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

Зачем?

genetic
selection
variation
traits
environments
evolutionary
fitness
species
tolerance
natural
evolution
resistance
leaf
plant
populations
flowering
responses
population
ecological
growth
time
suggest
thaliana
interactions
effects
quantitative
evidence
important
adaptive
shape
data
differences
results
patterns
phenotypic
damage
genes
loci
Arabidopsis
across
significant
constraints
potential
whether
trait
response
also
QTL
found
estimates
environment
molecular
adaptation
experiment
rate
herbivory
frequency
potentially
within
low
temperature
conditions
mechanisms
covariances
gene
dynamics
recovery
may
differentiation
vernalization
FRI
common
predicted
affect
gradients
genotype
accessions
effect
hederacea
single
phenotypes
associated
correlations
performance
Ipomoea
among
studies
mean
density
ecology
clinal
information
role
model
although
lines
large
neutral
suggesting
cline
provide
levels
capensis
influence
presence
depending
soil
different
leaves
shade
plans
FLC
using
selective
biology
hypothesis
interaction
field
analysis
number
little
multivariate
consistent
insect
lead
alleles
likely
light
induced
plants
rhizobia
absence
analyses
plasticity
latitudinal
study
symbiotic
coevolution
treatments
glory
costs
herbivores
higher
clines
well
structure
avoidance
correlated
mutualism
show
change
locus
methods
North
many
Using
exploitative
approach
benefits
frost
use
often
dependent
variance
major
endangered
directions
significantly
detected
environmental
community
cold
either
floral
one
two
herbivore
underlying
genotypes
expression
size
negative
basis
approaches
However
pattern
photoperiod
indicate
high
coevolution
symbiotic

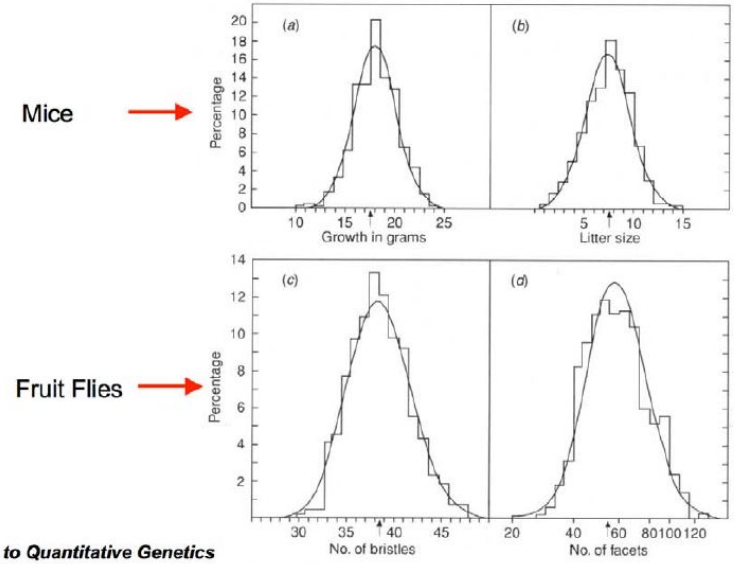
Примерный план на 1 часть (очень примерный)

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

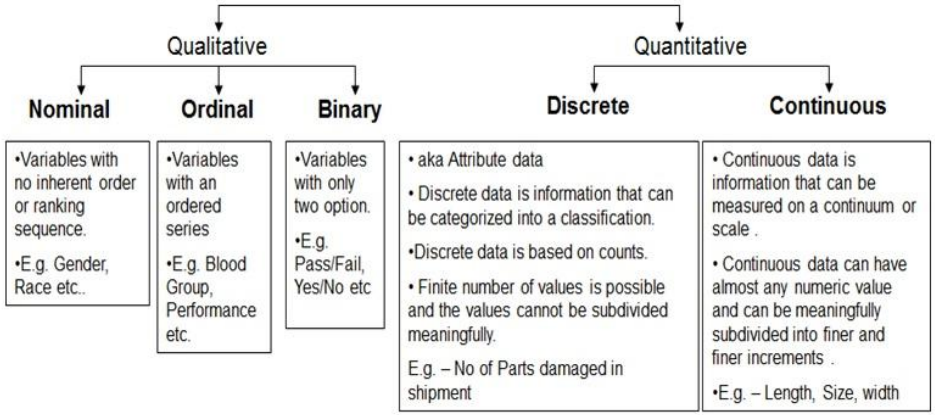
Признаки

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

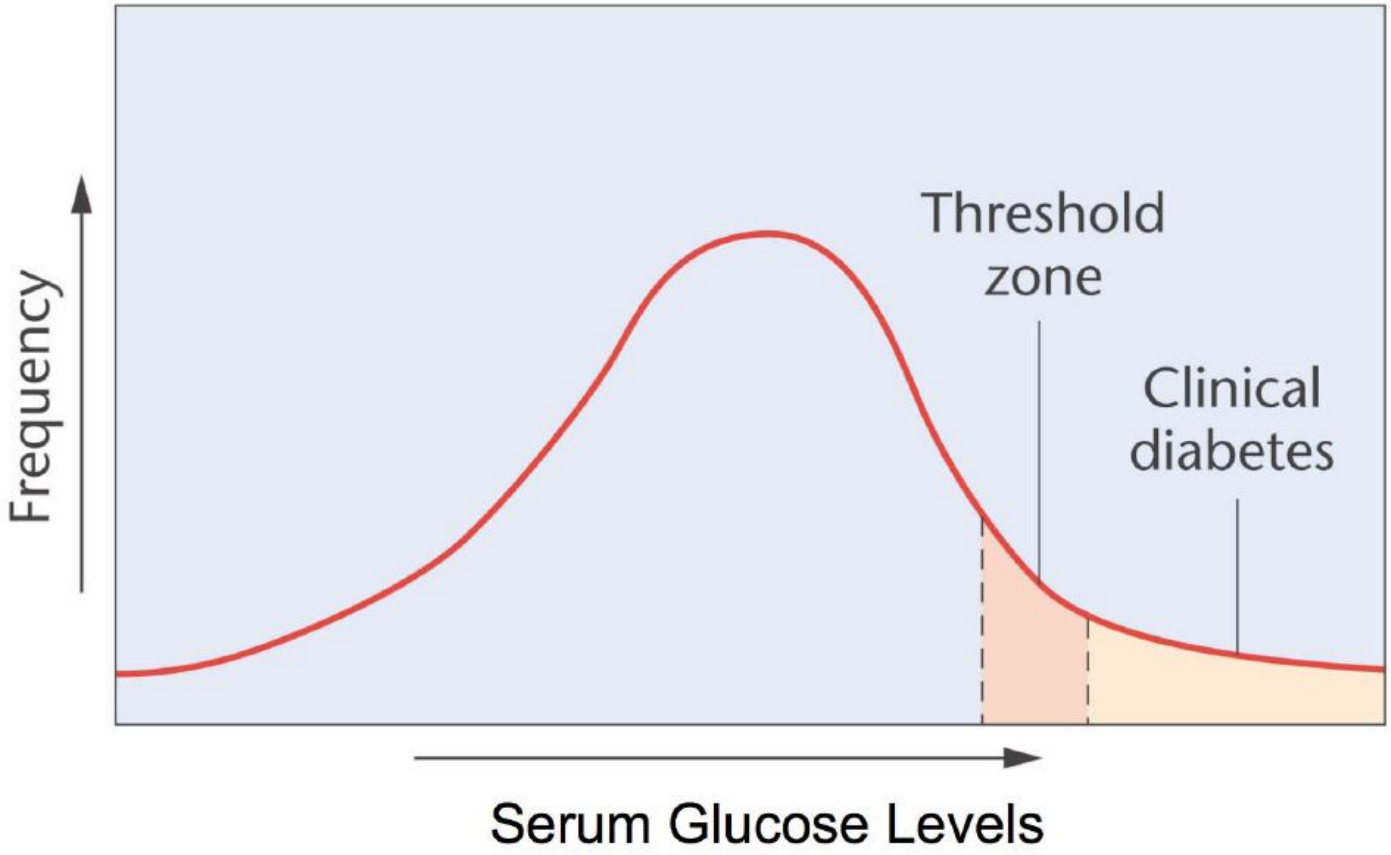
…



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



Качественные признаки глубоко в душе количественные



Придадим немного формализма

Самый простой случай: биаллельный полиморфизм

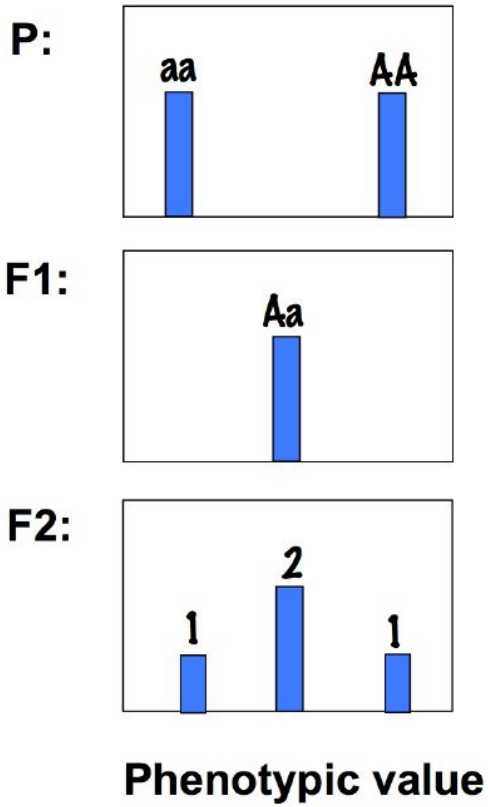
Familiar model	Additive model
one allele is dominant (uppercase)	Active allele (uppercase)
other allele is recessive (lowercase)	Inactive allele (lower case)

Расщепление по одному гену

If we do a Aa x 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

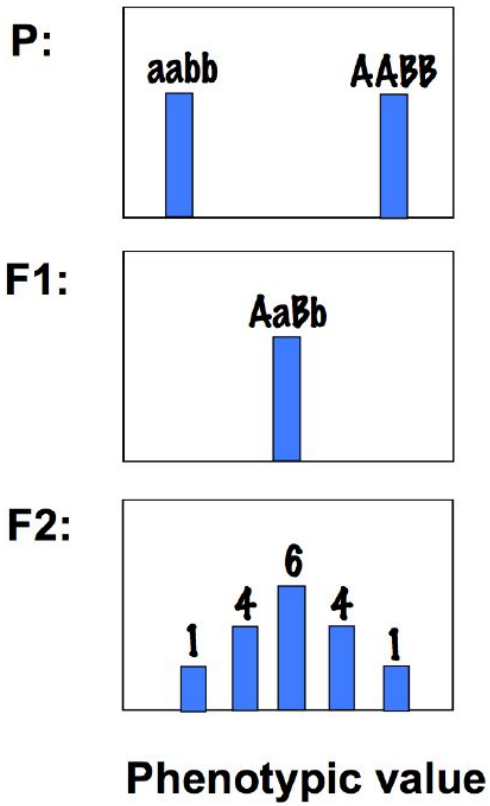
μ = average phenotype



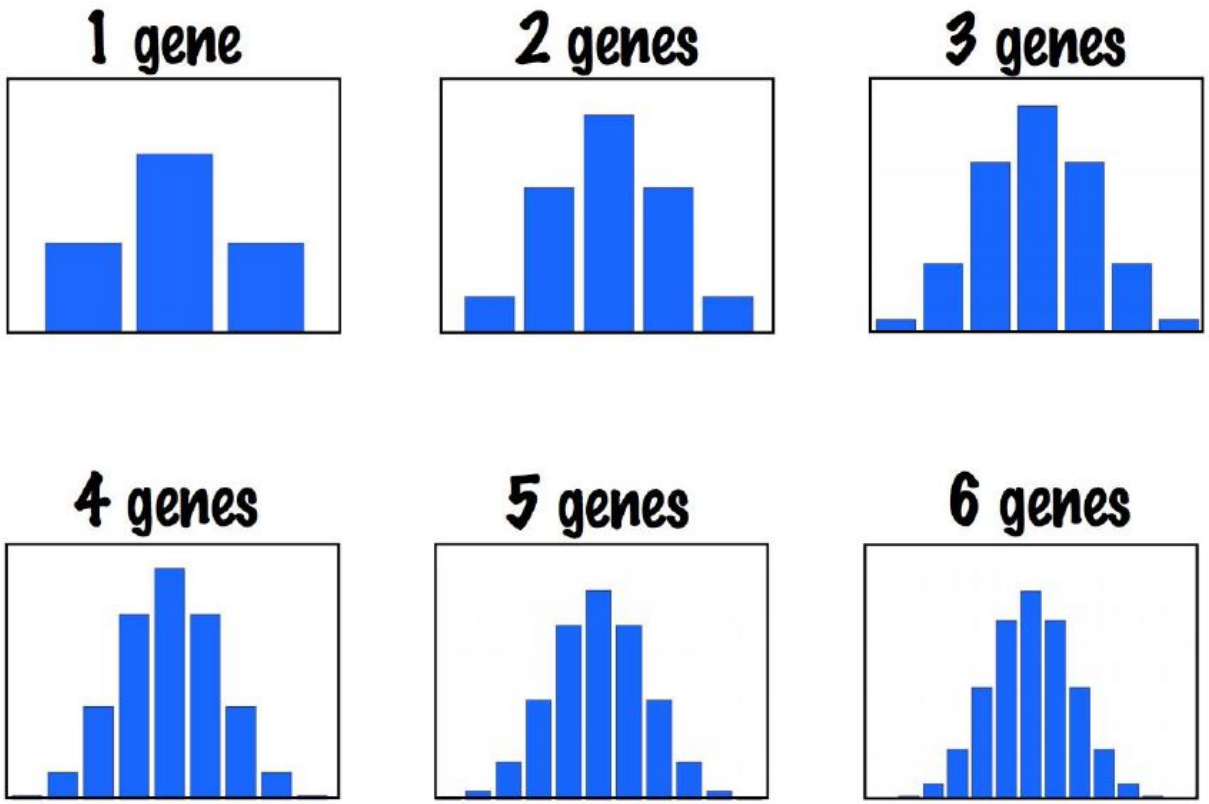
Расщепление в F2

If we do a AaBb x 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



Уходим от дискетности



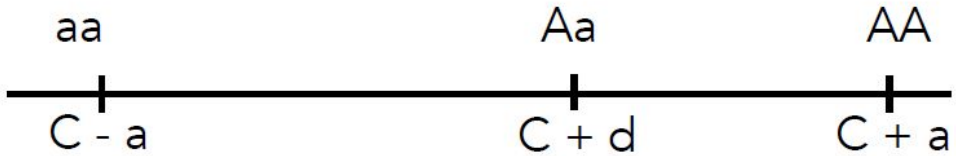
2n+1 phenotypic classes for n genes

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 = freq(A), q = 1-p = 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

Mean = $q^2(C - a) + 2pq(C + d) + p^2(C + a)$

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

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

Mean = $(1/4)(C - a) + (1/2)(C + d) + (1/4)(C + a)$

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

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,

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

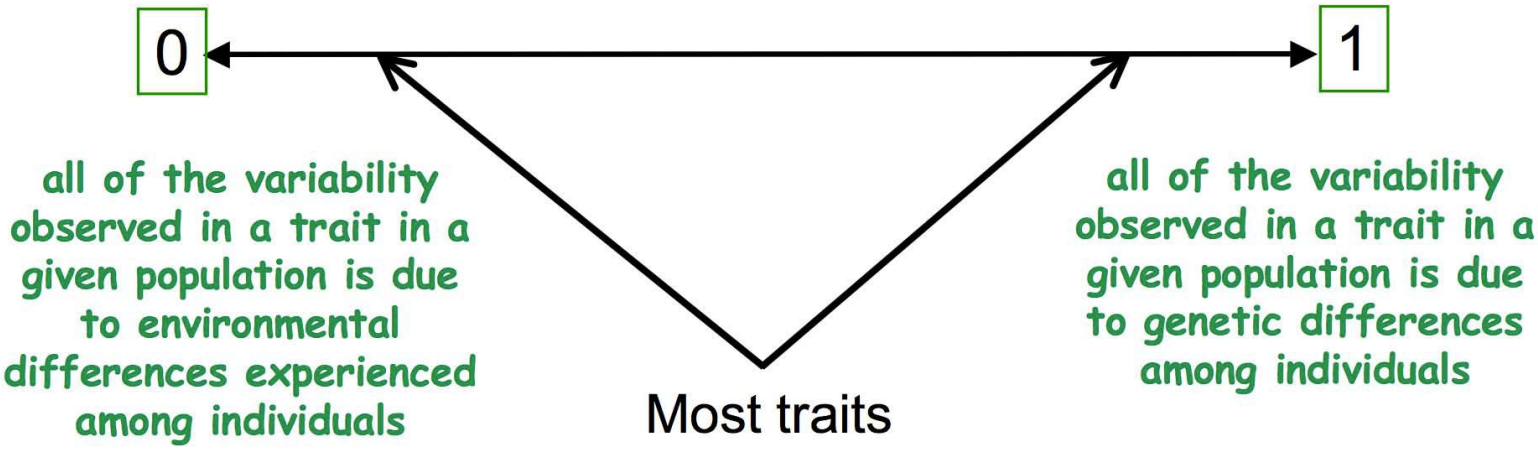
Ключевые концепции

- 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
- 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

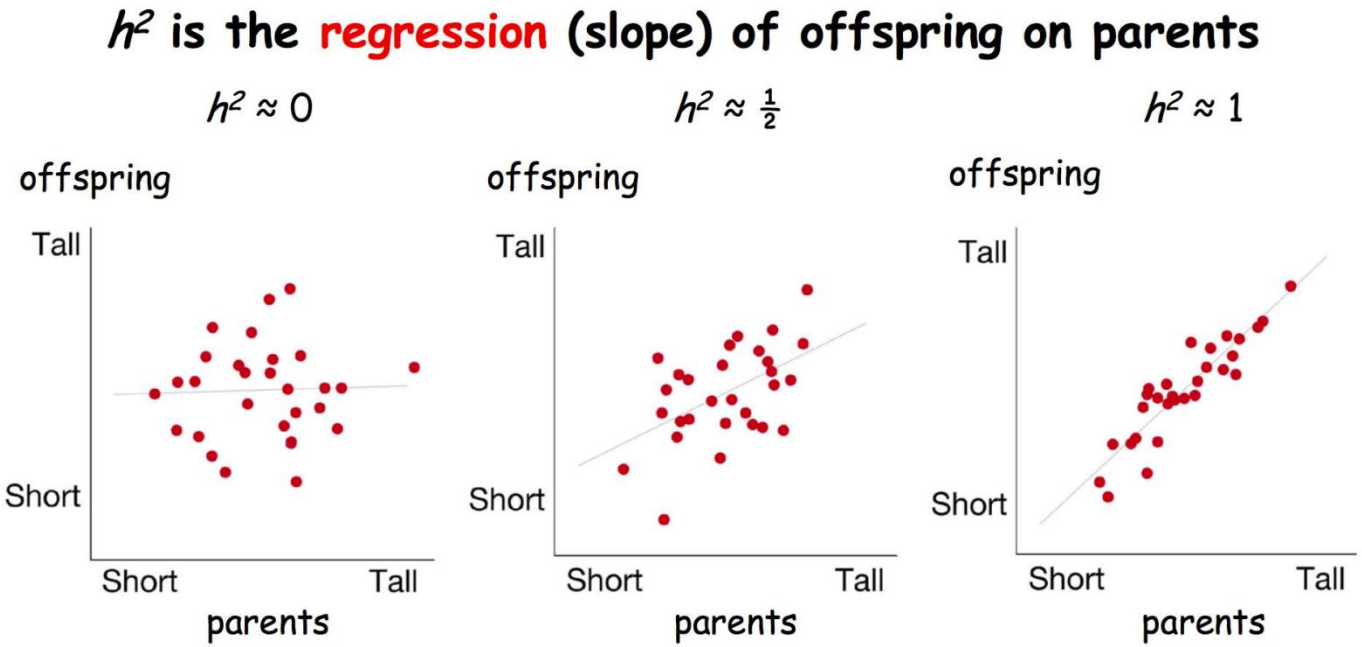
1) У конкретного индивида очень высокое значение признака, значит его наследуемость очень высокая!
   Нет, наследуемость это характеристика популяционная, не индивида!
2) Наследуемость признака равна 0, значит генетика не важна для этого признака.
   Нет, наследуемость ничего не говорит о генетической 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?



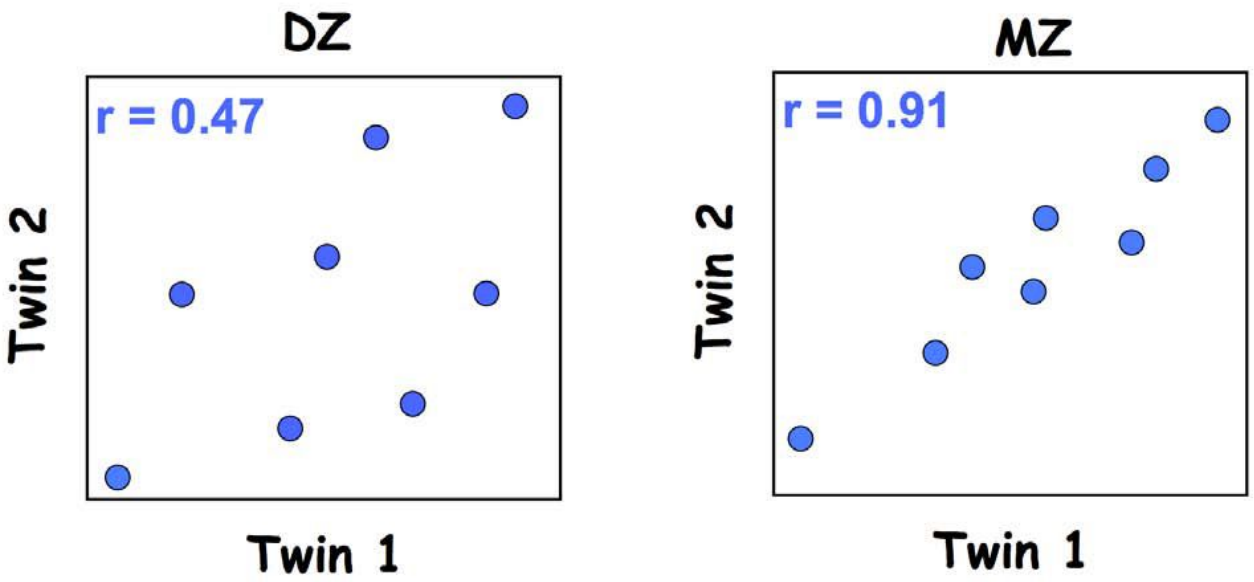
В общем случае: родственная структура – чем больше родственники похожи по определённому фенотипу – тем больше его наследуемость

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

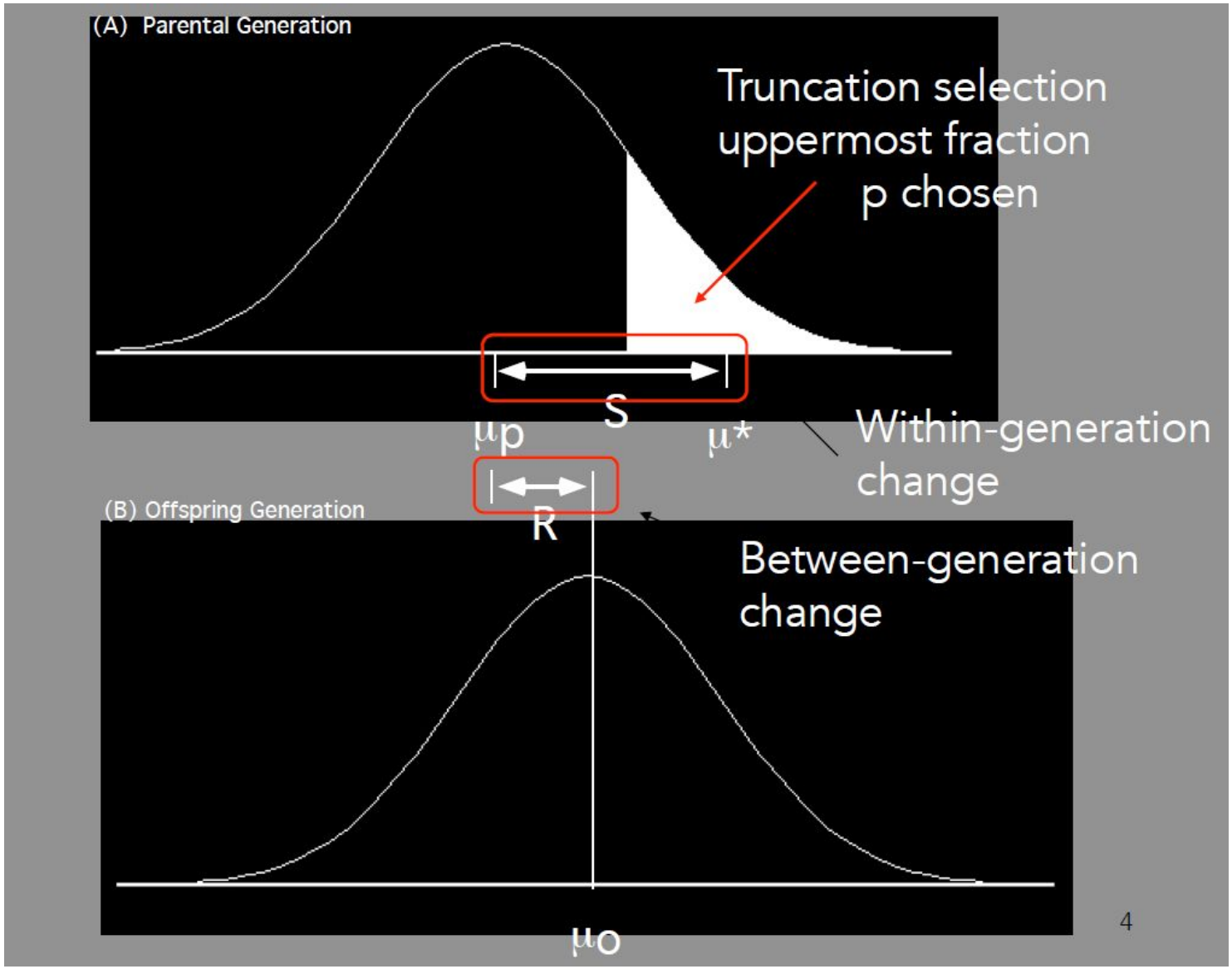
The Selection Differential and the Response to Selection

- 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
Truncation selection
uppermost fraction
p chosen
S
μp
μ*
Within-generation
change
R
(B) Offspring Generation
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)

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.
  - α_A = mean(offspring value given parent transmits A) - mean(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

Mean(A transmitted) = p(C + a) + q(C + d) = C + pa + qd

α_A = Mean(A transmitted) - μ = 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

Mean(a transmitted) = p(C + d) + q(C - a) = C - qa + pd

α_a = Mean(a transmitted) - μ = -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 h2 predicting response over several generations depends on:
- – The reliability of the initial h2 estimate
- – Absence of environmental change between generations
- – The absence of genetic change between the generation in which h2 was estimated and the generation in which selection is applied