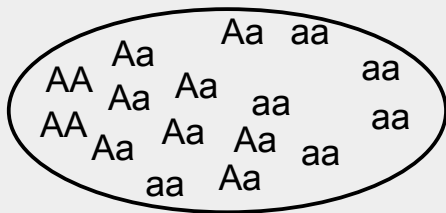


# Терминология генетики популяций

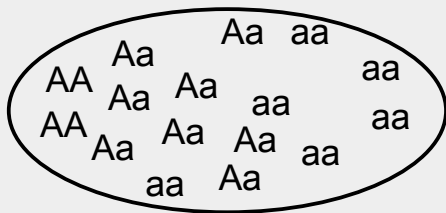
# Терминология генетики популяций

Популяция — группа организмов одного вида, частично или полностью (репродуктивно) изолированная от других таких групп.



# Терминология генетики популяций

Популяция — группа организмов одного вида, частично или полностью (репродуктивно) изолированная от других таких групп.

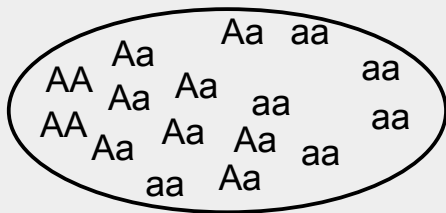


Частоты генотипов (genotype frequency):  $P_{AA}, P_{Aa}, P_{aa}$

Частоты аллелей (allele / gene frequency):  $p_A, p_a$  ( $q_a$ ).

# Терминология генетики популяций

Популяция — группа организмов одного вида, частично или полностью (репродуктивно) изолированная от других таких групп.



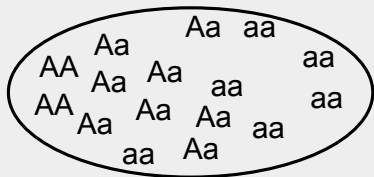
Частоты генотипов (genotype frequency):  $P_{AA}, P_{Aa}, P_{aa}$

Частоты аллелей (allele / gene frequency):  $p_A, p_a$  ( $q_a$ ).

$$p_A = P_{AA} + \frac{1}{2}P_{Aa}$$

$$p_a = \frac{1}{2}P_{Aa} + P_{aa}$$

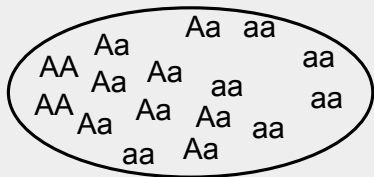
## Закон Харди-Вайнберга (Hardy-Weinberg law)



$$p_A = P_{AA} + \frac{1}{2}P_{Aa}$$

$$p_a = \frac{1}{2}P_{Aa} + P_{aa}$$

## Закон Харди-Вайнберга (Hardy-Weinberg law)



$$p_A = P_{AA} + \frac{1}{2}P_{Aa}$$

$$p_a = \frac{1}{2}P_{Aa} + P_{aa}$$

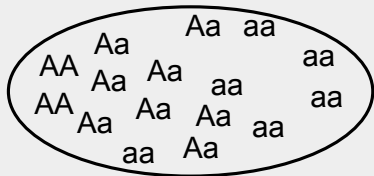
Частоты генотипов в следующем поколении:

$$P_{AA} = p_A \cdot p_A = p_A^2$$

$$P_{Aa} = p_A \cdot p_a + p_a \cdot p_A = 2p_A p_a$$

$$P_{aa} = p_a \cdot p_a = p_a^2$$

# Закон Харди-Вайнберга (Hardy-Weinberg law)



$$p_A = P_{AA} + \frac{1}{2}P_{Aa}$$

$$p_a = \frac{1}{2}P_{Aa} + P_{aa}$$

Частоты генотипов в следующем поколении:

$$P_{AA} = p_A \cdot p_A = p_A^2$$

$$P_{Aa} = p_A \cdot p_a + p_a \cdot p_A = 2p_A p_a$$

$$P_{aa} = p_a \cdot p_a = p_a^2$$

Частоты гамет:

$$p_A = P_{AA} + \frac{1}{2}P_{Aa} = p_A^2 + \frac{1}{2} \cdot 2p_A p_a = p_A(p_A + p_a) = p_A$$

$$p_a = \frac{1}{2}P_{Aa} + P_{aa} = \dots = p_a$$

# Закон Харди-Вайнберга

Относительные частоты аллелей в популяции постоянны, а частоты генотипов устанавливаются во втором поколении и позже сохраняются.



# Закон Харди-Вайнберга

Относительные частоты аллелей в популяции постоянны, а частоты генотипов устанавливаются во втором поколении и позже сохраняются.

## Следствия

- ▶ Зная частоты аллелей, мы можем рассчитать частоты генотипов.
- ▶ Частые аллели хранятся в гомозиготах, редкие — в гетерозиготах.
- ▶ Популяция не вырождается.

# Закон Харди-Вайнберга

Относительные частоты аллелей в популяции постоянны, а частоты генотипов устанавливаются во втором поколении и позже сохраняются.

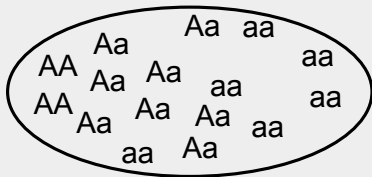
## Условия выполнения

1. Бесконечный размер популяции.
2. Случайность скрещивания (панмиксия).
3. Одинаковая частота генотипов у особей обоих полов.
4. Отсутствие мутаций.
5. Отсутствие иммиграции.
6. Отсутствие эмиграции или равномерная эмиграция особей разных генотипов.
7. Одинаковая фертильность особей всех генотипов.
8. Одинаковая выживаемость особей всех генотипов.

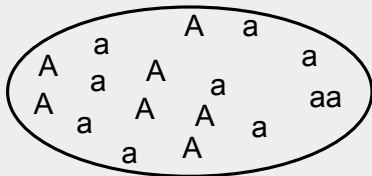
По Felsenstein, J. Theoretical evolutionary genetics. 2016.

# Варианты популяций:

1. диплоидная с короткой гаплоидной фазой;



2. гаплоидная без полового процесса;



3. гаплоидная с короткой диплоидной фазой.

# Силы, действующие на популяцию:

1. естественный отбор;
2. мутации;
3. рекомбинация;
4. дрейф генов;
5. миграция.

# Естественный отбор (natural selection) —

# Естественный отбор (natural selection) —

это дифференциальное воспроизведение генотипов.

# Естественный отбор (natural selection) —

это дифференциальное воспроизведение генотипов.

Гаплоидная популяция без полового процесса и пересечения поколений



$$N_A(t), N_a(t)$$

Абсолютная (дарвиновская) приспособленность

(fitness):  $W_A = v_A \cdot f_A$ ,  $v$  — выживание, а  $f$  — фертильность

# Естественный отбор (natural selection) —

это дифференциальное воспроизведение генотипов.

Гаплоидная популяция без полового процесса и пересечения поколений



$N_A(t), N_a(t)$

Абсолютная (дарвиновская) приспособленность (fitness):  $W_A = v_A \cdot f_A$ ,  $v$  — выживание, а  $f$  — фертильность

В следующем поколении:

$$N_A(t+1) = W_A \cdot N_A$$

$$\frac{p_A(t+1)}{p_a(t+1)} = \frac{W_A}{W_a} \cdot \frac{p_A(t)}{p_a(t)} = w_A \cdot \frac{p_A(t)}{p_a(t)} = (1+s) \frac{p_A(t)}{p_a(t)},$$

где  $w_A$  — относительная приспособленность,  
а  $s$  — коэффициент отбора в пользу  $A$ .



## Отбор в пользу и против генотипа

$$\frac{p_A}{p_a} = \frac{W_A}{W_a} \cdot \frac{p_A(t)}{p_a(t)} = w_A \cdot \frac{p_A(t)}{p_a(t)} = (1 + s) \frac{p_A(t)}{p_a(t)},$$

где  $w_A$  — относительная приспособленность,  
а  $s$  — коэффициент отбора в пользу  $A$ .

$$\frac{p_a}{p_A} = \frac{W_a}{W_A} \cdot \frac{p_a(t)}{p_A(t)} = w_a \cdot \frac{p_a(t)}{p_A(t)} = (1 - s) \frac{p_a(t)}{p_A(t)},$$

и это уже другая  $s$ !

# Мутации

$10^{-7}$

Пусть  $p$  — частота доминантной аллели  $A$ , причём  $A \rightarrow a$  с частотой  $m$  и  $a \rightarrow A$  с частотой  $M$ .

$$p^{t+1} = p^t(1 - m) + (1 - p^t)M$$

# Мутации

$$10^{-7}$$

Пусть  $p$  — частота доминантной аллели  $A$ , причём  $A \rightarrow a$  с частотой  $m$  и  $a \rightarrow A$  с частотой  $M$ .

$$p^{t+1} = p^t(1 - m) + (1 - p^t)M$$

$$\Delta p = p^t \cdot (1 - m) + (1 - p^t) \cdot M - p^t = -m \cdot p^t + (1 - p^t) \cdot M$$

# Мутации

$$10^{-7}$$

Пусть  $p$  — частота доминантной аллели  $A$ , причём  $A \rightarrow a$  с частотой  $m$  и  $a \rightarrow A$  с частотой  $M$ .

$$p^{t+1} = p^t(1 - m) + (1 - p^t)M$$

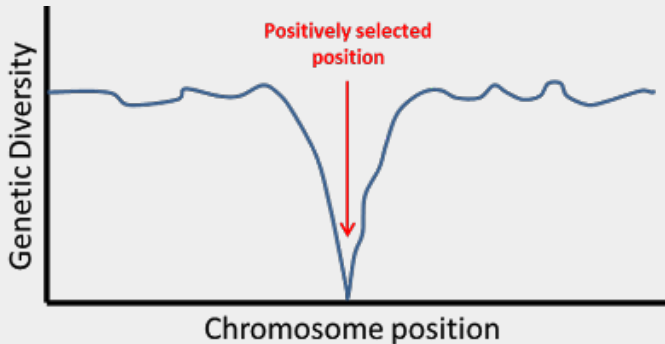
$$\Delta p = p^t \cdot (1 - m) + (1 - p^t) \cdot M - p^t = -m \cdot p^t + (1 - p^t) \cdot M$$

Состояние равновесия (equilibrium):  $\Delta p = 0$

$$p_e = \frac{M}{m + M}$$

# Рекомбинация: selective sweep

Снижение разнообразия около полезной мутации.



<http://genomeeee.blogspot.ru/2011/12/hard-selective-sweeps-dont-seem-to-be.html>

## Неравновесие по сцеплению (linkage disequilibrium, LD)

Разница между наблюдаемой и ожидаемой частотой гамет (или гаплотипов) для пары генов

$$D = P_{AB} - p_A * p_B$$

$D = 0$  — равновесие по сцеплению

Поправки:

$$\blacktriangleright D' = \frac{D}{D_{min}}, \begin{cases} D_{min} = \min(p_A p_B, q_a q_b), D \geq 0 \\ D_{min} = (p_A q_b, q_a p_B), D \leq 0 \end{cases}$$

$$\blacktriangleright r^2 = \frac{D^2}{p_A q_a p_B q_b}$$

# Поиск следов отбора в последовательностях

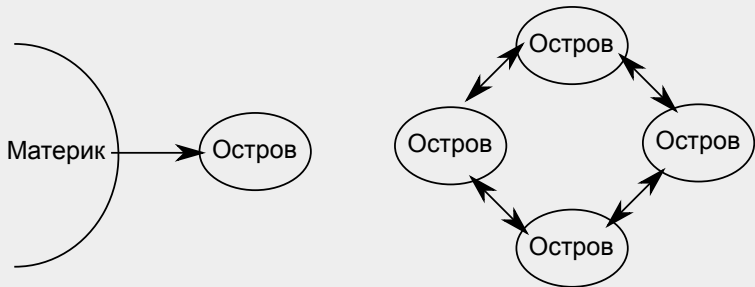
- ▶ Отношение частот несинонимичных и синонимичных замен ( $\frac{d_N}{d_S} = \frac{K_a}{K_s}$ );
- ▶ McDonald-Kreitman test (сравнение значений  $\frac{d_N}{d_S}$  внутри вида и в группе видов).
- ▶ Tajima's D (разница среднего числа попарных различий между особями и общего числа различий);
- ▶ Fay & Wu's H (разница числа попарных различий и числа гомозигот для вновь возникшей нейтральной мутации);
- ▶ EHH, extended haplotype homozygosity (определение длинных участков с высоким LD).

См. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annual review of genetics. 2013 Nov 23;47:97-120.

# Миграция

Поток генов

Две основные группы моделей:



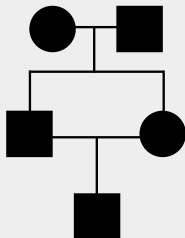
Может приводить к LD!



# Инбридинг (inbreeding) —

это преимущественное скрещивание с родственниками.

S. Wright



Коэффициент инбридинга ( $f$ ) —  
вероятность того, что две копии гена  
идентичны по происхождению

$$P_{AA} = p^2(1 - f) + pf$$

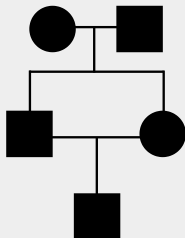
$$P_{Aa} = 2p(1 - p)(1 - f)$$

$$P_{aa} = (1 - p)^2(1 - f) + (1 - p)f$$

# Инбридинг (inbreeding) —

это преимущественное скрещивание с родственниками.

S. Wright



Коэффициент инбридинга ( $f$ ) —  
вероятность того, что две копии гена  
идентичны по происхождению

$$P_{AA} = p^2(1 - f) + pf$$

$$P_{Aa} = 2p(1 - p)(1 - f)$$

$$P_{aa} = (1 - p)^2(1 - f) + (1 - p)f$$

$f$  = коэффициент родства родителей

Следствие инбридинга — снижение гетерозиготности.

Изменение средней приспособленности:

$$w = w_0 - 2fp(1 - p)[w_{Aa} - (w_{AA} + w_{aa})]$$

## $F_{ST}$ и другие коэффициенты

$$P_{AA} = \pi^2(1 - F) + \pi F$$

$$P_{Aa} = 2\pi(1 - \pi)(1 - F)$$

$$P_{aa} = (1 - \pi)^2(1 - F) + (1 - \pi)F$$

$$\pi = cp_1 + (1 - c)p_2$$



$$(1 - F) = (1 - f)(1 - \theta), \text{ где } f = cf_1 + (1 - c)f_2$$

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

$$\theta = F_{ST} = \frac{\sigma_{\pi}^2}{\pi(1 - \pi)}$$

## $F_{ST}$ и другие коэффициенты

$$\theta = F_{ST} = \frac{\sigma_{\pi}^2}{\pi(1 - \pi)}$$

Часть разнообразия, которая объясняется различными частотами аллелей в субпопуляциях.

## $F_{ST}$ и другие коэффициенты

$$\theta = F_{ST} = \frac{\sigma_{\pi}^2}{\pi(1 - \pi)}$$

Часть разнообразия, которая объясняется различными частотами аллелей в субпопуляциях.

Способы оценки:

- ▶ метод моментов;
- ▶ ML;
- ▶ BI.

## $F_{ST}$ и другие коэффициенты

$$\theta = F_{ST} = \frac{\sigma_{\pi}^2}{\pi(1 - \pi)}$$

Часть разнообразия, которая объясняется различными частотами аллелей в субпопуляциях.

Способы оценки:

- ▶ метод моментов;
- ▶ ML;
- ▶ BI.

Статистики, основанные на  $F_{ST}$ :

- ▶  $\Phi_{ST}$  (учитывает гаплотипы);
- ▶  $R_{ST}$  (для микросателлитов);
- ▶  $Q_{ST}$  (для количественных данных).

## $F_{ST}$ и другие коэффициенты

$$\theta = F_{ST} = \frac{\sigma_{\pi}^2}{\pi(1 - \pi)}$$

Часть разнообразия, которая объясняется различными частотами аллелей в субпопуляциях.

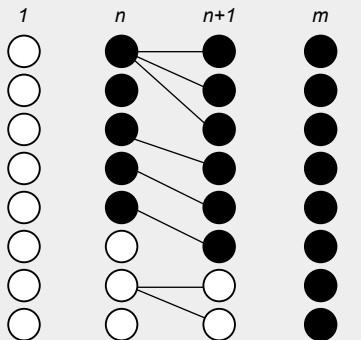
Применение  $F_{ST}$  и родственных статистик:

- ▶ оценка скорости миграции;
- ▶ демографическая история популяции;
- ▶ поиск локусов, подверженных отбору;
- ▶ криминалистика.

# Дрейф генов (genetic drift) —

это изменение частот аллелей в популяции за счёт случайных событий.

Поколения:



фиксация (fixation)  
«чёрной» аллели



# Эффективная численность популяции

$N_e$  — это число особей, передающих свои гены в следующее поколение.

Почти всегда  $N_e \ll N_c$  (номинальная численность)

Причины:

- ▶ разное число особей двух полов;
- ▶ селективные скрещивания;
- ▶ разная фертильность разных генотипов;
- ▶ другие отступления от «идеальности» популяции.

# Эффективная численность популяции

$N_e$  — это число особей, передающих свои гены в следующее поколение.

Почти всегда  $N_e \ll N_c$  (номинальная численность)

Причины:

- ▶ разное число особей двух полов;
- ▶ селективные скрещивания;
- ▶ разная фертильность разных генотипов;
- ▶ другие отступления от «идеальности» популяции.

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

$\Theta = 4\mu N_e \Leftrightarrow N_e = \frac{\Theta}{4\mu}$ , где  $\Theta$  — нуклеотидное разнообразие, а  $\mu$  — скорость мутирования.

## Силы, действующие на популяцию:

1. естественный отбор;
2. мутации;
3. рекомбинация;
4. дрейф генов;
5. миграция.

Соотношение зависит от собственных коэффициентов и эффективной численности популяции.

В маленькой популяции дрейф важнее отбора.

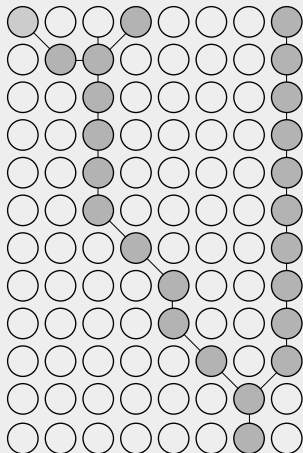
# Программы для анализа структуры популяции и её динамики

- ▶ PAML
- ▶ DNASP
- ▶ Arlequin
- ▶ Structure и fastStructure
- ▶ Admixture
- ▶ Geneland
- ▶ poppr
- ▶ pegas
- ▶ strataG
- ▶ *etc.*

# Коалесценция —

схождение к общему предку (время до события коалесценции обычно считают в поколениях)

J. Kingman



# Коалесценция

J. Kingman

Популяция Райта-Фишера: гаплоидная, размер постоянный, поколения разделены.

$$P(t_k) = \left(\frac{1}{N}\right)\left(1 - \frac{1}{N}\right)^{k-1}$$

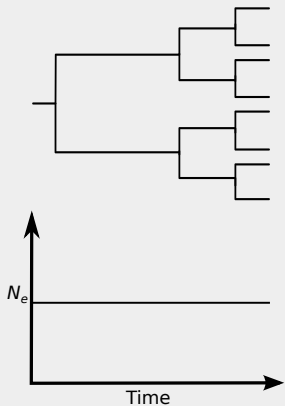
$$P(t_1) = \frac{n(n-1)}{2N}$$

$$P(t_k) = \left(\frac{n(n-1)}{2N}\right)\left(1 - \frac{n(n-1)}{2N}\right)^{k-1}$$

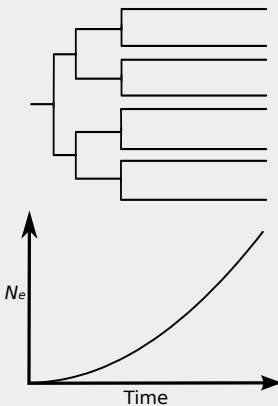
Ожидаемое время схождения всех линий —  $2N$   
( $n \ll N$  и  $N \rightarrow \infty$ ).

# Связь динамики $N_e$ и частоты событий коалесценции

Constant Population Size

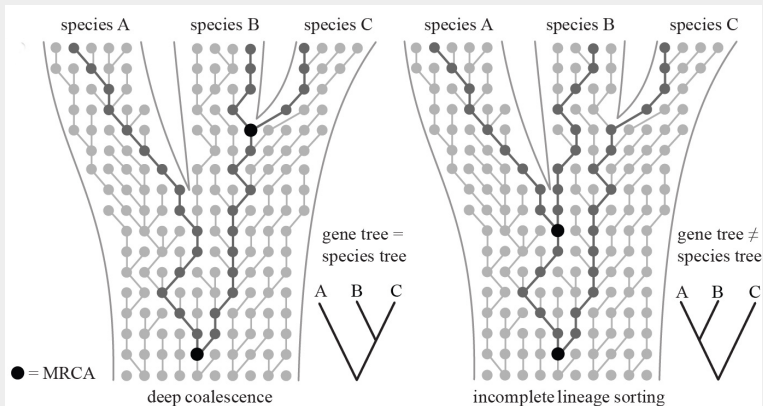


Exponential Growth



По <https://commons.wikimedia.org/w/index.php?curid=25233219> (By Erik Volz - Given to me by the author, CC BY 2.5)

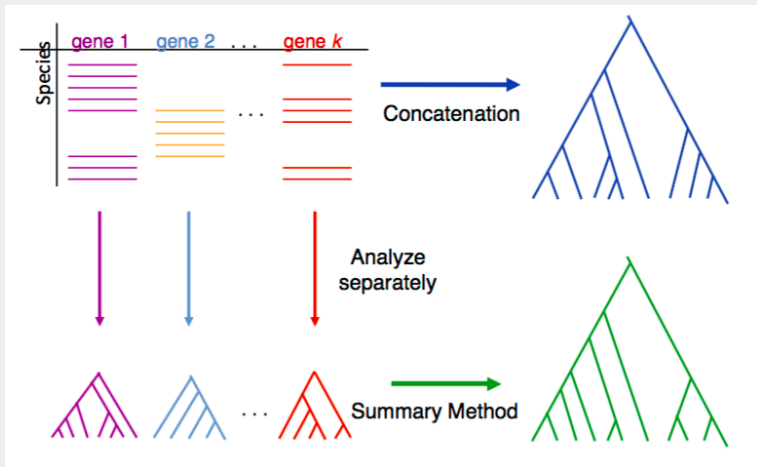
# Неполное разделение линий (incomplete lineage sorting)



Leliaert F *et al.* (2014) DNA-based species delimitation in algae. *European Journal of Phycology* 49: 179-196.

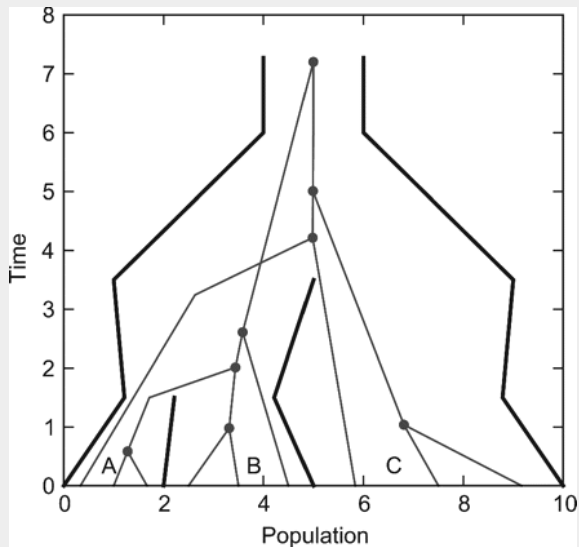


# Конфликтующие деревья генов



<https://t.co/75waIXWBKq>

# Multispecies coalescent



Heled J, Drummond AJ. Bayesian inference of species trees from multilocus data. *Molecular biology and evolution*. 2010 Mar 1;27(3):570-80.

# Программы для анализа структуры популяции и её динамики-2

- ▶ **Lamarc**
- ▶ \*Beast (часть **BEAST**)
- ▶ **BATWING**
- ▶ **BP&P** (часть PAML)
- ▶ *etc.*