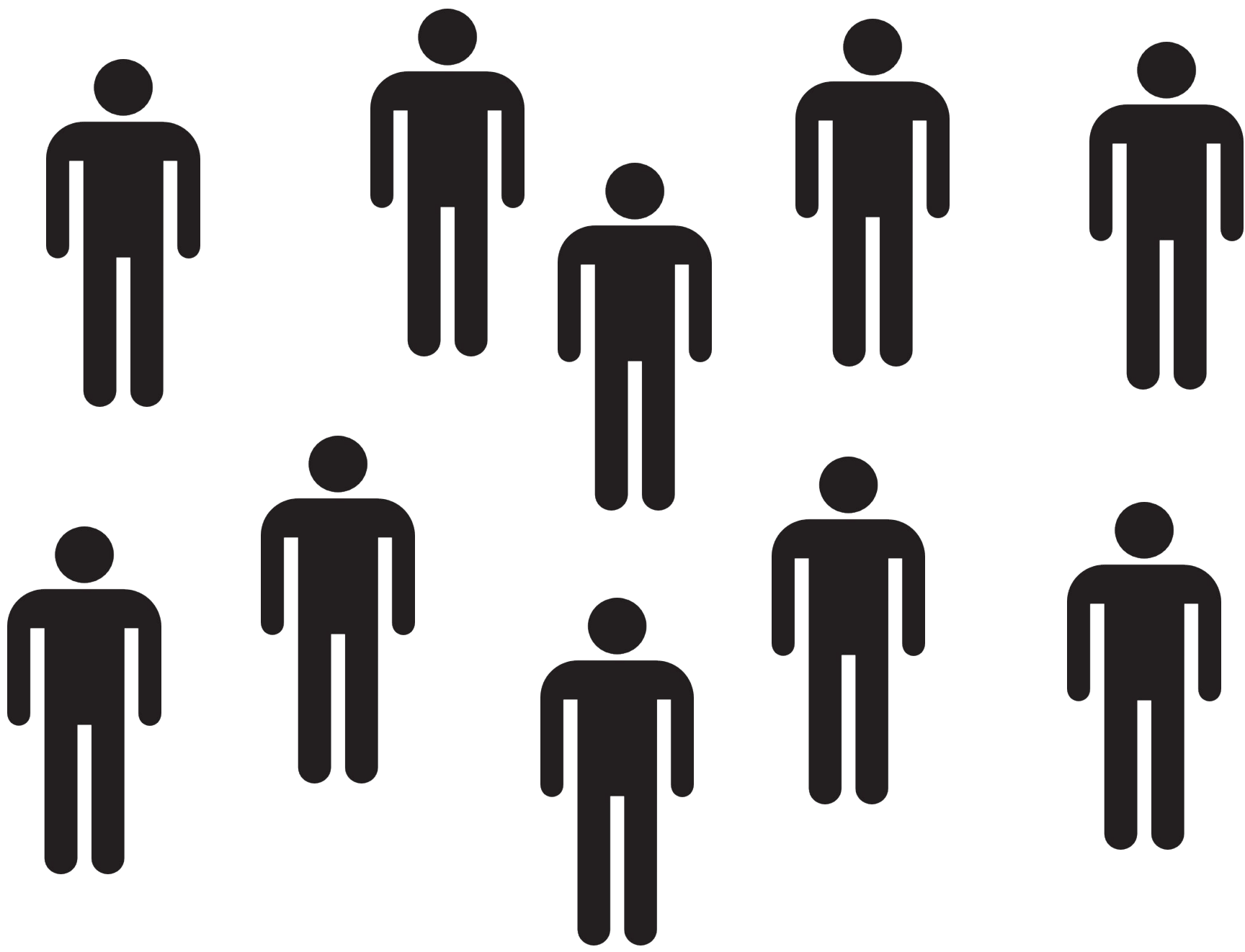
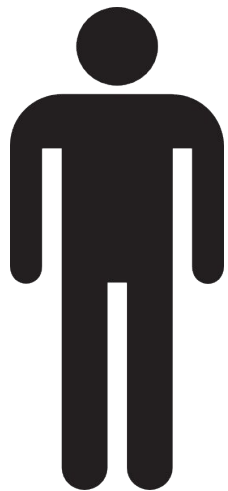


Генетическая структура популяции

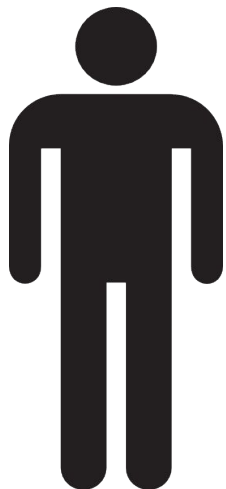
Описание генетической
структуры популяции



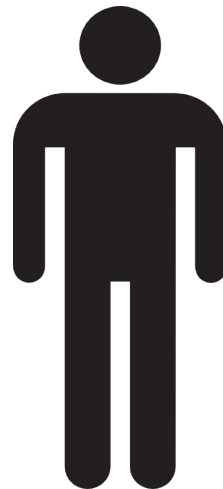
G/G



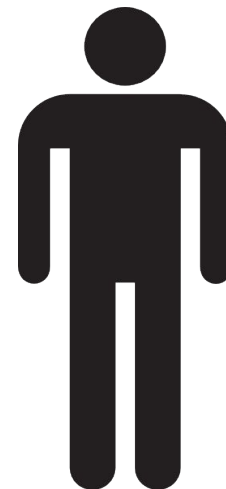
G/A



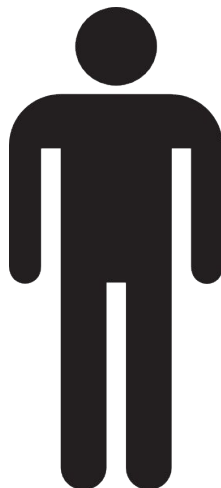
G/G



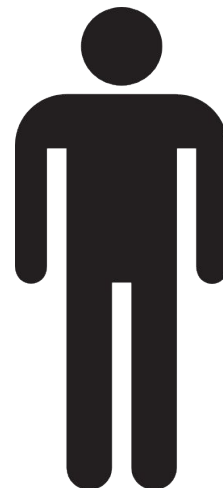
G/A



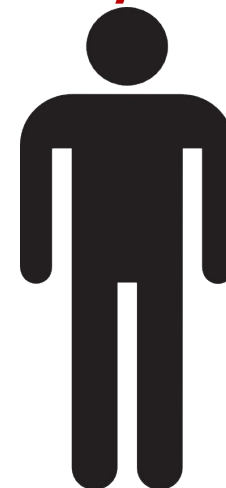
G/G



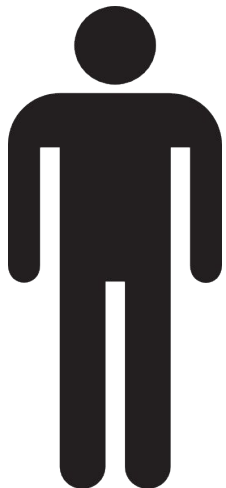
G/A



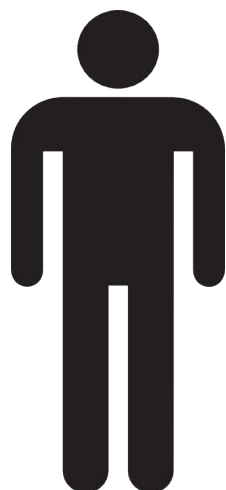
G/G



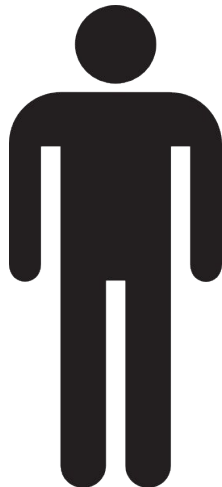
G/G



A/A



G/G



G/G

G/A

G/G

G/A

G/G

A/A

G/G

G/A

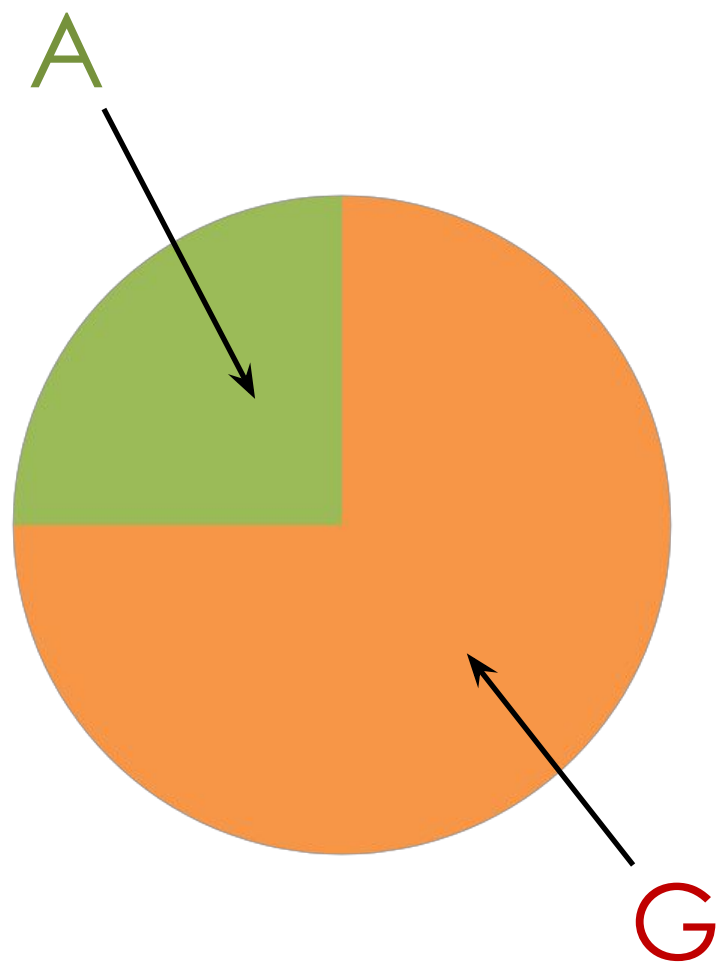
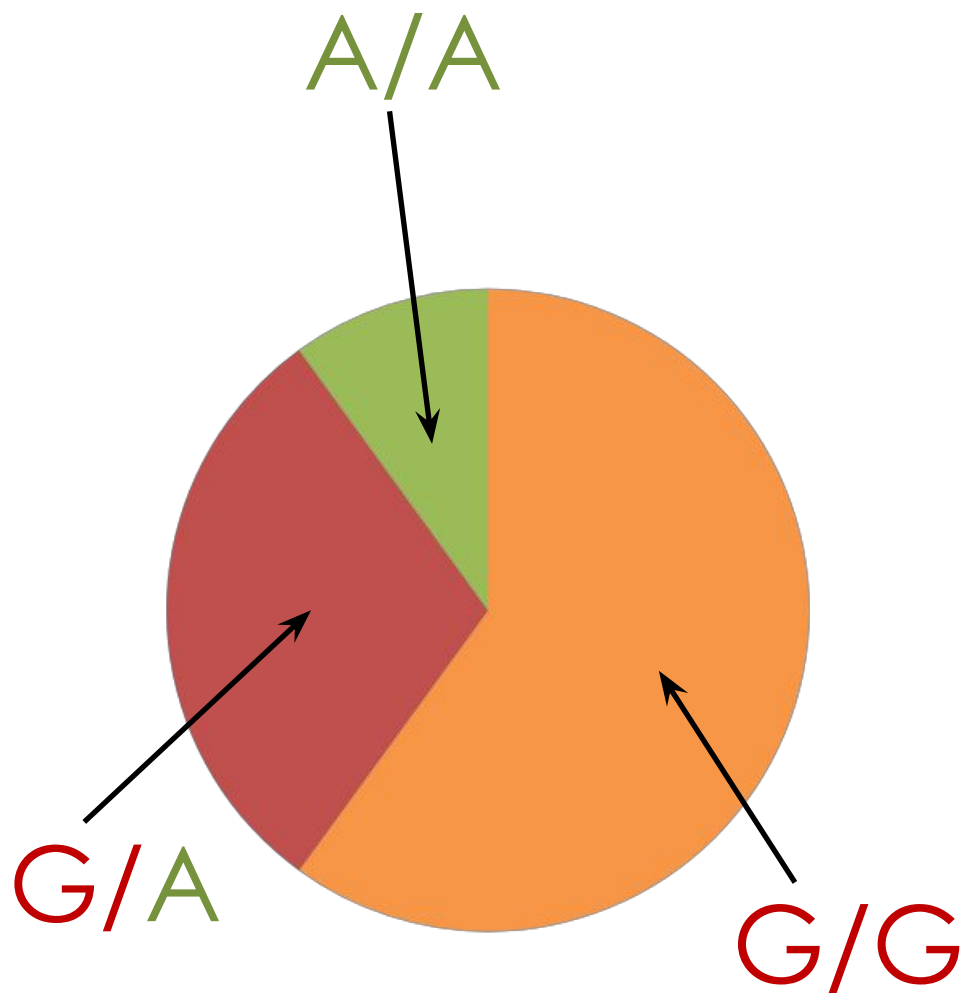
G/G

G/G

Генотип	Число образцов	Частота
G/G	6	$6/10 = 60\%$
G/A	3	$3/10 = 30\%$
A/A	1	$1/10 = 10\%$

Генотип	Число образцов	Частота
G/G	6	$6/10 = 60\%$
G/A	3	$3/10 = 30\%$
A/A	1	$1/10 = 10\%$

Аллель	Число аллелей	Частота
G	$6*2+3 = 15$	$15/20 = 75\%$
A	$1*2+3 = 5$	$5/20 = 25\%$



Вопрос 1:

Насколько наши оценки
близки к частотам в
генеральной совокупности?

SD, SE & CI для частоты

p – частота аллеля A_1 ,

n – размер выборки (аллелей)

$$SD = \sqrt{p(1 - p)}$$

$$SE = \sqrt{p(1 - p)/n}$$

$$CI = p \pm z^* SE \text{ (если } n \text{ большое)}$$

95%-CI частот (n небольшое)

• Модифицированный метод Вальда

$$p' = (n * p + z^2 / 0,5) / (n + z^2)$$

$$z = 1,96$$

$$w = z * \sqrt{p'(1 - p') / (n + z^2)}$$

$$CI = p' \pm w$$

Вопрос 2:

Как соотносятся между собой частоты аллелей и генотипов?

Сэр Харди, как
узнать частоты
генотипов в
следующем
поколении?

Эх, биологи...
Надо было учить
математику, сэр
Пённет!





Равновесие Харди-Вайнберга

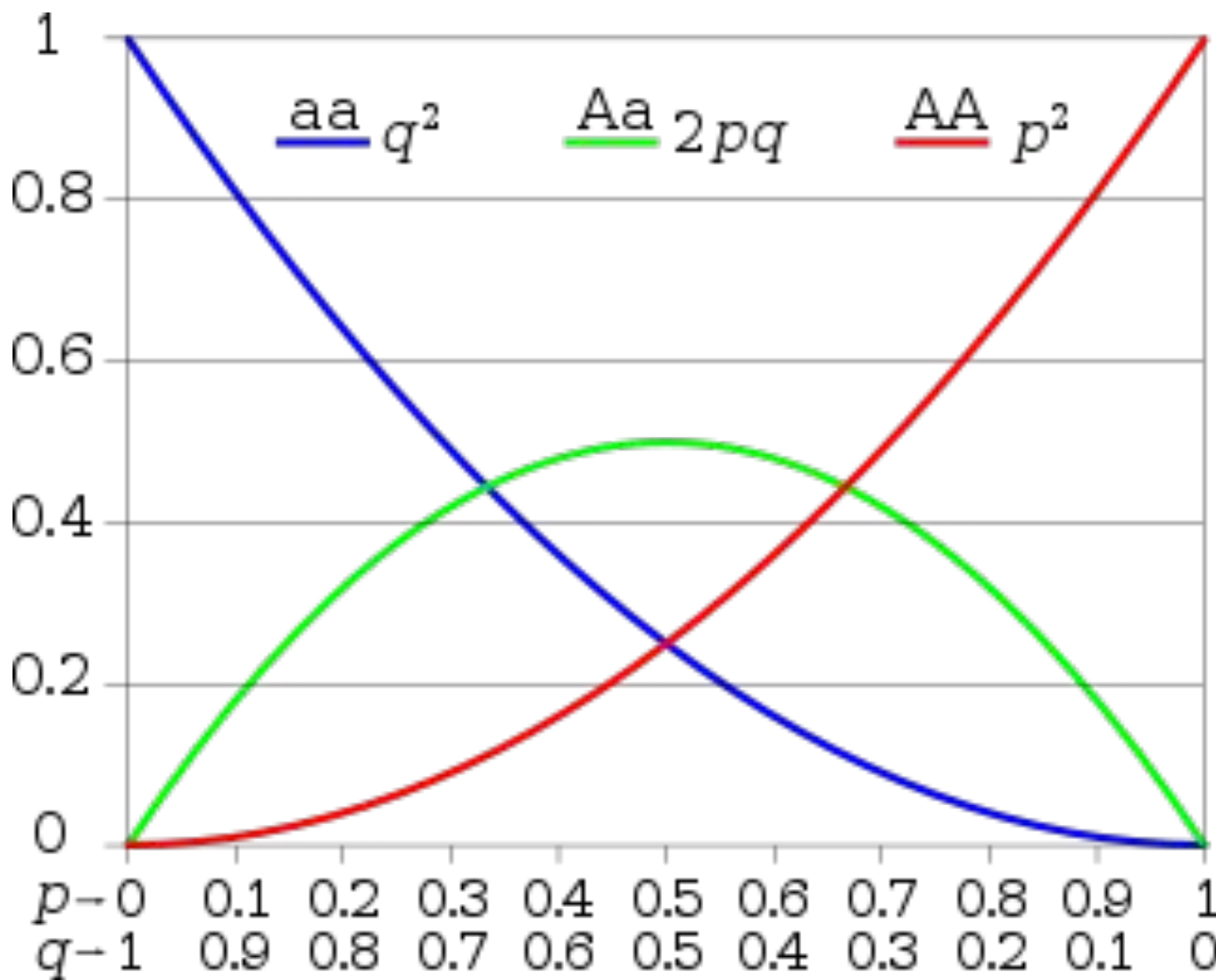
$$p = f_0(A_1A_1) + 1/2f_0(A_1A_2);$$

$$q = f_0(A_2A_2) + 1/2f_0(A_1A_2);$$

		Самцы	
		p	q
Самки	p	p^2	pq
	q	pq	q^2

$$f_1(A_1A_1) = p^2; f_1(A_2A_2) = q^2; f_1(A_1A_2) = 2pq$$

Равновесие Харди-Вайнберга



Задача

На остров попало 10 особей A_1A_1 и 40 A_1A_2 (поровну самцов и самок).

Какими будут частоты генотипов через одно и два поколения?

Задача

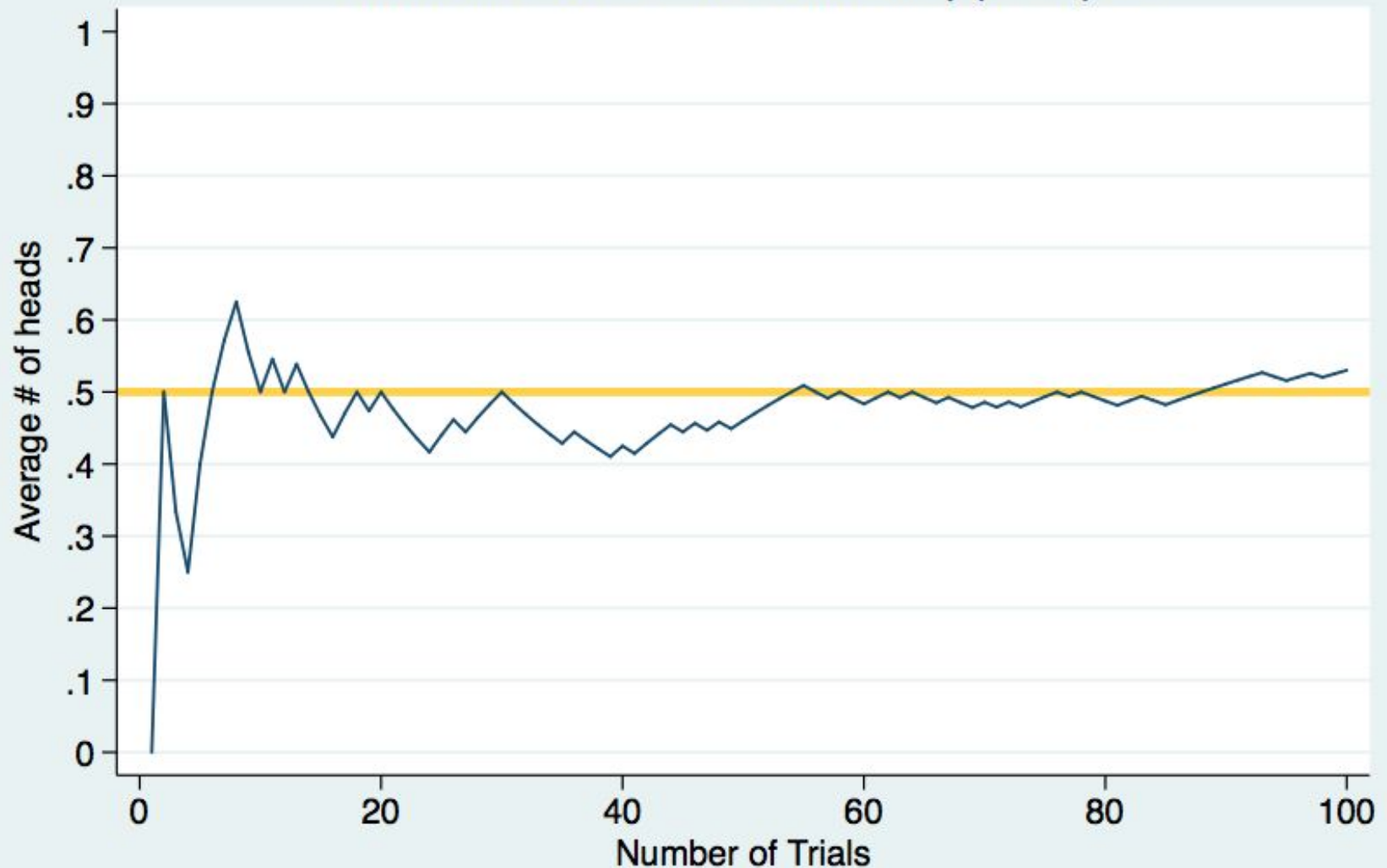
На остров попало 10 особей A_1A_1 и 40 A_1A_2 (поровну самцов и самок).

Какими будут частоты генотипов через одно и два поколения?

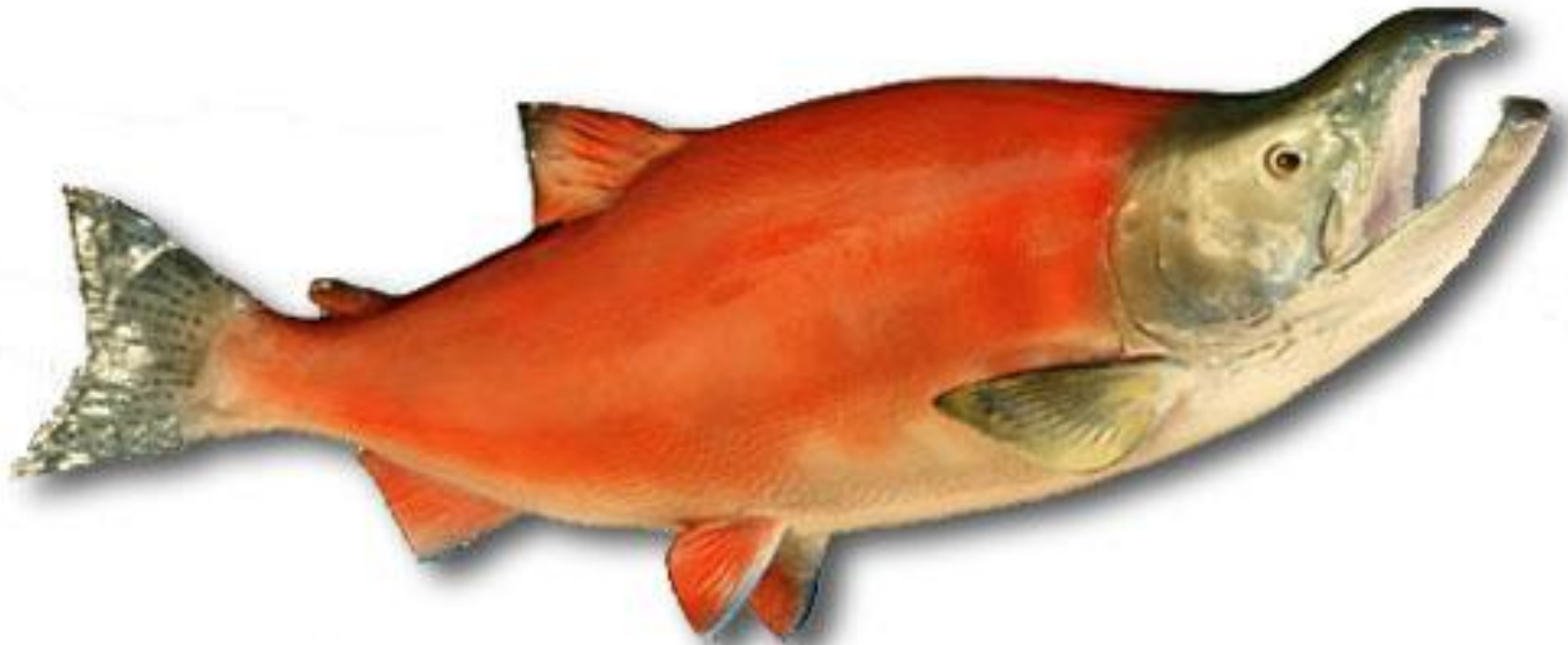
0,36 A_1A_1 0,48 A_1A_2 0,16 A_2A_2

Большая популяция

Average number of heads
from trial 1 to 100 of 1 coin with $p(\text{head})=.5$



Неперекрывающиеся поколения



Одинаковые частоты у обоих полов

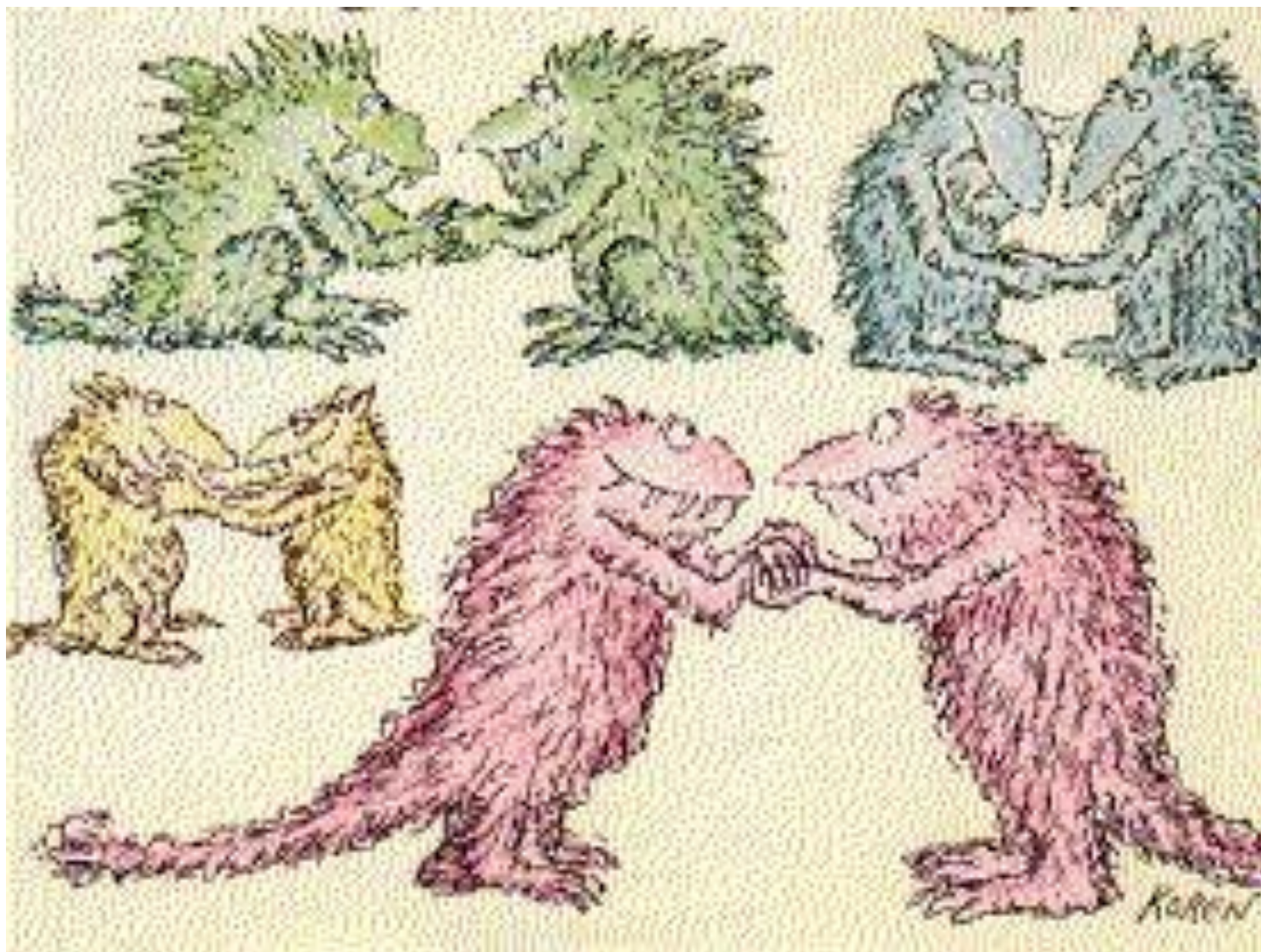


$$p + q$$

Частоты аллелей не меняются со временем

- a. мутации
- b. отбор
- c. поток генов
- d. дрейф генов

Панмиксия



Koren © 1977 The New Yorker Magazine, Inc.

Ассортативное скрещивание

- а. Инбридинг/аутбридинг
- б. Внутренняя структура популяции
- с. Ассортативность по определенному фенотипу

Множественный аллелизм

$$f(A_i A_i) = p_i^2$$

$$f(A_i A_j) = 2p_i p_j$$

Сцепление с полом

Самки: 20 $X^{A1}X^{A1}$ 10 $X^{A1}X^{A2}$
20 $X^{A2}X^{A2}$

Самцы: 50 $X^{A1}Y$

$$f(A1) = ?$$

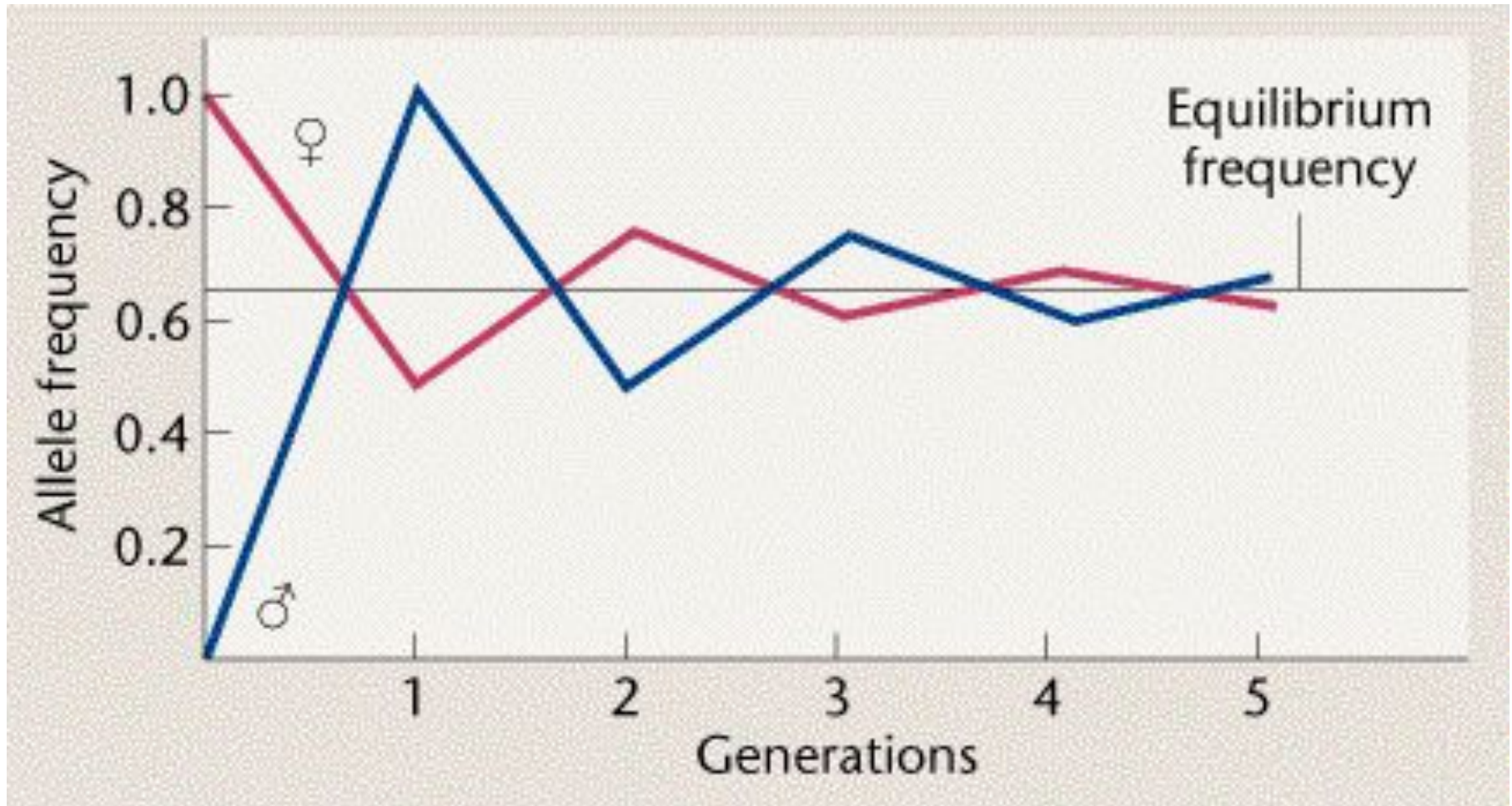
Сцепление с полом

Самки: 20 $X^{A1}X^{A1}$ 10 $X^{A1}X^{A2}$
20 $X^{A2}X^{A2}$

Самцы: 50 $X^{A1}Y$

$$f(A1) = \frac{(2 \cdot 20 + 1 \cdot 10 + 1 \cdot 50)}{(2 \cdot 50 + 1 \cdot 50)} = \frac{2}{3}$$

Сцепление с полом



Несколько локусов

		♂			
		AB	Ab	aB	ab
♀	AB				
	Ab				
	aB				
	ab				

$$f(A) = p, f(a) = q, f(B) = m, f(b) = n$$

Несколько локусов

		♂			
		AB	Ab	aB	ab
♀	AB	p^2m^2	p^2mn	pqm^2	$pqmn$
	Ab	p^2mn	p^2n^2	$pqmn$	pqn^2
	aB	pqm^2	$pqmn$	q^2m^2	q^2mn
	ab	$pqmn$	pqn^2	q^2mn	q^2n^2

$$f(A) = p, f(a) = q, f(B) = m, f(b) = n$$

Вопрос 3:

Что это все значит?

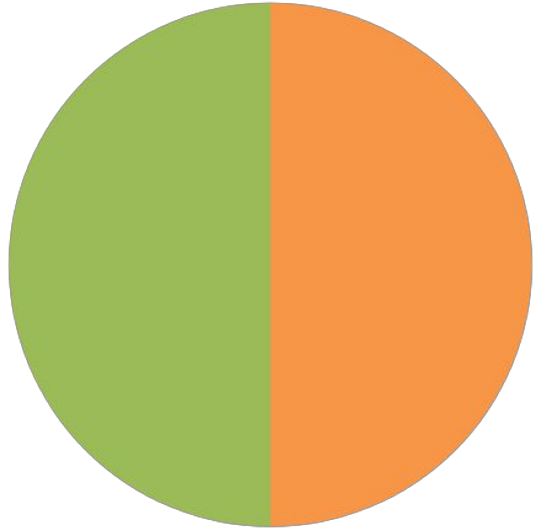
Вопрос 3:

~~Что это все значит?~~

Какие выводы можно
делать из частот аллелей и
генотипов?

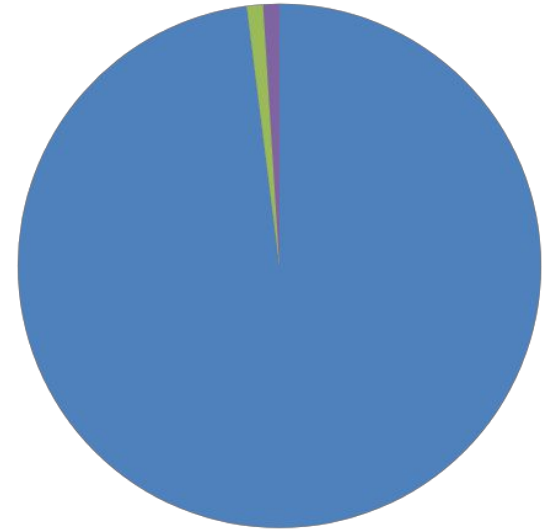
Вопрос 3.1:

Каково генетическое
разнообразие популяции?



$$f(A1) = 0,5$$

$$f(A2) = 0,5$$



$$f(A1) = 0,98$$

$$f(A2) = 0,01$$

$$f(A3) = 0,01$$

MAF

Частота более редкого аллеля
для биаллельных локусов

Генное разнообразие Nei

- $$H_e = 1 - \sum_{i=1}^n p_i^2$$

где n – число аллелей локуса,

p_i – частота i -го аллеля

ожидаемая гетерозитность

Поправка для ДИПЛОИДНЫХ ЛОКУСОВ

- $$\widehat{H}_e = \frac{2N}{2N - 1} \left(1 - \sum_{i=1}^n p_i^2 \right)$$

где N – число организмов в
выборке

n – число аллелей локуса,

p_i – частота i -го аллеля

Поправка для гаплоидных ЛОКУСОВ

- $$\widehat{H}_e = \frac{N}{N-1} \left(1 - \sum_{i=1}^n p_i^2 \right)$$

где N – число организмов в
выборке

n – число аллелей локуса,

p_i – частота i -го аллеля

Аллель 1	T	G	C	G	T	A	C	G	p ₁
Аллель 2	T	G	T	G	T	A	C	G	p ₂
Аллель 3	T	G	C	G	T	G	C	G	p ₃
Аллель 4	T	G	T	G	C	A	C	G	p ₄
Аллель 5	T	A	C	A	T	A	C	G	p ₅

Нуклеотидное разнообразие

-

$$\pi = \sum_{i,j}^n p_i p_j d_{ij}$$

где p_i и p_j – частоты соответствующих аллелей

d_{ij} – доля нуклеотидных позиций, по которым различаются аллели i и j

Вопрос 3.2:

Выполняется ли
равновесие ХВ?

AA	Aa	aa
30	60	110

χ^2 или точный тест Фишера

Квантили $\chi^2_{\alpha,n}$													
	0,01	0,025	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95
1	0,0002	0,0010	0,0039	0,0158	0,0642	0,1485	0,2750	0,4549	0,7083	1,0742	1,6424	2,7055	3,8415
2	0,0201	0,0506	0,1026	0,2107	0,4463	0,7133	1,0217	1,3863	1,8326	2,4079	3,2189	4,6052	5,9915
3	0,1148	0,2158	0,3518	0,5844	1,0052	1,4237	1,8692	2,3660	2,9462	3,6649	4,6416	6,2514	7,8147
4	0,2971	0,4844	0,7107	1,0636	1,6488	2,1947	2,7528	3,3567	4,0446	4,8784	5,9886	7,7794	9,4877
5	0,5543	0,8312	1,1455	1,6103	2,3425	2,9999	3,6555	4,3515	5,1319	6,0644	7,2893	9,2364	11,0705
6	0,8721	1,2373	1,6354	2,2041	3,0701	3,8276	4,5702	5,3481	6,2108	7,2311	8,5581	10,6446	12,5916
7	1,2390	1,6899	2,1673	2,8331	3,8223	4,6713	5,4932	6,3458	7,2832	8,3834	9,8032	12,0170	14,0671
8	1,6465	2,1797	2,7326	3,4895	4,5936	5,5274	6,4226	7,3441	8,3505	9,5245	11,0301	13,3616	15,5073
9	2,0879	2,7004	3,3251	4,1682	5,3801	6,3933	7,3570	8,3428	9,4136	10,6564	12,2421	14,6837	16,9190

Ген А, $p = 0,75$

Ген В, $p = 0,43$

Ген С, $p = 0,14$

Ген D, $p = 0,48$

Ген E, $p = 0,56$

Ген F, $p = 0,68$

Ген G, $p = 0,38$

Ген H, $p = 0,20$

...

Ген X, $p = 0,04$

Вопрос 3.3:

Случайны ли скрещивания
в популяции?

Инбридинг

- Коэффициент инбридинга

$$F = \frac{f(Aa)_E - f(Aa)_o}{f(Aa)_E}$$

где $f(Aa)_o$ – доля гетерозигот в выборке
 $f(Aa)_E$ – ожидаемая частота
гетерозигот $2pq$

Эффект Валунда

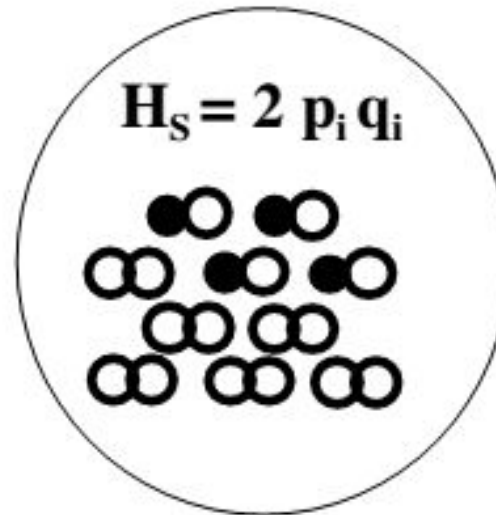
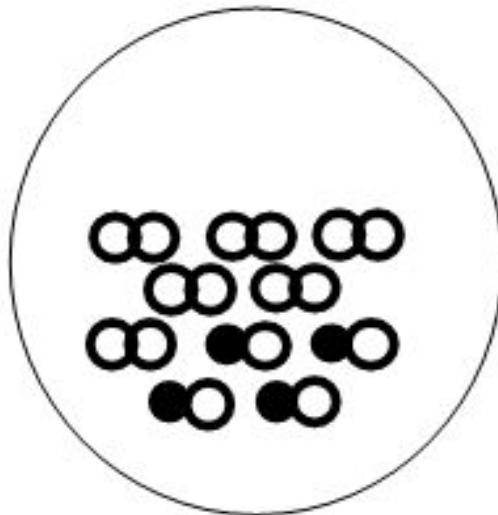
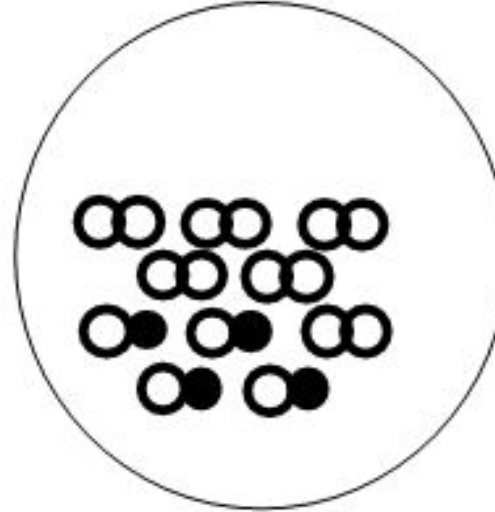
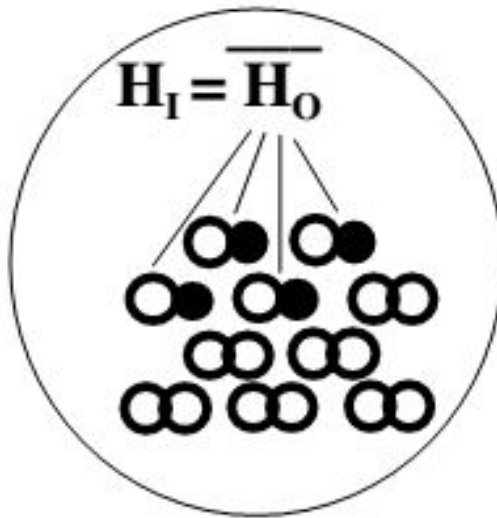
Поп1 $A_1A_1 - 0,49$; $A_1A_2 - 0,42$; $A_2A_2 - 0,09$

Поп2 $A_1A - 0,25$; $A_1A_2 - 0,5$; $A_2A_2 - 0,25$

Что если их объединить?

Тоже снижение гетерозиготности

F-статистика Райта



$$H_T = 2 \bar{p} \bar{q}$$

F-статистика Райта

$$\bullet F_{IT} = \frac{H_T - \overline{H_I}}{H_T}$$

общий индекс фиксации

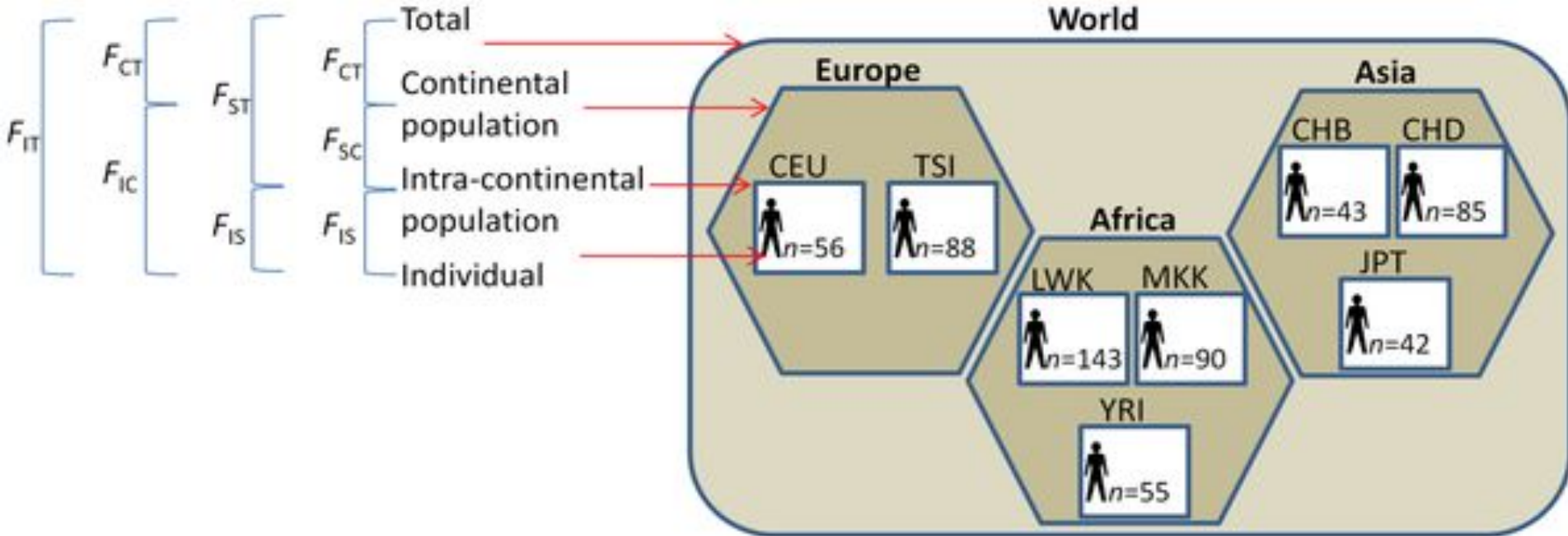
$$F_{IS} = \frac{H_S - H_I}{H_S}$$

коэффициент инбридинга

$$F_{ST} = \frac{H_T - \overline{H_S}}{H_T}$$

индекс фиксации
(между субпопуляциями)

$$(1 - F_{IT}) = (1 - F_{IS}) * (1 - F_{ST})$$



Chromosome type	SNPs #	Hierarchical F -statistics		
		Continental population	Intra-continental population	Individual
		0.12	0.13	0.13
Autosomal	1,100,484		0.01	0.01
				~0

Elhaik E (2012) Empirical Distributions of F_{ST} from Large-Scale Human Polymorphism Data. PLoS ONE 7(11): e49837.

doi:10.1371/journal.pone.0049837

<http://127.0.0.1:8081/plosone/article?id=info:doi/10.1371/journal.pone.0049837>

Другое определение F_{ST}

- $$F_{ST} = \frac{V_p}{p(1-p)}$$

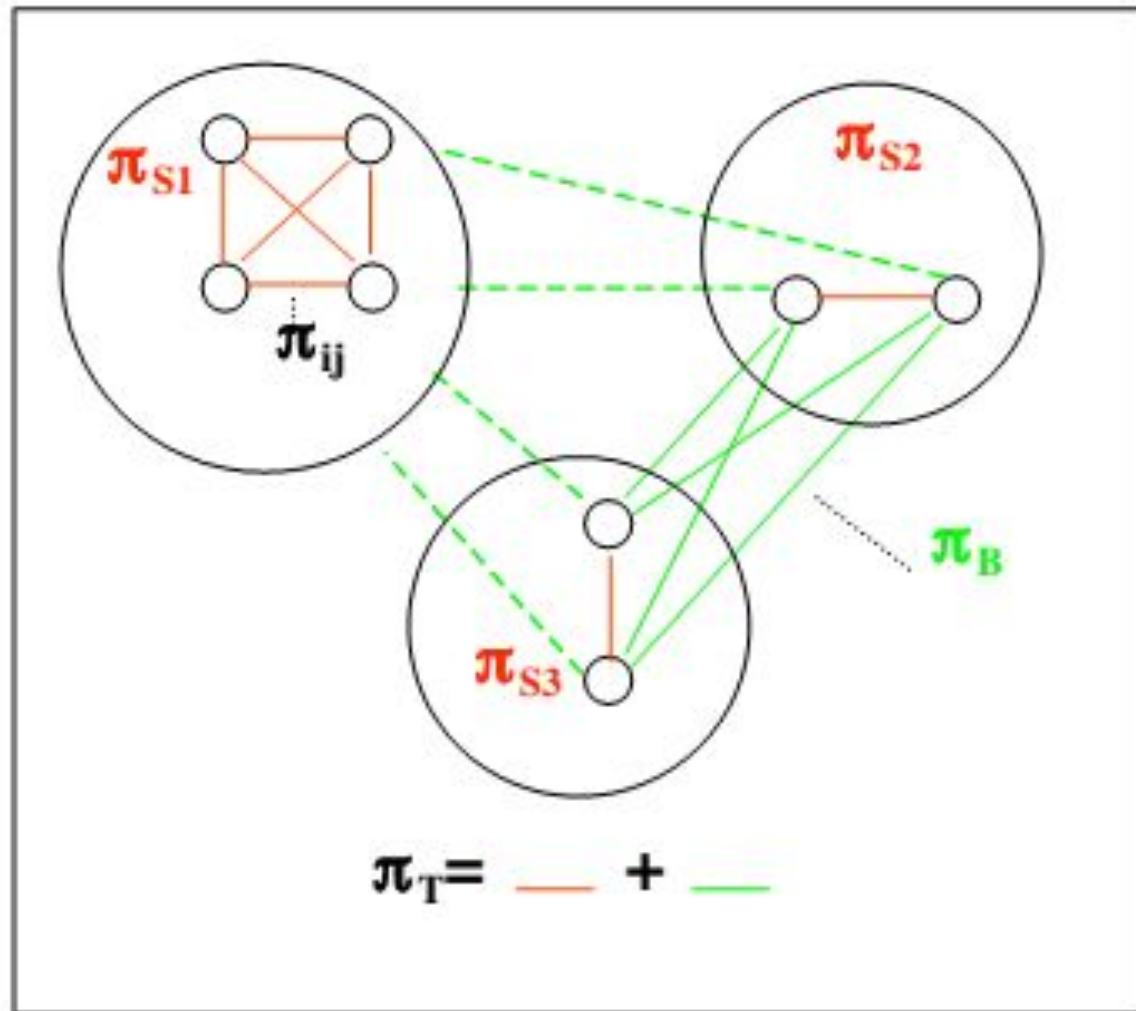
где p – частота аллеля во всей популяции,

V_p – дисперсия частоты аллеля между
субпопуляциями

Для двух аллелей

Αναλογ F_{ST}

$$\bullet \Phi_{ST} = \frac{\pi_T - \overline{\pi_S}}{\pi_T}$$



Вопрос 3.3:

Насколько различаются
популяции между собой?

Ген. дистанции между популяциями

$$F_{ST}$$

$$\Phi_{ST}$$

$$D = -\ln\left(\frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}\right)$$

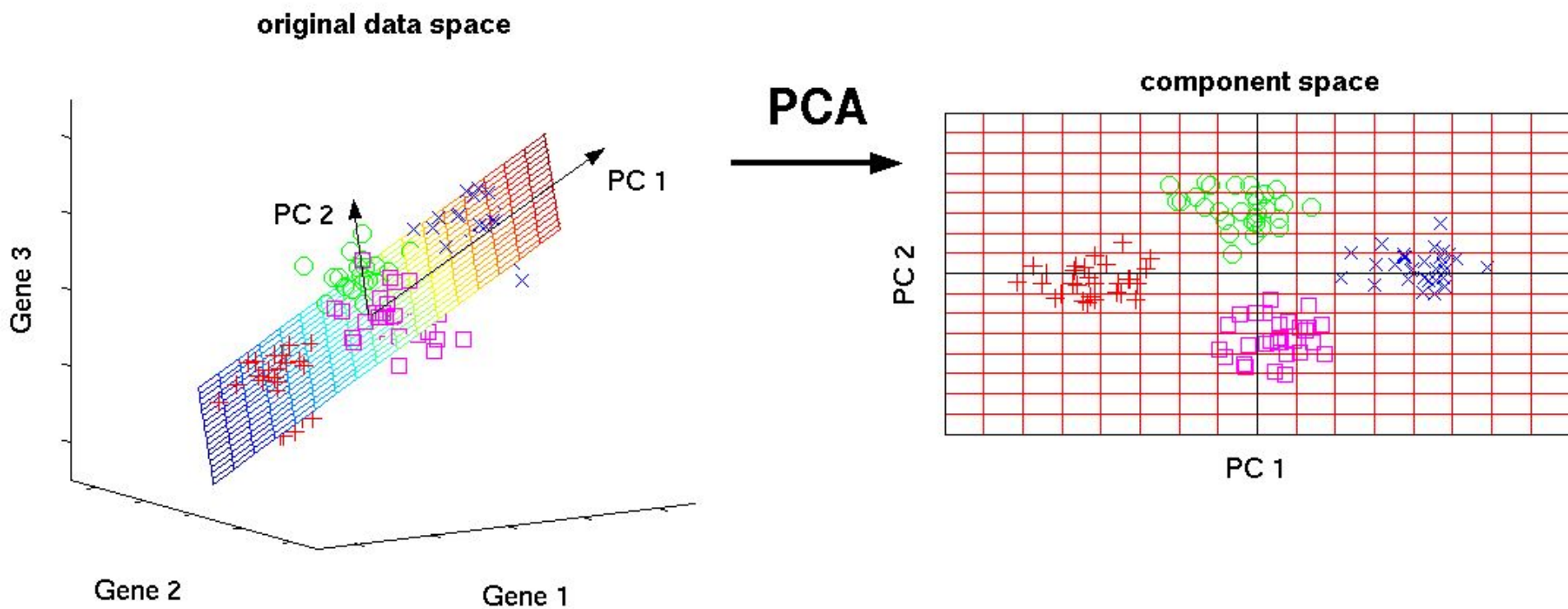
Стандартная генетическая дистанция
Nei

x_i – частота i -го аллеля в популяции x

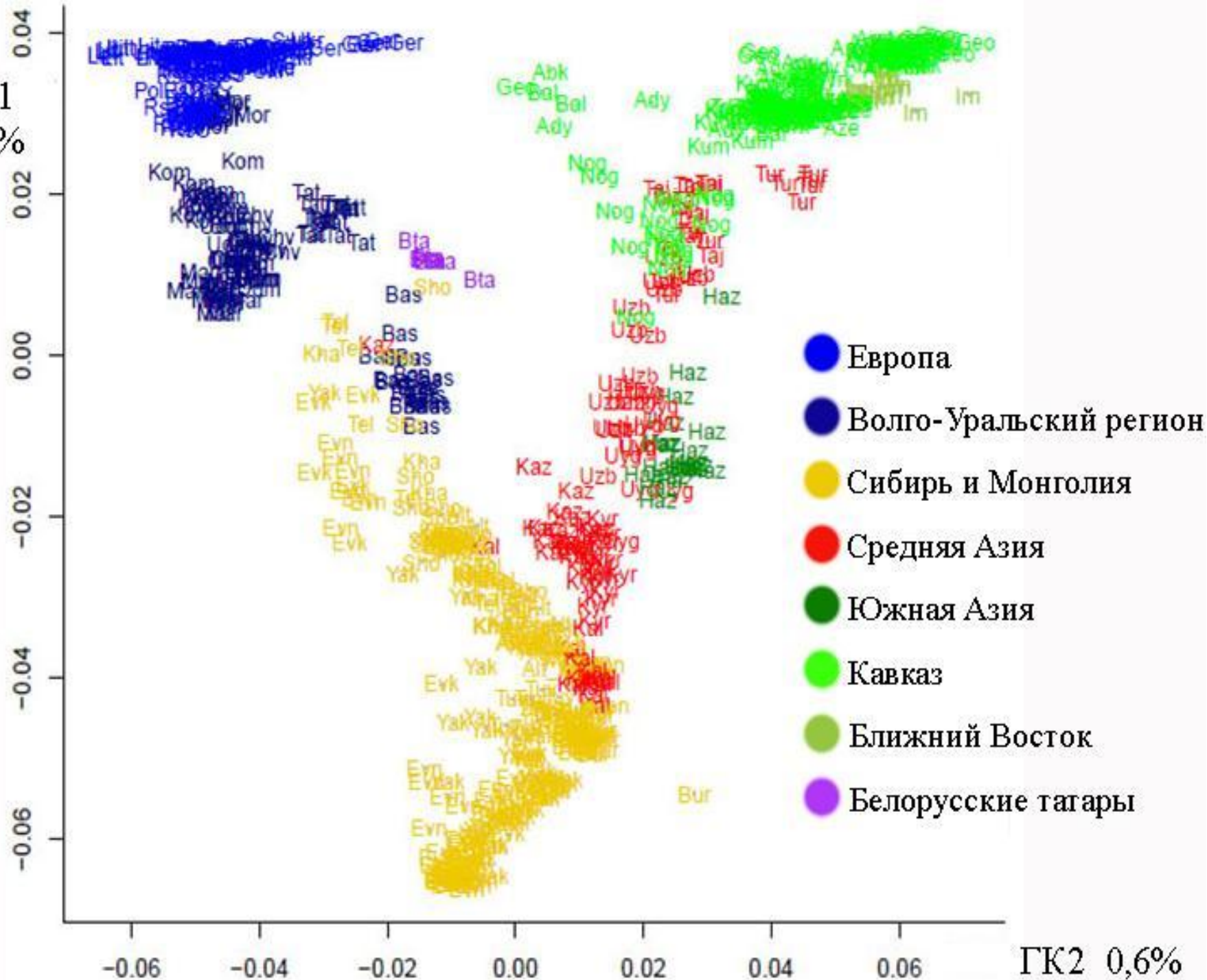
Анализ главных компонент

	A1	A2	A3	A4	A5	A6	A7
Поп1	x11	x12	x13	x14	x15	x16	x17
Поп2	x21	x22	x23	x24	x25	x26	x27
Поп3	x31	x32	x33	x34	x35	x36	x37
Поп4	x41	x42	x43	x44	x45	x46	x47
Поп5	x51	x52	x53	x54	x55	x56	x57
Поп6	x61	x62	x63	x64	x65	x66	x67
Поп7	x71	x72	x73	x74	x75	x76	x77
Поп8	x81	x82	x83	x84	x85	x86	x87
Поп9	x91	x92	x93	x94	x95	x96	x97
Поп10	x101	x102	x103	x104	x105	x106	x107

Анализ главных компонент



ГК1
3,7%



ГК2 0,6%

Кластерный анализ



How many clusters?

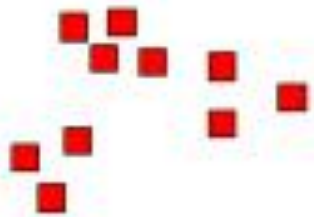
Кластерный анализ



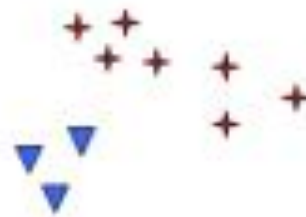
How many clusters?



Six Clusters



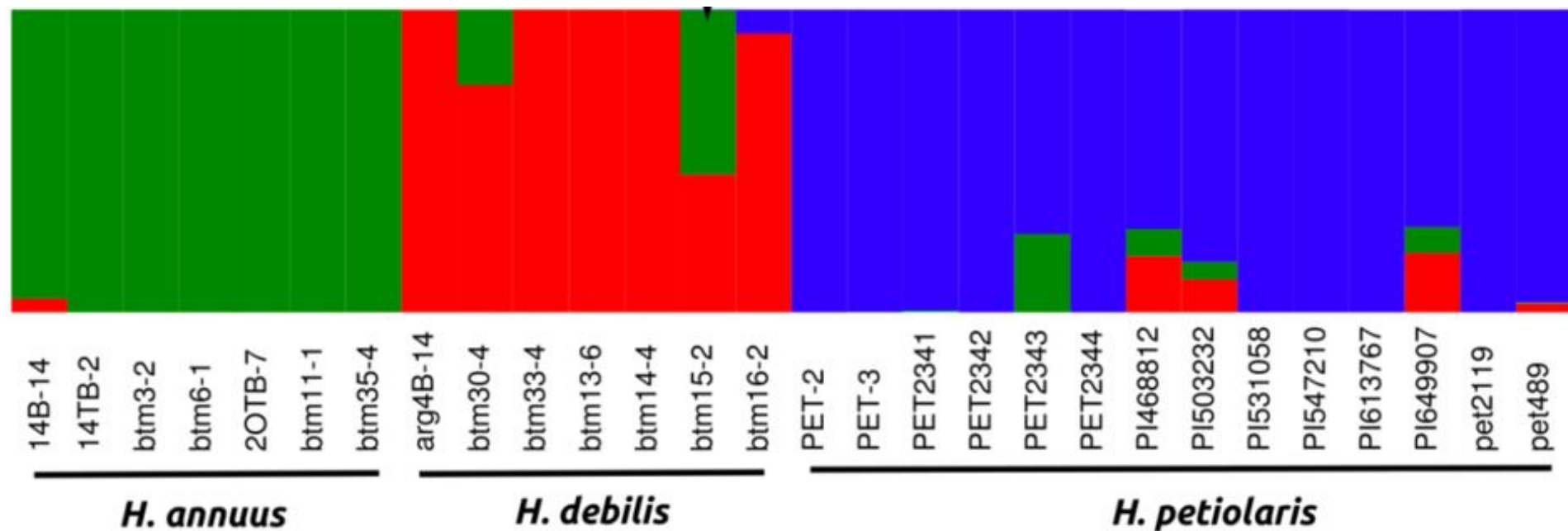
Two Clusters



Four Clusters



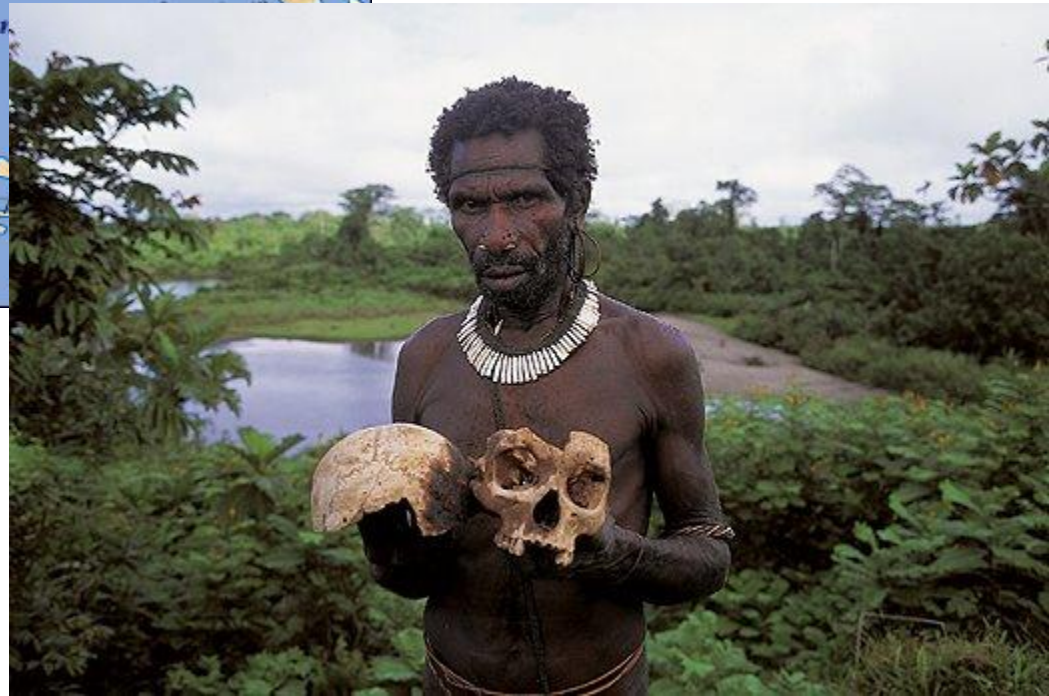
Кластерный анализ



Вопрос 3.4:

Действует ли
естественный отбор?

Пример с куру



Группа	Генотип по кодону 129 гена белка PRNP			р- значе- ние
	Met/Met	Met/Val	Val/Val	
Жен. старше 1950 г.р.	16	86	23	$2,1 \times 10^{-5}$
Муж. старше 1960 г.р.	34	111	60	0,15
Молодые М и Ж	52	136	94	0,8