

# Экология и эволюция

Игорь Мартынович Рузин

Доцент, в.н.с.

[ivan4995@gmail.com](mailto:ivan4995@gmail.com)

[igor.rouzine@iephb.ru](mailto:igor.rouzine@iephb.ru)



# Plan

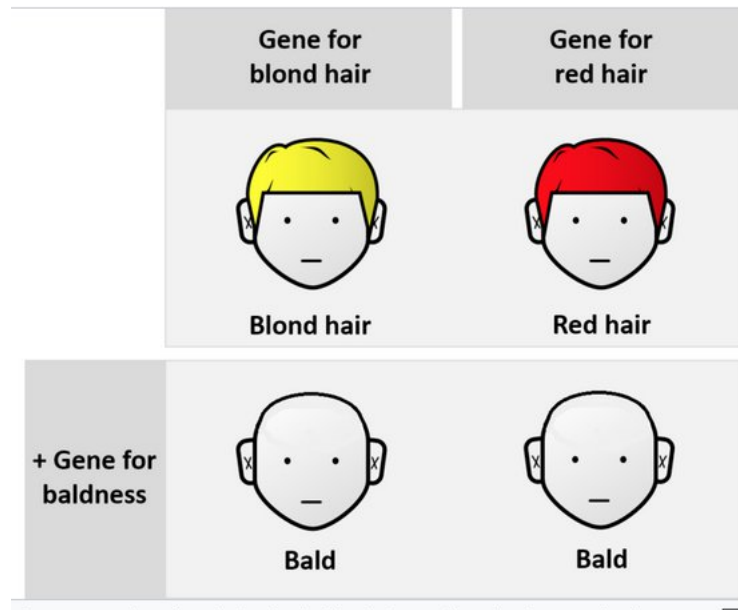
- і Что такое эпистаз?
- і Универсальная мера эпистаза
- і Проблемы с обнаружением
- і Метод 3 локусного обнаружения
- і Применение к вирусу гриппа

Эпистаз

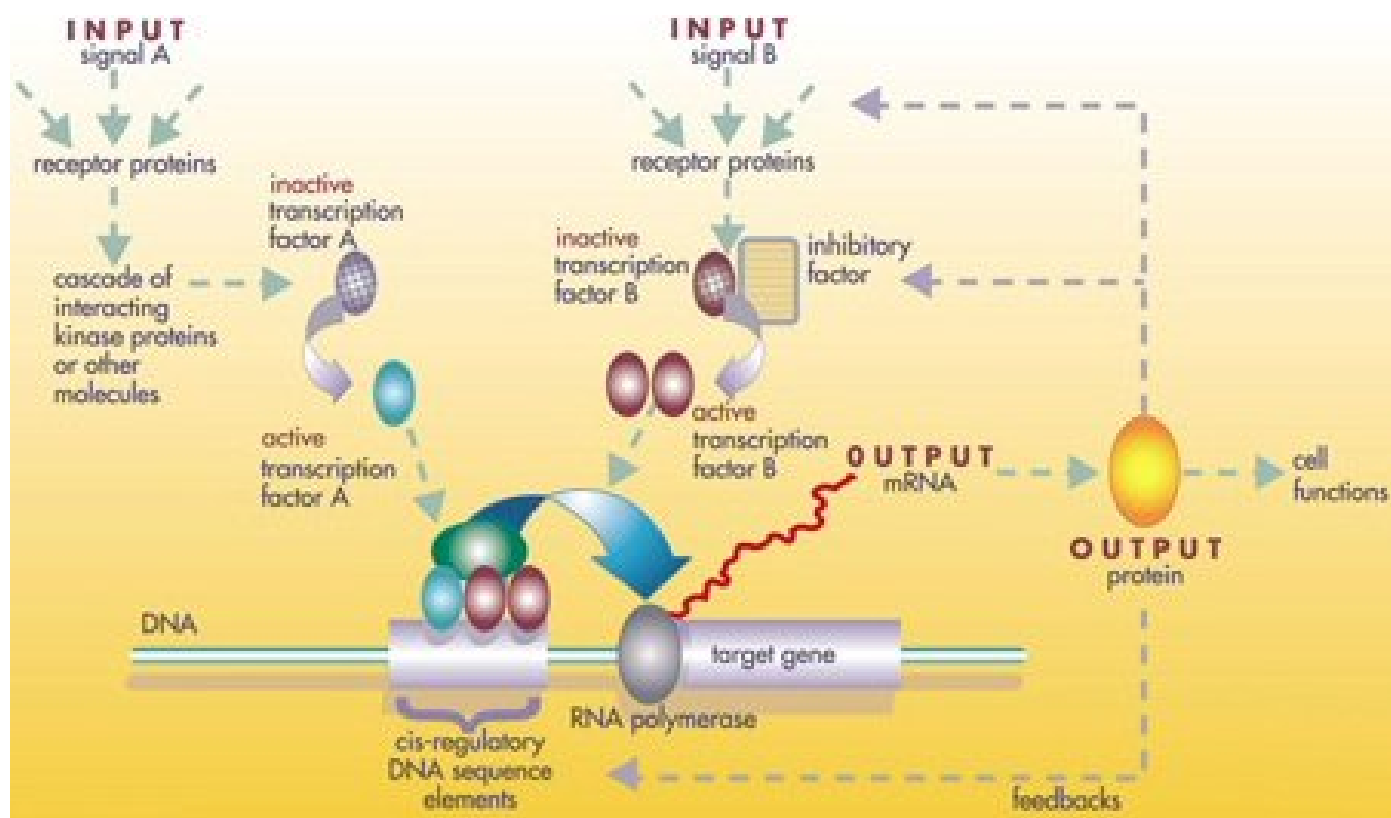


# Общее определение

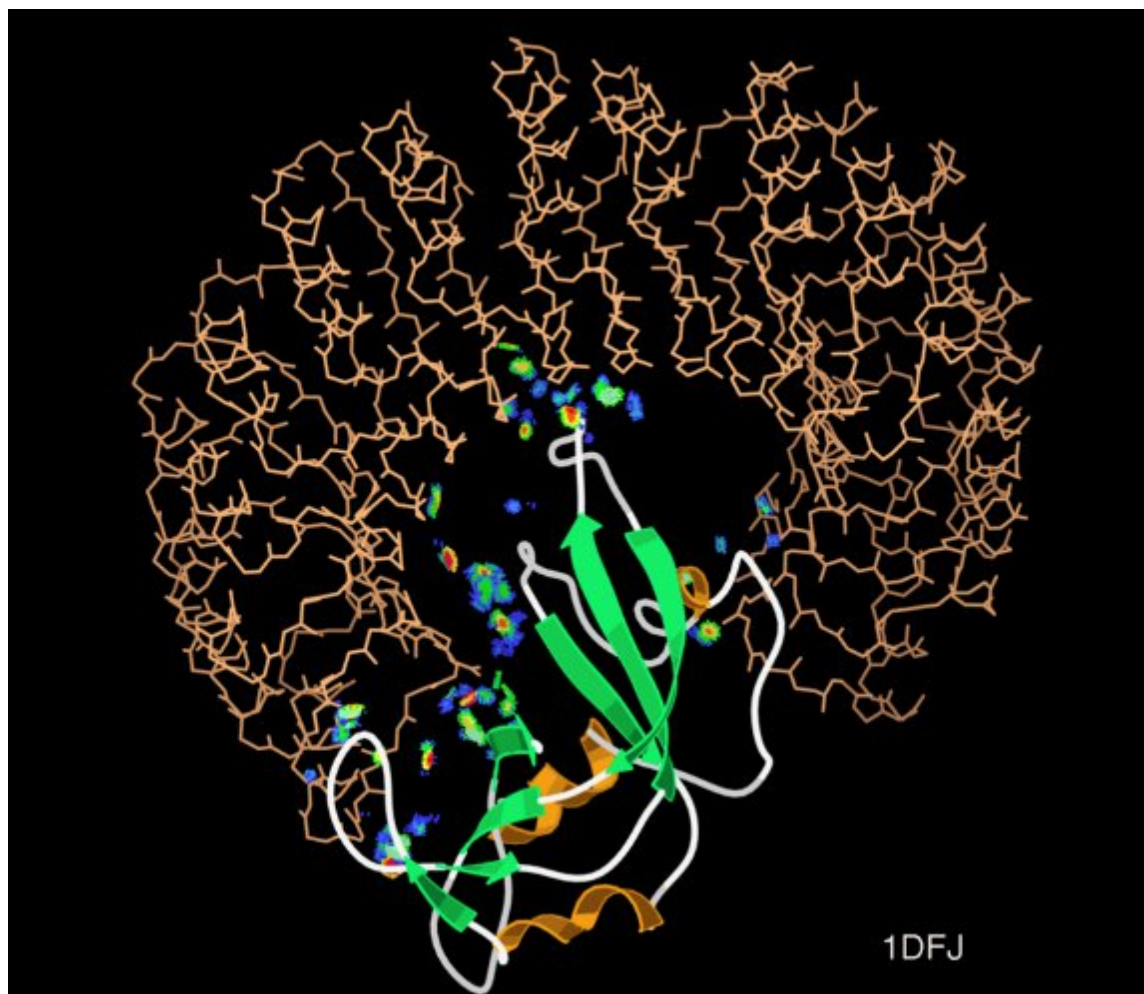
Эффект одной мутации зависит от аллели на другом локусе



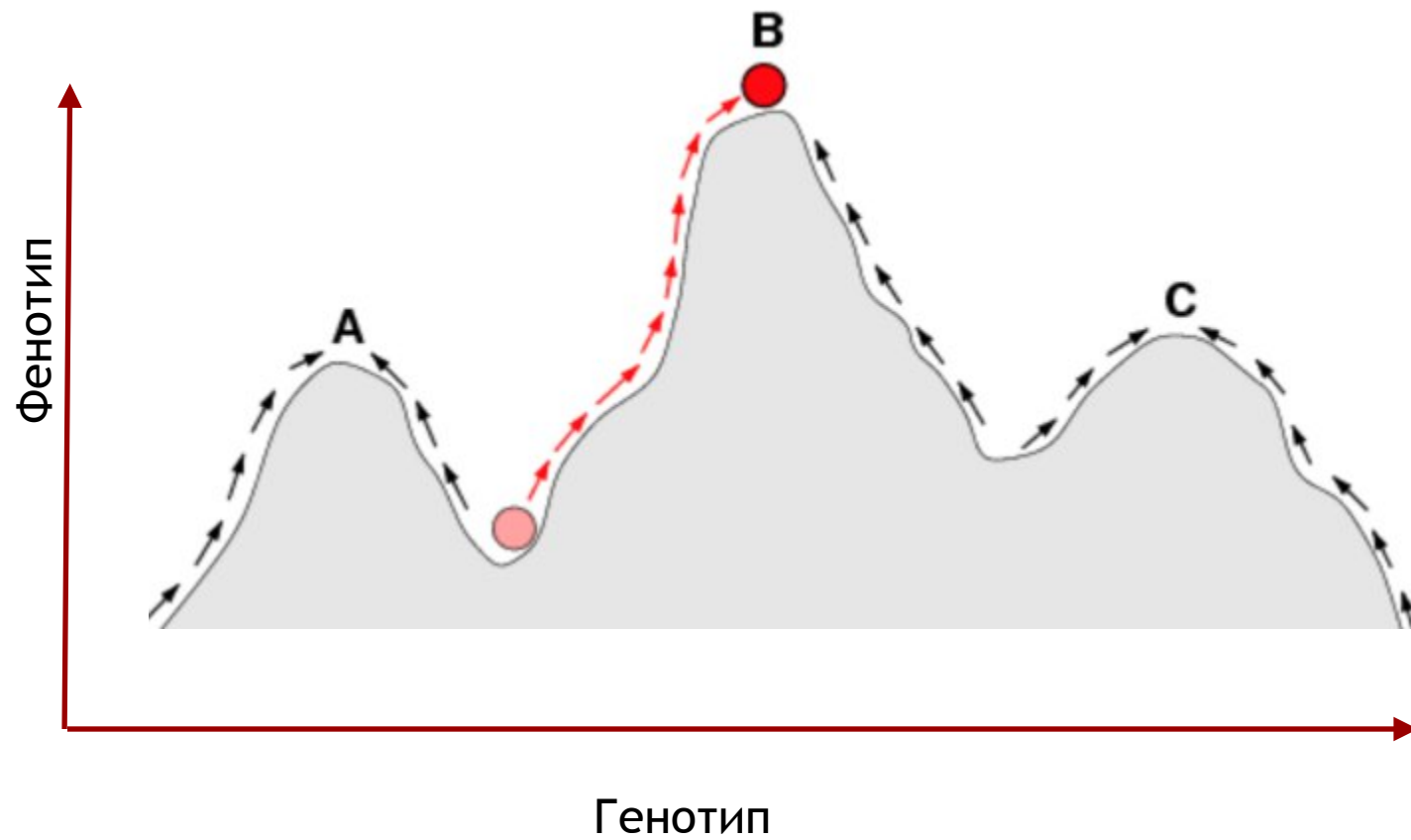
# Ex 1: взаимодействие генов



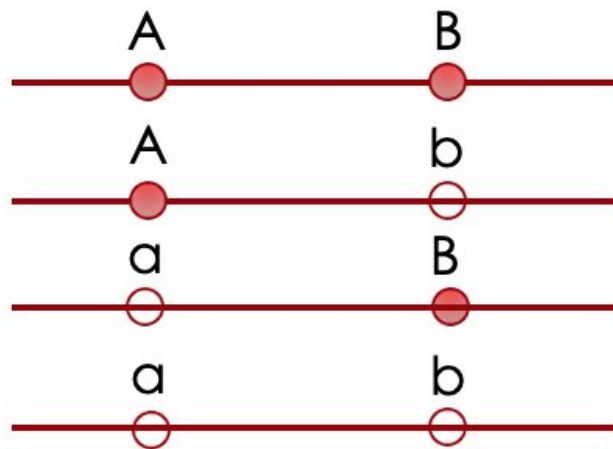
## Ex 2: взаимодействие белков



# Фенотип и генотип



# Two-locus model of epistasis



log fitness

$$W_{00} = 0$$

$$W_{10} = -s_1$$

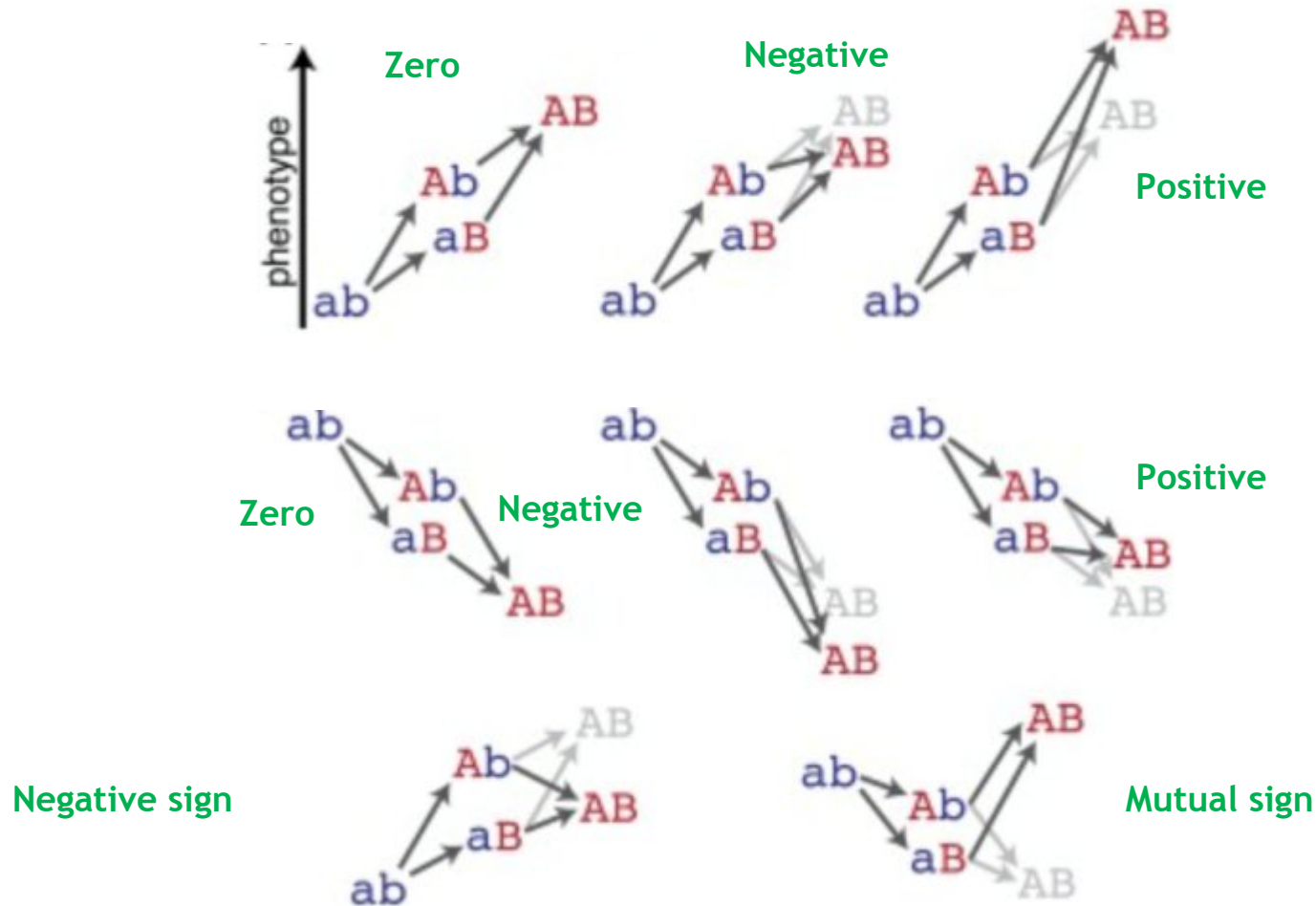
$$W_{01} = -s_2$$

$$W_{11} = -(s_1 + s_2)(1 - E)$$

$E$  - dimensionless epistasis (degree of compensation)



# Парный эпистаз: генетическое определение неаддитивный фенотип (фитнесс)



# Много локусов



і N бинарных последовательностей  $\{K_i\}=0 \text{ or } 1$

і Фитнес  $W\{K_i\}$  
$$W = -\sum_{i=1}^L s_i K_i - \sum_{i<j}^L s_{ij} K_i K_j$$

$$s_{ij} = E_{ij}(s_i + s_j)T_{ij} \quad s_i > 0$$

і Среднее число потомства  $A_{\text{exp}}(W)$  «разломанная палка»

і Случайная мутация

і Рекомбинации нет

$$T_{ij} = 0 \text{ or } 1$$

Матрица взаимодействий

$$E_{ij}$$

Относительная величина эпистаза

$$W = - (s_1 + s_2)(1 - E_{12})$$

$$E = 0 \text{ and } 1$$

Нет эпистаза и 100%

# Меры эпистаза



# Меры 2 локусной корреляции



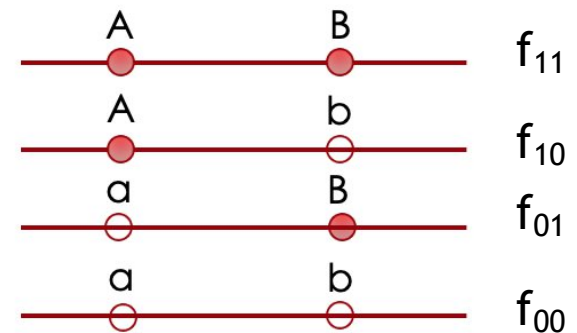
$$D = f_{ij} - f_i f_j$$

Коэффициент Пирсона

$$r = D / \sqrt{f_i(1 - f_i)f_j(1 - f_j)}$$

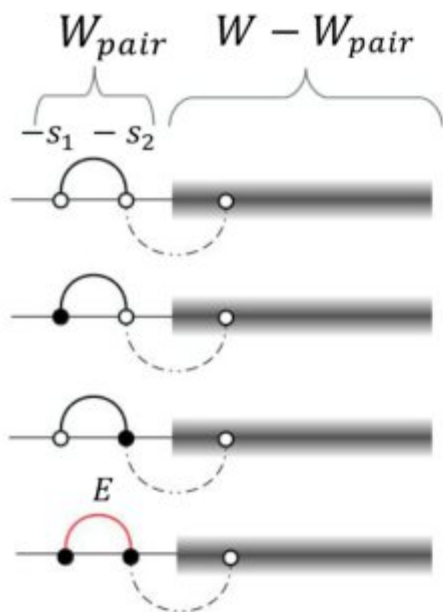
«Универсальный след»

$$\text{UFE} = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)}$$



Все они обнуляются если  $f_{ij} = f_j f_i$

# Вывод величины эпистаза через средние доли гаплотипов в приближении квазиравновесия



$$\begin{aligned}
 W_{00} &= 0 \\
 W_{10} &= -s_1 \\
 W_{01} &= -s_2 \\
 W_{11} &= -(s_1 + s_2)(1 - E)
 \end{aligned}$$

C

$$S(W - W_{\text{pair}}) \approx S(W) - \beta W_{\text{pair}}$$

Вывод на доске

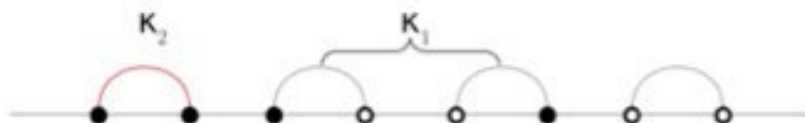
D

$$\begin{aligned}
 f_{00} &= \alpha e^{S(W)} \\
 f_{10} &= f_{00} e^{-\beta s_1} \\
 f_{01} &= f_{00} e^{-\beta s_2} \\
 f_{11} &= f_{00} e^{-\beta(1-E)(s_1+s_2)}
 \end{aligned}$$

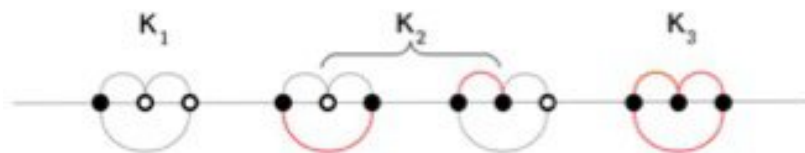
$$E = 1 - \frac{\log\left(\frac{f_{11}}{f_{00}}\right)}{\log\left(\frac{f_{10}f_{01}}{f_{00}^2}\right)}$$

≡

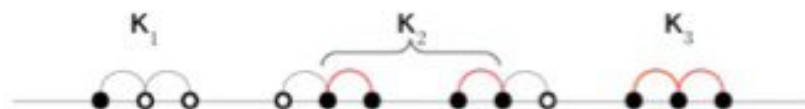
# Примеры топологии сети



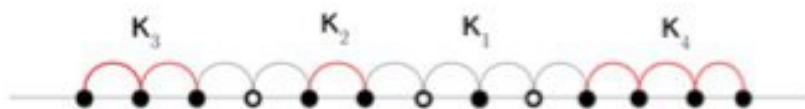
Арчи



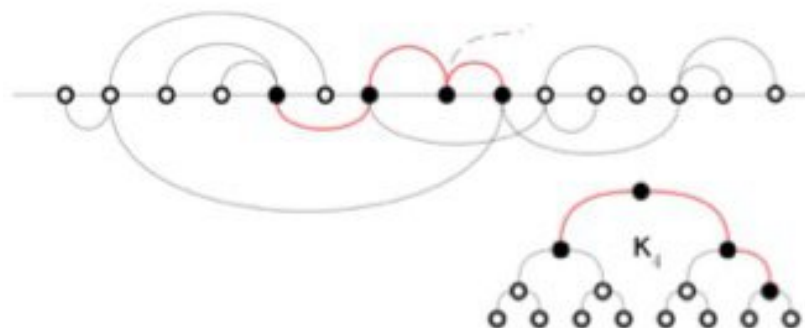
Треугольники



Двойные арчи



Цепь



Бинарное дерево

# Общий подход для однородной топологии



$$W \equiv -s_0 f_0 L = -s_0 \sum_{i=1}^{i_{\max}} k_i (i - 2E b_i)$$

Фитнесс фиксирована

$$e^S = \prod_{i=1}^{i_{\max}} C_{L_i}^{k_i} (n_i)^{k_i}$$

Число комбинаций, максимум по  $k_i$



$$f = \frac{1}{L} \sum_i i k_i \ll 1$$

Доля аллелей

$$f_{11} = \frac{1}{L_{\text{pair}}} \sum_i k_i b_i$$

Доли гаплотипов

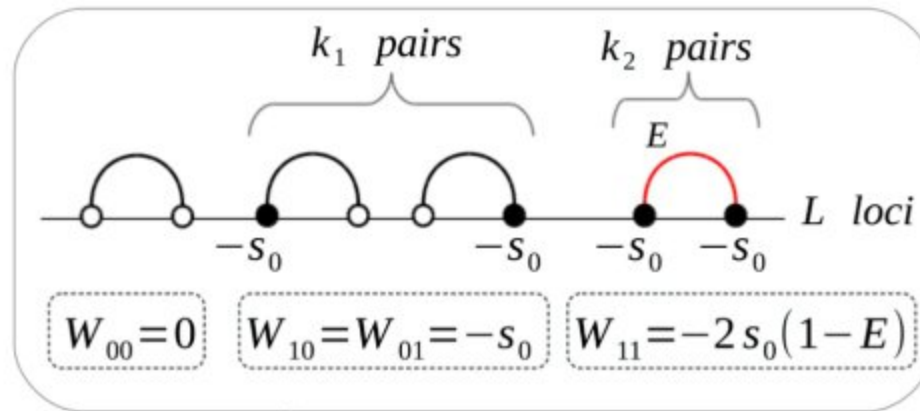
$$f_{10} = f_{01} = f - f_{11}$$

$$\text{UFE} = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)}$$

$$D_{11} = \frac{f_{11}}{f^2}, \quad D_{10} = \frac{f_{10}}{f(1-f)},$$

Корреляторы

# Пример: изолированные пары



Домашнее задание!

**A** Total fitness  $f_0 L \equiv -\frac{W}{s_0} = k_1 + 2(1-E)k_2$

**B** Entropy  $S(k_1, k_2) = \log(C_L^{k_1} C_{L/2}^{k_2})$

**C**  $L \gg k_1, k_2$

**D**  $dS \approx dk_1 \log\left(\frac{1}{f_{10}}\right) + dk_2 \log\left(\frac{1}{f_{11}}\right)$

**E**  $dk_1 = -2(1-E)dk_2$

**F**  $f_{11} = f_{01}^{2(1-E)}$   $dS=0$  at SS

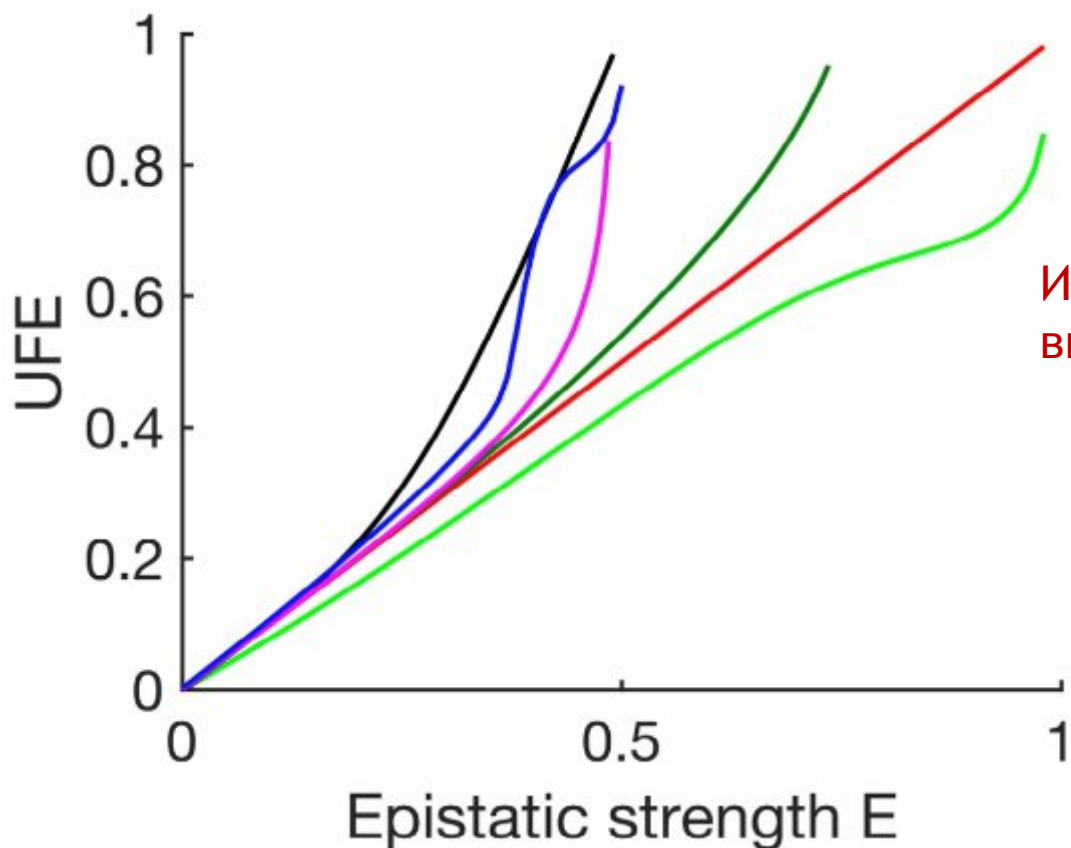
**G** Mutant frequency  $f = f_{01} + f_{11}$

**D**  $f_{10} = f_{01} = k_1/L$ ,  $f_{11} = 2k_2/L$  are haplotype frequencies



# UFE и топология сети

$$UFE = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)}$$



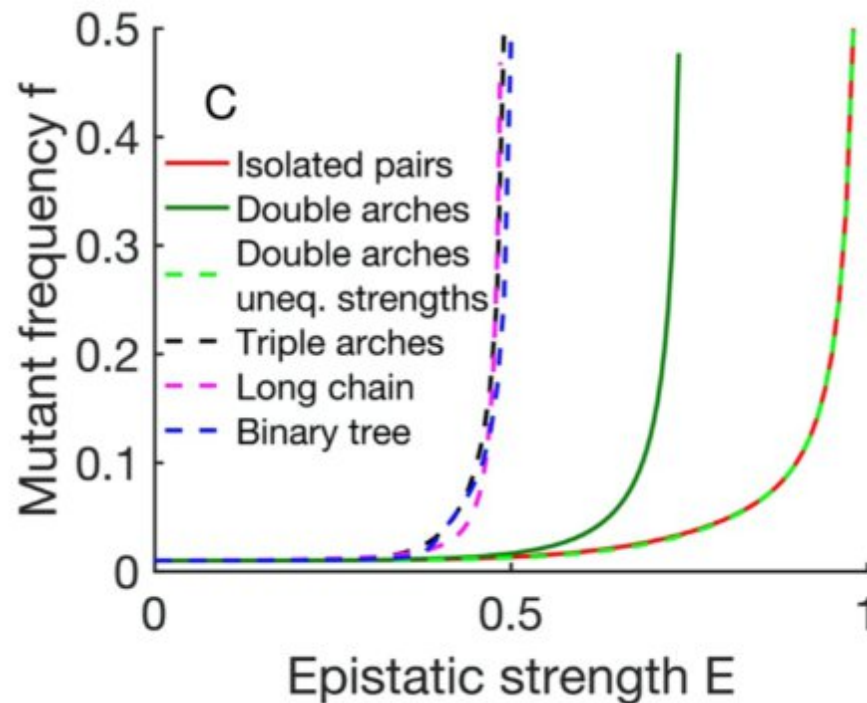
- арки
- двойные арки
- они же, с неравной величиной
- треугольники
- цепь
- бинарное дерево

Интервал где кластеры размером выше двух малочисленны:

$$UFE = E, E < E_{UFE}$$

$$E_{UFE} = \min_{i \geq 2} \left[ \frac{i-2}{2(b_i-1)} \right] \leq 1/2$$

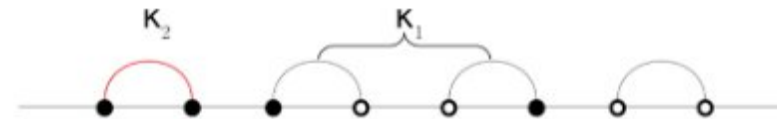
# Доля аллелей и критическая точка



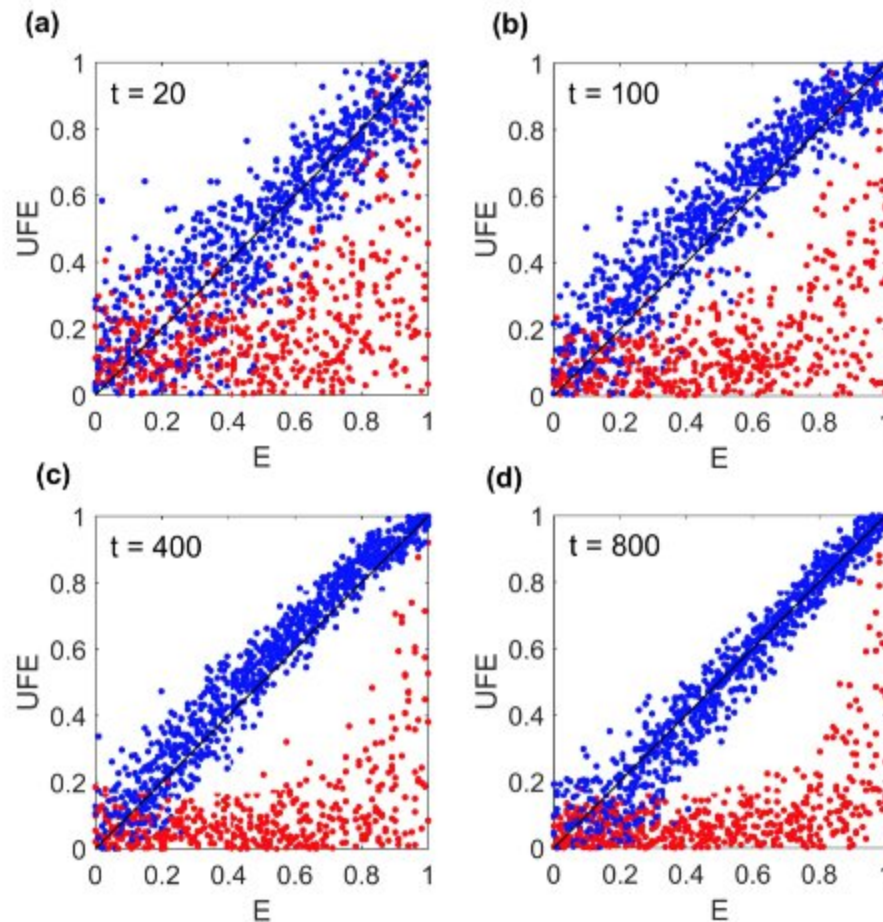
Один из кластеров дает нулевой вклад в фитнесс

$$E_c = \min_i [i / (2b_i)]$$

# Проверка UFE на моделированных последовательностях



$$UFE = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)}$$



$$L = 300, s_0 = 0.05, N = 500, \mu L = 0.5, f(0) = 0.5$$

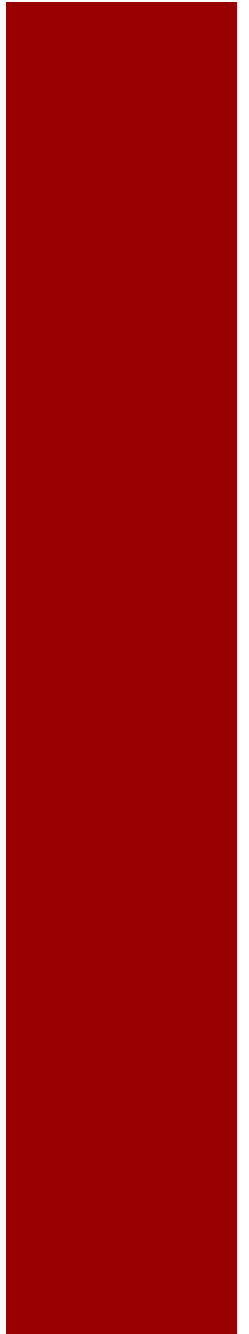
# Выводы I



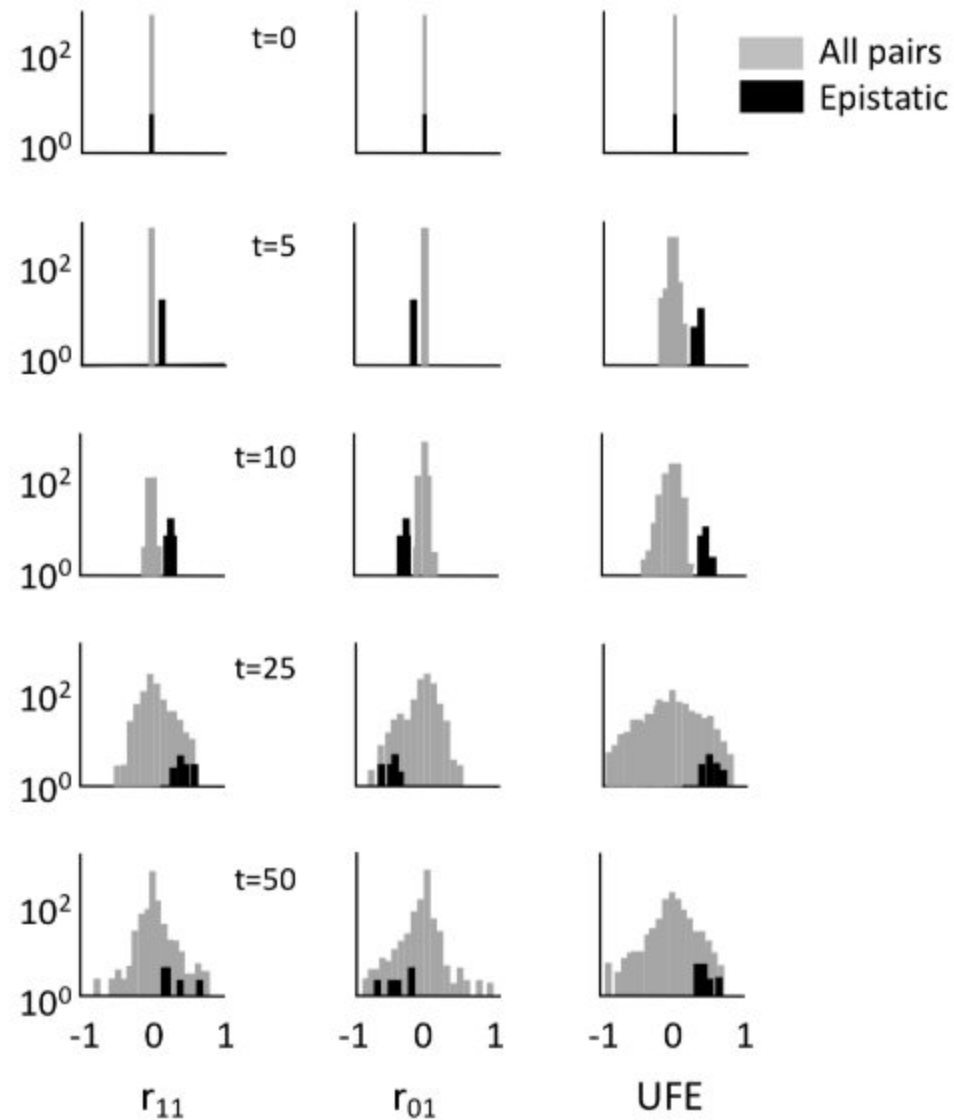
- і Эпистаз это неаддитивность лог фитнес по мутациям из за биологического взаимодействия локусов (фенотипа)
- і Классифицируется одним безразмерным параметром,  $E$
- і Вводится формально добавкой недиагонального члена в фитнес
- і Приближение квазиравновесия работает неплохо
- і  $UFE = E$  при достаточном усреднении и изолированных взаимодействиях
- і Более сложную топологию можно учесть

# Иголка в стогу сена

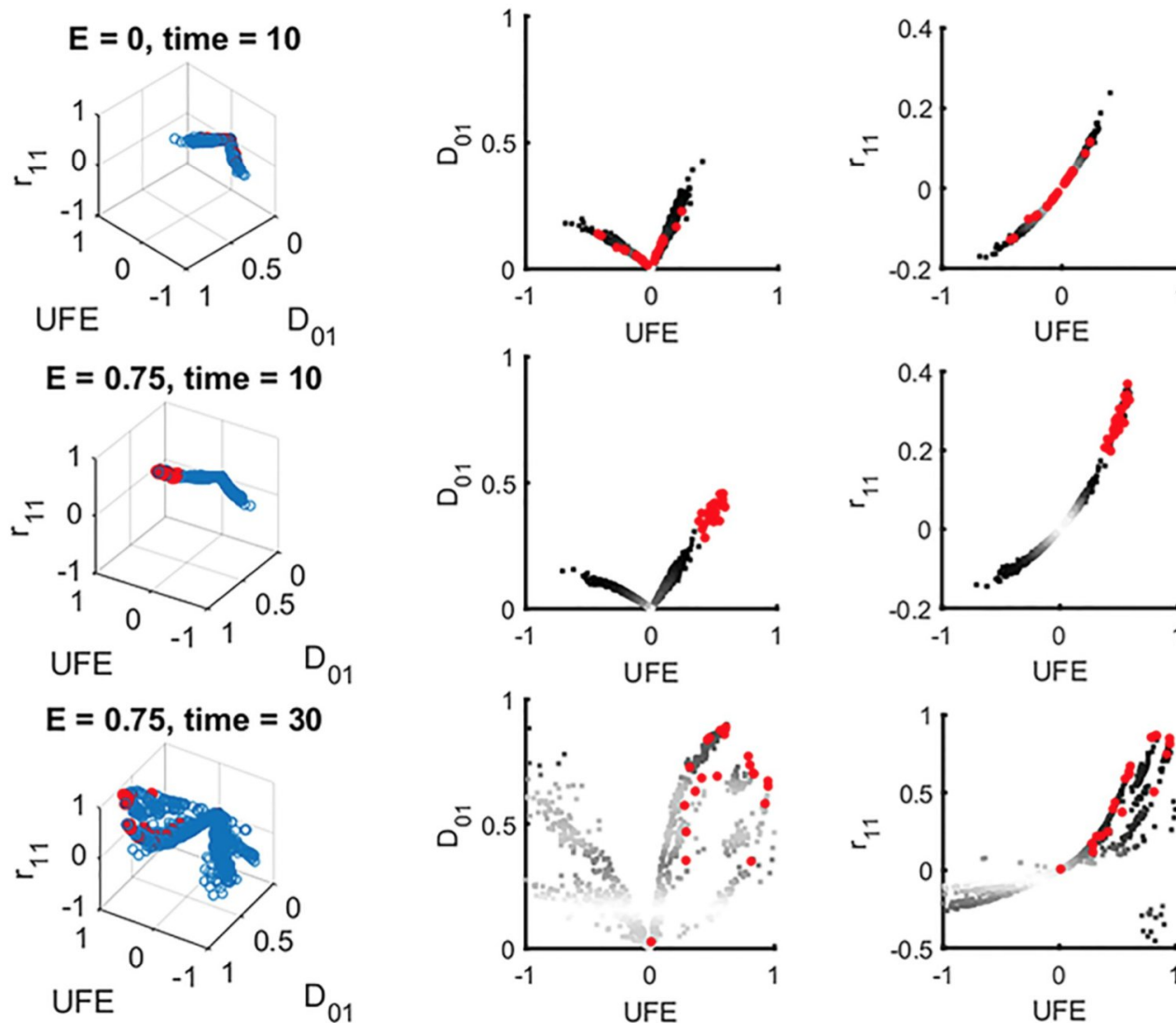
Генетическое сцепление прячет эпистаз



# Эпистаз затмевается генетическим сцеплением



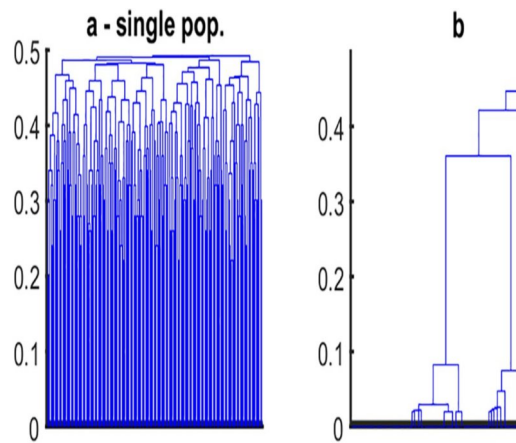
# Using a few measures together does not help at all



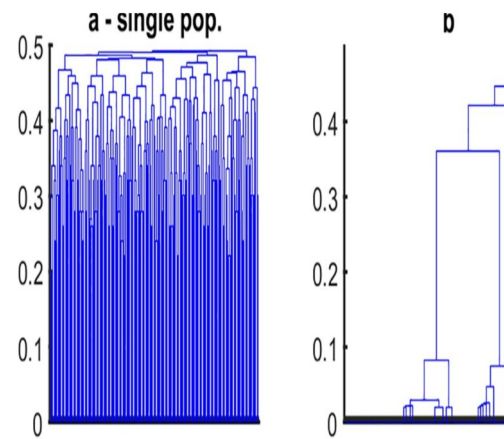
# Причина: стохастическая родословная



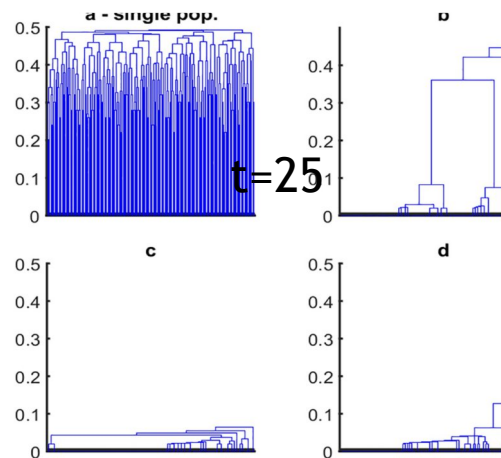
t=0



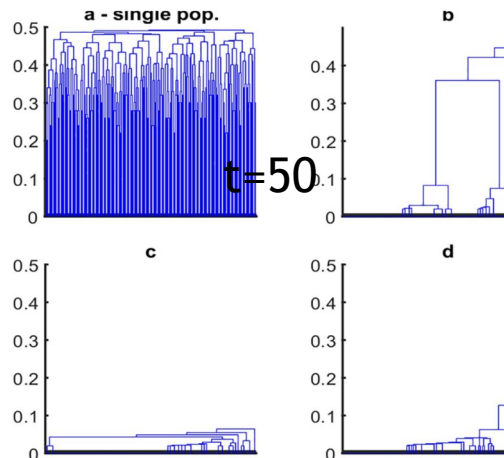
t=5



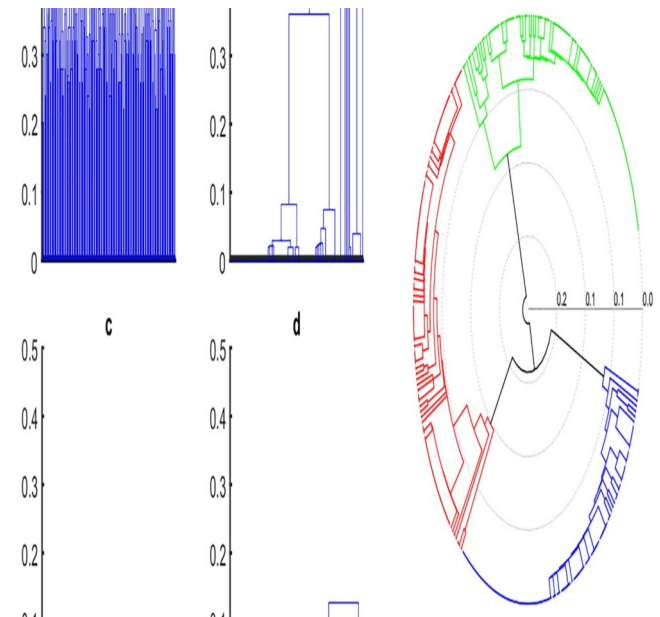
t=25



t=50

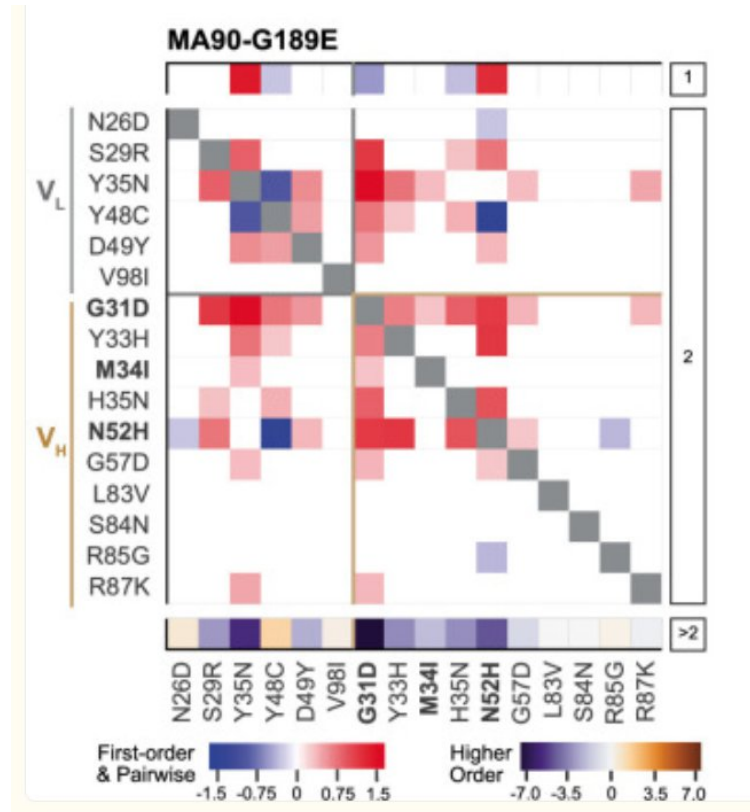


3 популяции





# Escape mutations from SARS-CoV-2, enrichment



Epistasis or linkage?

Hierarchical sequence-affinity landscapes shape the evolution of breadth in an anti-influenza receptor binding site antibody. [Angela M Phillips](#) [Daniel P Maurer](#) [Caelan Brooks](#), [Thomas Dupic](#), [Aaron G Schmidt](#), [Michael M Desai](#)



# Как же обнаружить эпистаз?

- і Усреднить  $f_{ij}$  по многим независимым популяциям
- і Парные корреляции (UFE, Pearson coefficient)
- і Для длинных геномов,  $L \gg 1$ , слишком много популяций надо!
- і Проклятие размерности:  $L$  пар среди  $L^2/2$  ВОЗМОЖНЫХ

# Популярный метод: Квази-равновесие генетического сцепления (Кимура 1967)



- і Популяция очень большая
- і Есть сильная рекомбинация и эпистаз, в противоборстве
- і Предположение: эффектов генсцепления нет совсем
- і Пишем функцию распределения типа модели Изинга для частот гаплотипов и подогнать параметры к генетическим данным
- і Проблемы: работает частично для далеких видов, слишком много параметров для подгонки
- і Обзор: Neher RA, Shraiman BI. Statistical genetics and evolution of quantitative traits. Rev of Modern Physics. 2011; 83:1283-300

# Трех-локусный метод

вблизи точки полной компенсации

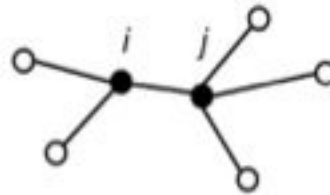


# Два шага

Усредним  $f_{ij}$  по 20-200 популяциям

Pair-wise association analysis

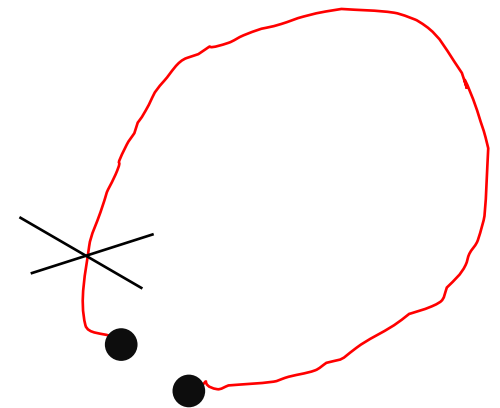
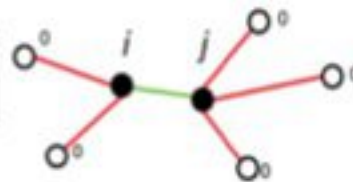
$$UFE_{ij} = 1 - \frac{\log(f_{11}/f_{00})}{\log(f_{01}f_{10}/f_{00}^2)}$$



Для каждой пары, берем только геномы с 0 на одном из соседних локусов

Removing indirect links

$$UFE_{ij0} = \min \left[ 1 - \frac{\log(f_{110}/f_{000})}{\log(f_{010}f_{100}/f_{000}^2)} \right]$$

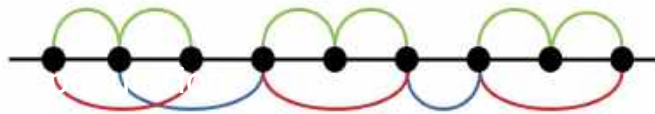


... и найти минимум по соседям

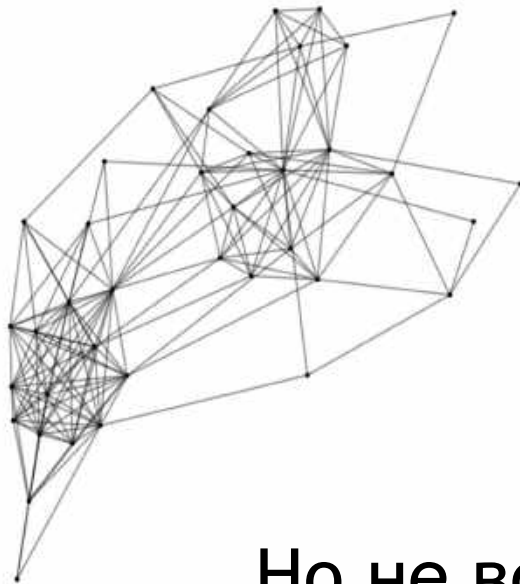
# Усреднение $f_{ij}$ по популяциям убирает большинство ложно-положительных корреляций



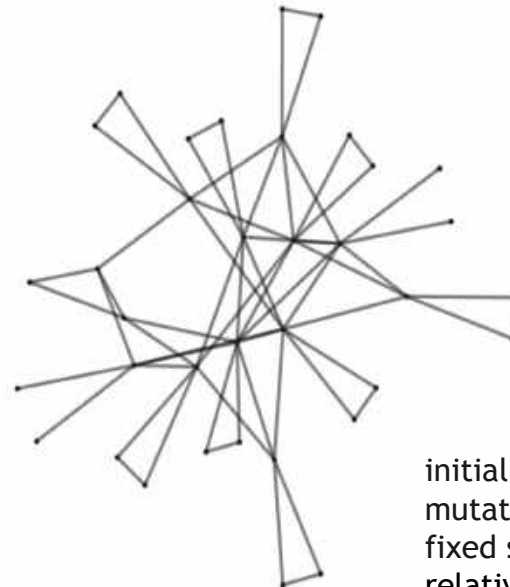
L=40, 200 популяций



— Epistatic link  
— Indirect link  
— Stochastic LD

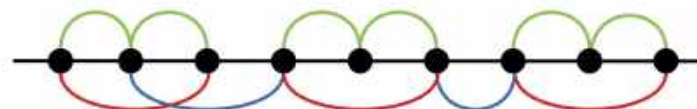
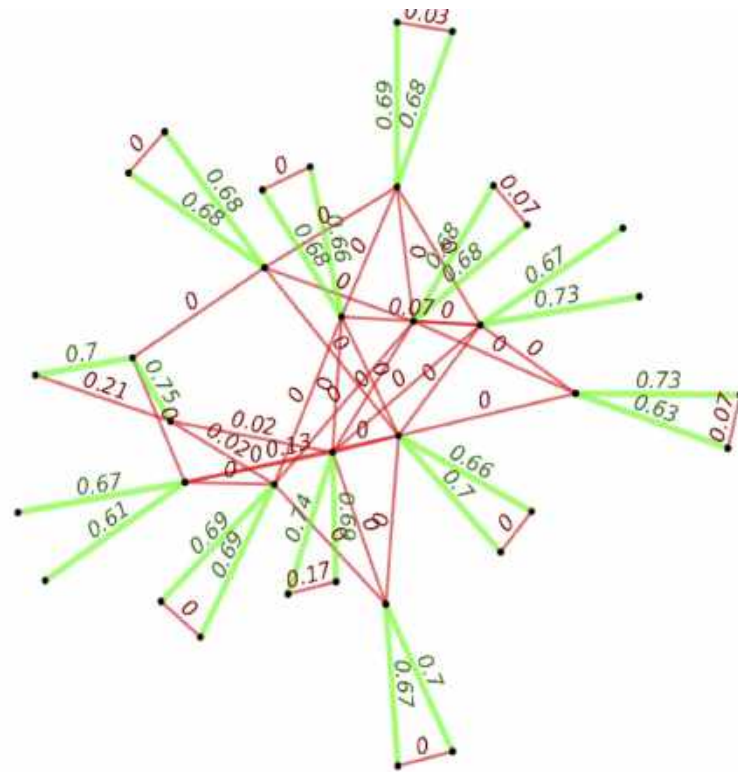
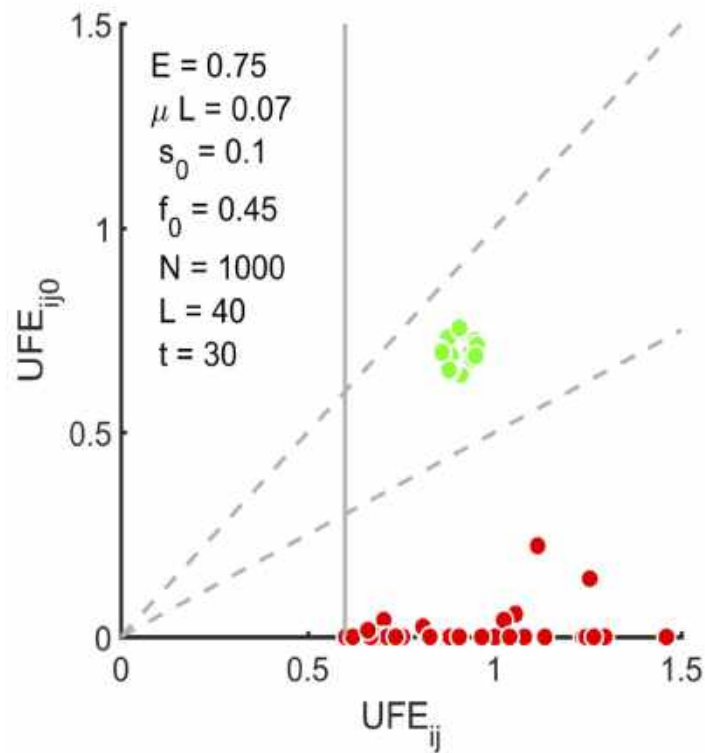


Но не все!



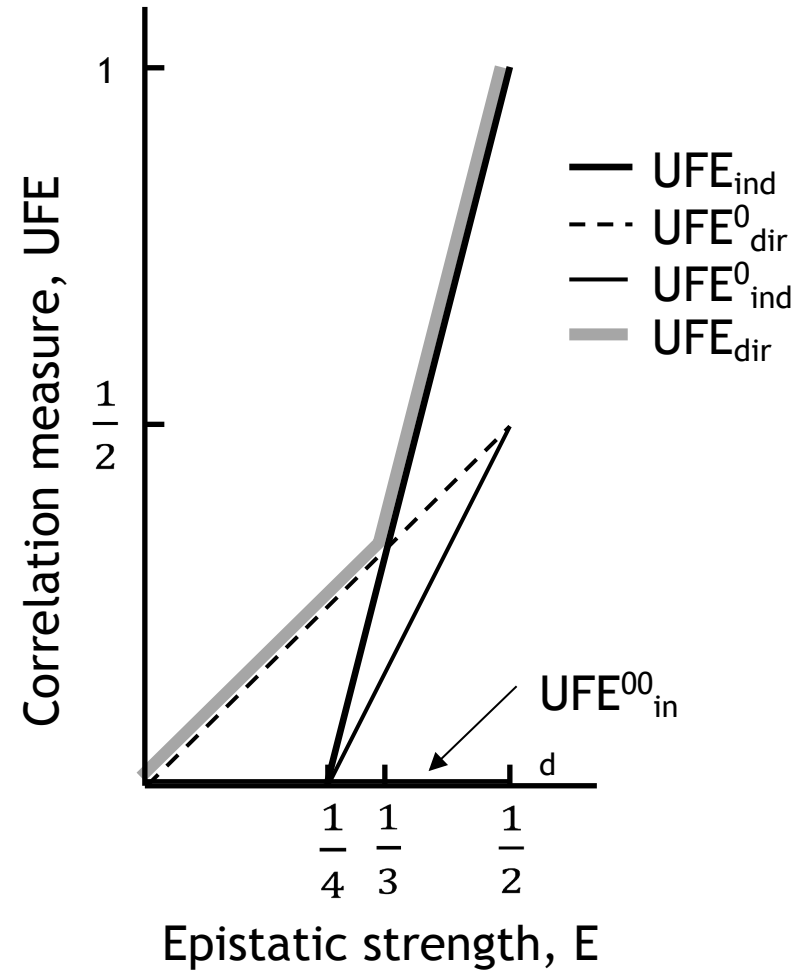
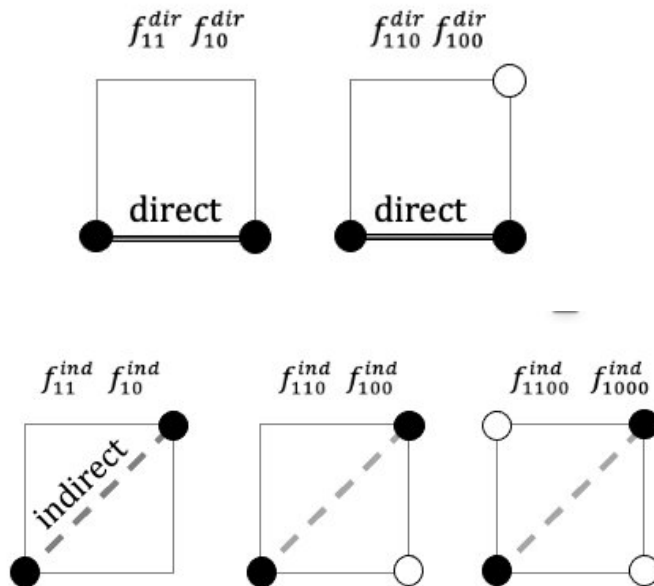
initial allele frequency  $f_0=0.45$   
mutation rate per genome  $\mu L=0.07$   
fixed selection coefficient  $s=0.1$   
relative epistatic strength  $E=0.75$   
 $N=1000$  individuals  
 $L=40$  sites

# 100% успех!



- Epistatic link
- Indirect link
- Stochastic LD

# Связность 2: 4-локуса





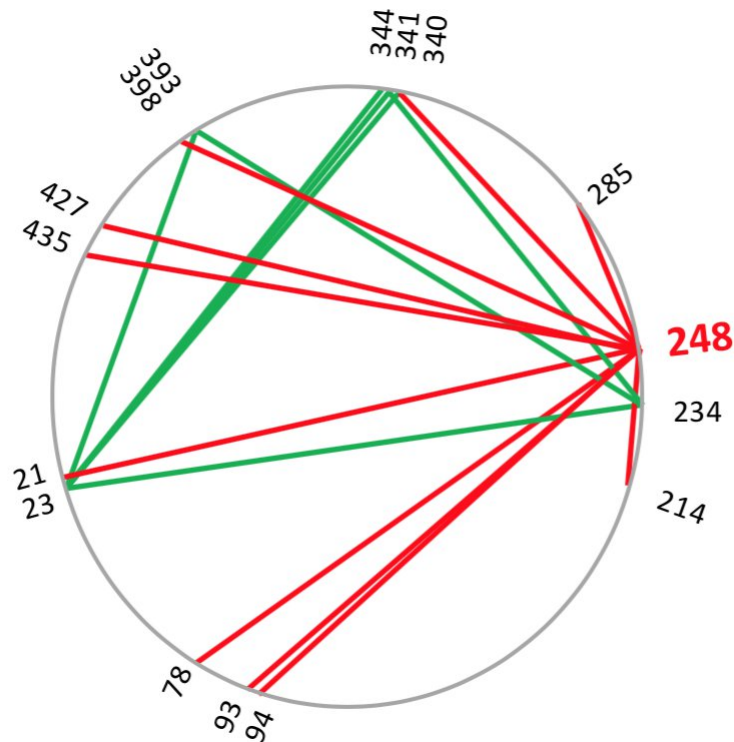
# Применение к нейроминидазе вируса гриппа А Н1N1



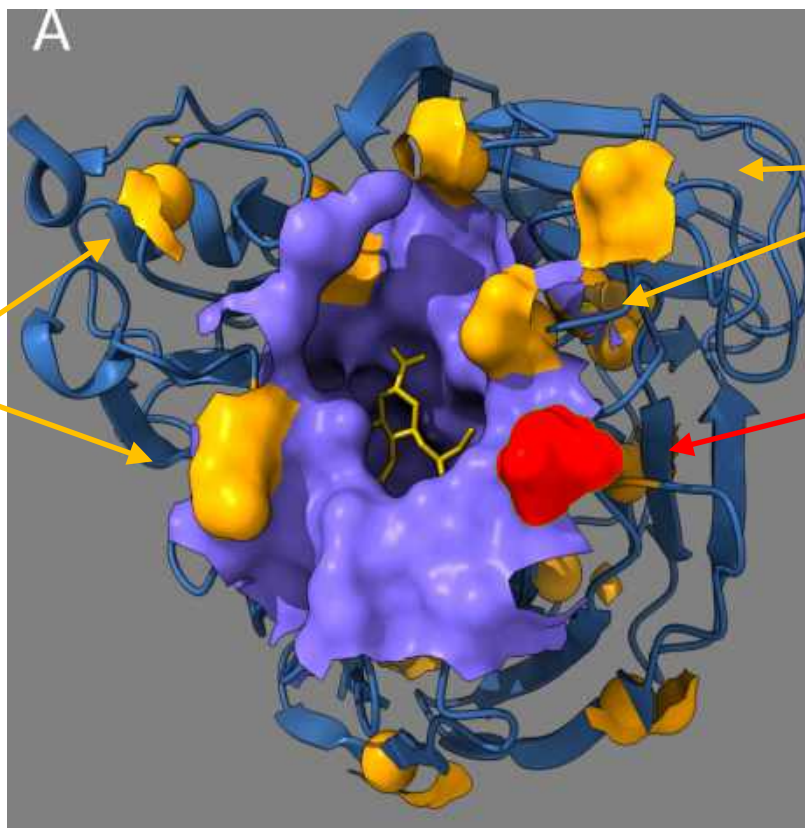
РНК данные 2005-2010

2009: новый пандемический штамм с 20% разницей

Примерно равное количество каждого штамма



# Структура нейроминидазы вируса гриппа



# Выводы II

- і След эпистаза затмевается корреляциями от эффектов генетического сцепления
- і Усреднение долей всех гаплотипов по 20-200 независимых популяций сильно уменьшает эти фальшивые связи
- і Но оставшиеся все равно прячут эпистаз
- і Новый трех-локусный метод, при простой топологии сети, может обнаружить эпистаз с высокой надежностью
- і Он простой, общий, быстрый, и был успешно применен к штаммы пандемии гриппа 2009

# Спасибо

