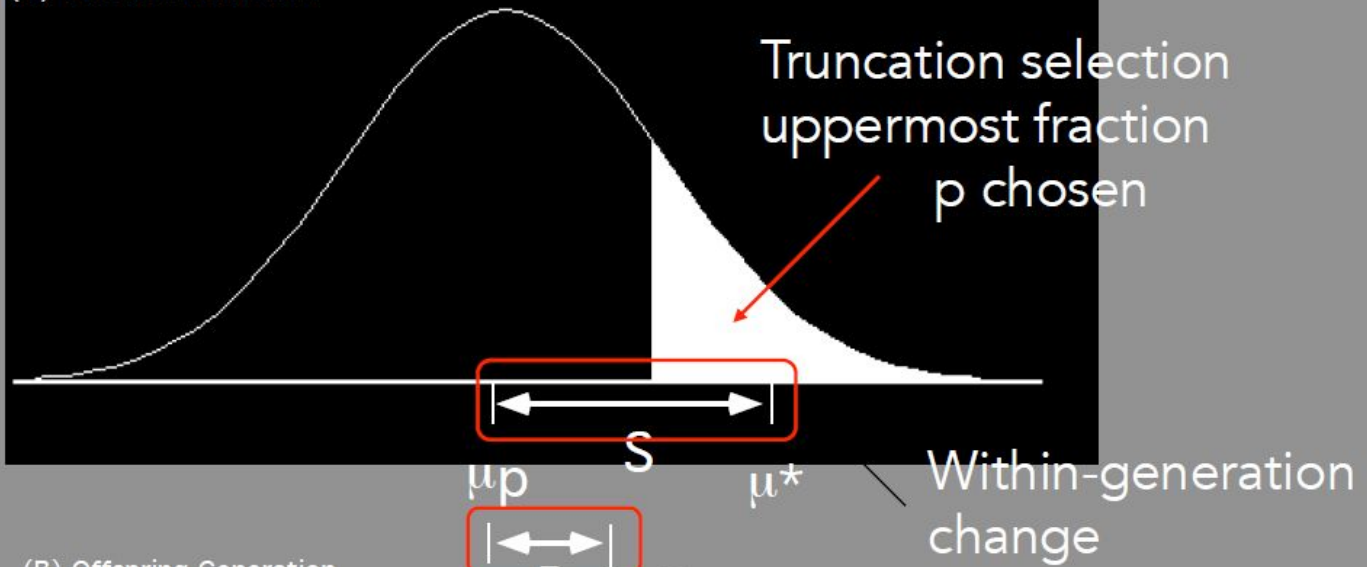
- The selection differential S measures the within-generation change in the mean

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

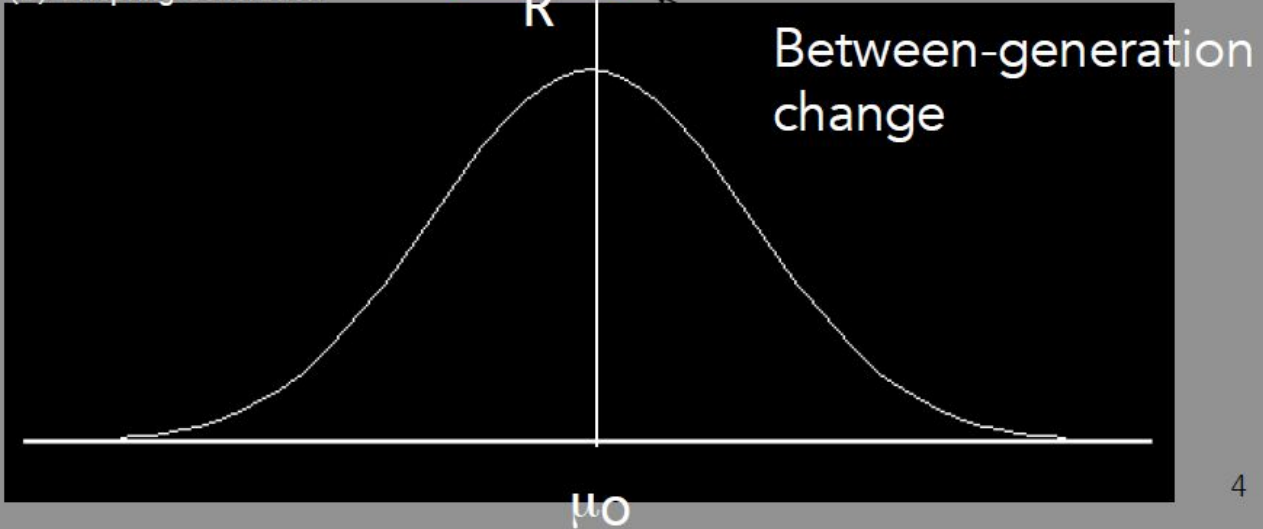
- The response R is the between-generation change in the mean

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

(A) Parental Generation



(B) Offspring Generation



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:

$$R = h^2 S$$

The Breeders' Equation (Jay Lush)

Средний эффект аллеля

- The average effect α_A of an allele A is defined by the difference between offspring that get allele A and a random offspring.
 - $\alpha_A = \text{mean}(\text{offspring value given parent transmits A}) - \text{mean}(\text{all offspring})$
 - Similar definition for α_a .

Consider the average effect of allele A when a parent is randomly-mated to another individual from its population

Suppose parent contributes A

Allele from other parent	Probability	Genotype	Value
A	p	AA	C + a
a	q	Aa	C + d

$$\text{Mean(A transmitted)} = p(C + a) + q(C + d) = C + pa + qd$$

$$\alpha_A = \text{Mean(A transmitted)} - \mu = q[a + d(q-p)]$$

Now suppose parent contributes a

Allele from other parent	Probability	Genotype	Value
A	p	Aa	C + d
a	q	aa	C - a

$$\text{Mean(a transmitted)} = p(C + d) + q(C - a) = C - qa + pd$$

$$\alpha_a = \text{Mean(a transmitted)} - \mu = -p[a + d(q-p)]$$

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