

Слайд 1.

Настоящий раздел посвящен введению в предмет молекулярной эволюции.

Слайд 2.

Под термином эволюция в науке понимается изменение чего-либо во времени. Предметом теории молекулярной эволюции является изучение закономерностей эволюционных изменений генетических макромолекул в живых организмах. В основе лежит эволюционная теория - комплекс знаний об общих закономерностях и движущих силах исторического развития живой природы. Основой эволюционной теории служит утверждение о том, что все ныне существующие организмы произошли от ранее существовавших путем длительного их изменения под воздействием внешних и внутренних факторов.

Слайд 3.

Основными задачами теории молекулярной эволюции являются изучение закономерностей эволюции генетических макромолекул, а также реконструкция эволюционной истории генов и организмов.

Слайд 4.

При решении этих задач используются результаты исследований в других областях науки: палеонтологии, генетики, молекулярной биологии, биофизики, математики и информатики.

Слайд 5.

Наибольшее влияние на развитие идей эволюции оказало учение Ч. Дарвина, которое было сформулировано им в работе "Происхождение видов" в 1859 году.

В основе учения Дарвина лежали следующие положения.

В пределах каждого вида существует изменчивость по морфологическим, физиологическим и др. признакам. Эти вариации случайные.

Особь дает потомство, которое наследует признаки родителей. Эффективность воспроизводства особей называют приспособленностью.

Организмы дают большее потомство, чем это необходимо для простого воспроизводства численности вида.

Из-за ограниченности внешних ресурсов возникает естественный отбор, который приводит к выживанию и размножению наиболее приспособленных особей.

В целом, теорию эволюции Дарвина можно рассматривать как "хаос с обратной связью".

Слайд 6.

Обратимся теперь к объектам исследования молекулярной биологии. В настоящем разделе мы будем рассматривать особенности эволюции ДНК и белков. Эти молекулы выполняют специфические функции. ДНК является носителем генетической информации. Белки выполняют структурные, регуляторные, каталитические функции. Специфика этих молекул (их структурная организация и функции) определяет и особенности их эволюции. Поэтому важной особенностью белков является эволюция их структур. Мы так же коснемся особенностей эволюции белков в составе полных геномов организмов.

Слайд 7.

Согласно основной догме молекулярной биологии, генетическая информация реализуется по принципу ДНК->РНК->белок. Процесс формирования молекулы РНК называется транскрипцией, затем на рибосоме происходит синтез белка по матричной РНК.

Слайд 8.

Генетическая информация записана в последовательности ДНК сложным образом. ДНК может содержать часть, которая кодирует другие генетические макромолекулы – РНК и белки. Доля кодирующей части ДНК варьируется от организма к организму и у человека составляет менее 2%. Некодирующая часть может содержать регуляторные районы, повторяющиеся участки. Функция некоторых районов ДНК пока неизвестна.

Слайд 9.

Ключевым в биологии является понятие гена. Его определение изменялось с накоплением молекулярно-биологических знаний. Согласно современным представлениям ген - это

сегмент ДНК, который обуславливает фенотип или функцию. В отсутствии данных о функции ген может быть охарактеризован последовательностью, транскриптом или по гомологии. Количество генов в разных организмах различно. Так, например, у человека предполагается около 30000 генов.

Слайд 10.

Гены кодируют белки, РНК и имеют сложную структуру. На рисунке представлено схематически строение гена эукариот. В структуре гена можно выделить кодирующие участки (экзоны), некодирующие участки (интроны), регуляторные районы (промотор). В кодирующей части есть стартовый кодон, определяющий начало аминокислотной последовательности и стоп-кодон, который ее завершает.

Слайд 11.

Последовательность белка кодируется в последовательности матричной РНК по правилам генетического кода. Каждая тройка нуклеотидов (кодон), кодирует аминокислоту. Генетический код является вырожденным, т.е. несколько кодонов могут кодировать одну аминокислоту. Разные аминокислоты могут кодироваться разным числом кодонов. Например, триптофану соответствует один кодон – UGG, а лейцину – 6. Отметим, что тип кодируемой аминокислоты определяется чаще всего тремя первыми нуклеотидами и слабо зависит от третьей позиции. Генетический код так же содержит три стоп кодона, которые прекращают трансляцию белка.

Слайд 12.

Отдельные изменения генотипа называются мутациями. Мутация - основа наследственной изменчивости в живой природе. Мутации могут приводить к изменениям больших сегментов ДНК. К таким мутациям относятся транслокации - перенос гена в другое место генома; дупликации - удвоение участка гена; инверсия - поворот участка гена на 180 градусов; делеции – удаление участка гена.

Существует так же класс точечных мутаций, которые сводятся к заменам одного нуклеотида в последовательности ДНК.

Слайд 13.

Точечные замены нуклеотидов могут быть нескольких типов.

Во-первых, мутация может приводить к замене аминокислоты в последовательности белка (миссенс-мутация). Мутация может приводить к возникновению стоп-кодона (нонсенс-мутация). В третьих, вставки или делеции 1-2 нуклеотидов могут приводить к сдвигу рамки считывания белка. В четвертых, вставка или делеция трех нуклеотидов может приводить к вставке/делеции одной аминокислоты. На рисунке приведены несколько примеров мутаций разных типов, которые вызывают наследственное заболевание "кистозный фиброз".

Слайд 14.

Перейдем теперь к математическим моделям, описывающим эволюцию последовательностей ДНК, РНК и белков. Цель заключается в наиболее полном описании процесса одиночных нуклеотидных/аминокислотных замен.

В основе большинства моделей лежит теория Марковских цепей с конечным числом состояний. Полнота описания достигается как правило за счет усложнения моделей и переходу к большему числу параметров модели.

Слайд 15.

Одним из важнейших параметров теории молекулярной эволюции является эволюционное расстояние между последовательностями. Это расстояние определяется при сравнении нуклеотидных (аминокислотных) последовательностей и в общем случае зависит от того, какая модель лежит в основе описания эволюционного процесса. Эволюционное расстояние определяет взаимосвязь между различиями на уровне последовательностей ДНК/РНК/белков и временем (геологическим) эволюции видов. Очевидно, знание такой зависимости позволяет производить датировку молекулярных событий (расхождение видов) на основе сравнения их последовательностей макромолекул.

В простейшем случае, когда замены редки (или время эволюции мало) можно предположить, что число замен в паре последовательностей прямо пропорционально времени их эволюции. На рисунке показана родительская последовательность (изменившиеся позиции показаны синим) и две дочерних, которые после дивергенции видов от родительской эволюционировали независимо (изменения последовательности показаны красным). Заметим, что расстояние между двумя дочерними последовательностями отражает удвоенное время, прошедшее с момента дивергенции (при условии, что скорости накопления замен в двух последовательностях были одинаковы). Для определения времени эволюции в такой модели достаточно знать долю p несовпадающих символов в двух последовательностях. Расстояние будет вычисляться как $d = -\log_e(1-p)$. Зная скорость замен α можно вычислить время эволюции, или зная d и t можно вычислить скорость замен. Недостатком такого подхода является его неприменимость на больших временах эволюции, когда сказываются повторные и обратные замены нуклеотидов.

Слайд 16.

Более точному расчету эволюционных расстояний способствует построение более точных математических моделей эволюции и их теоретический анализ.

Начнем с рассмотрения простейшей модели рассмотрим модель Джукса-Кантора для последовательности нуклеотидов бесконечной длины. Эту модель можно описать одним параметром – вероятностью замены α одного нуклеотида на другой за единичный промежуток времени. Этот параметр одинаков для всех возможных пар замен (представленных в виде схемы слева). Такая модель была предложена Джуксом и Кантором и поэтому носит их имя. Отметим, что замены в позиции происходят случайным образом, независимо от других позиций последовательности. Поэтому вероятности наблюдать нуклеотиды можно оценить по частотам их встречаемости в последовательности.

Рассмотрим, как зависит вероятность наблюдать нуклеотид А в момент времени $t+1$ - $P_A(t+1)$, при условии, что в момент времени t частота нуклеотида А известна и равна $P_A(t)$.

$$P_A(t+1) = (1 - 3\alpha)P_A(t) + \alpha[1 - P_A(t)] = -4\alpha P_A(t) + \alpha$$

Здесь первое слагаемое отражает уменьшение частоты нуклеотида А за счет мутаций в любой из трех других нуклеотидов (A->X), второе слагаемое отражает увеличение частоты нуклеотида А за счет мутаций трех других нуклеотидов (X->A). Если перейти к непрерывному времени, то это выражение сведется к дифференциальному уравнению

$$\frac{dP_A(t)}{dt} = -4\alpha P_A(t) + \alpha$$

Слайд 17.

Решение этого уравнения записывается как

$$P_A(t) = 0.25 + (P_A(0) - 0.25)e^{-4\alpha t}$$

Где $P_A(0)$ – вероятность наблюдать нуклеотид А в момент времени $t=0$. Если $P_A(0) = 1$, то $P_A(t) = 0.25 + 0.75e^{-4\alpha t}$ (верхний график на рисунке слева), Если $P_A(0) = 0$, то

$P_A(t) = 0.25 - 0.25e^{-4\alpha t}$ (нижний график на рисунке слева). Отметим ряд важных особенностей модели эволюции. Во-первых, при большом времени эволюции, частоты всех нуклеотидов стремятся к своим равновесным значениям, одинаковым и равным $1/4$.

Во вторых, зависимость $P_{AA}(t) = P_{ii}(t)$ (вероятность нуклеотида остаться неизменным через промежуток времени t) не зависит от типа нуклеотида. В третьих, вероятность нуклеотида мутировать на любой другой $P_{ij}(t)$ так же не зависит от конкретного типа нуклеотида.

Слайд 18.

Известно, что в ДНК частоты замен зависят от типа нуклеотидов. В частности частоты замен типа $A \leftrightarrow T$ и $A \leftrightarrow G, C$ различаются. Замены типа $A \leftrightarrow T$ и $G \leftrightarrow C$ называются трансверсиями, типа $A \leftrightarrow G, C$ – транзициями. Различие частот трансверсий и транзиций было выявлено при сравнении ДНК разных организмов. Для описания таких различий Кимура модифицировал модель нуклеотидных замен, введя в нее дополнительный параметр. В модели Кимуры вероятности транзиций α не равны вероятностям трансверсий β . Из этого вытекает изменение динамики замен. Во-первых, поведение модели зависит от двух параметров. Во вторых, если вероятности P_{ii} не зависят от типа нуклеотида i , то P_{ij} уже зависят от того, к какому типу относится замена (транзиция или трансверсия). Однако, неизменным остается экспоненциальный характер зависимости частот нуклеотидов от времени, который приводит к тому, что через бесконечно большой промежуток времени частоты нуклеотидов примут стационарные значения.

Слайд 19.

В общем случае вероятность одномоментных замен нуклеотидов может быть описана т.н. матрицей замен. Размер этой матрицы 4×4 , она описывает линейное преобразование вектора P частот нуклеотидов за короткий промежуток времени (см. выражения для частот слева). В Векторной форме это преобразование записывается как $p' = Mp$, где p – вектор частот в момент времени t , p' – вектор частот в момент времени $t' = t + dt$. Свойства матрицы замен таковы: сумма ее элементов по строкам составляет 1; эволюция за n дискретных промежутков времени равна умножению на матрицу M^n . Если $M = \text{const}$, то существуют равновесные частоты, которые находятся из уравнения

$$p = Mp.$$

Это уравнение можно получить из того факта, что в равновесном состоянии частоты нуклеотидов не зависят от времени и поэтому умножение вектора частот p на матрицу M (что соответствует единичному шагу эволюции) оставляет его неизменным.

Слайд 20.

Рассмотрим матрицу замен для модели Джукса-Кантора.

Матрица скоростей нуклеотидных замен M_0 в этом случае будет иметь одинаковые недиагональные элементы, равные α , и диагональные элементы, равные $1 - 3\alpha$. Вектор равновесных частот имеет одинаковые компоненты, равные 0.25.

Вероятности замен в зависимости от времени были приведены на слайде 19, они образуют две группы зависимостей – P_{ii} (диагональные элементы) и P_{ij} (недиагональные элементы). Теперь мы рассмотрим две бесконечные последовательности нуклеотидов, которые разделены временем эволюции t , пусть частоты нуклеотидов в них все равны 0.25, матрицу вероятностей аминокислотных замен мы рассчитали в явном виде по формулам Джукса-Кантора. Как ли рассчитать долю совпадающих пар, которую можно ожидать при сравнении двух этих последовательностей?

Отметим, что вероятностям совпадений соответствуют диагональные элементы матрицы, а так как частоты нуклеотидов равны (и в одной и в другой последовательности), то для оценки доли совпадений необходимо умножить вероятность нуклеотида на вероятность того, что во второй последовательности ему встретится идентичный (P_{ii}) и просуммировать это выражение по всем 4м нуклеотидам. Поскольку их частоты и диагональные элементы равны, то долю совпадающих пар будет вычисляться как

$$q = 4(0.25 + 0.25e^{-4\alpha t}) \cdot 0.25 = (0.25 + 0.25e^{-4\alpha t}).$$

Тогда параметр αt можно отсюда выразить

$$\text{как } \alpha t = -1/4 \log(4q/3 - 1/3) = -1/4 \log(1 - 4p/3), \text{ а время эволюции как } t = \frac{-1}{4\alpha} \log(1 - \frac{4}{3} p)$$

. Мы видим, что оно однозначно связано с долей накопленных замен p , однако эта зависимость является более сложной, чем в случае модели Пуассона (поскольку учитывает двойные и обратные замены нуклеотидов).

Слайд 21.

Что касается эволюционного расстояния d , то его обычно принято выражать в среднем числе замен на время эволюции t . В модели Джукса-Кантора за единичное время может быть в среднем накоплено 3α замен (сумма недиагональных элементов в матрице скоростей замен M_0), то эволюционное расстояние оценивается как $3\alpha t =$

$d = -\frac{3}{4} \log_e \left(1 - \frac{4}{3} p\right)$. Более сложное выражение получается для модели Кимуры, которая определяется двумя параметрами. Поэтому при определении эволюционного расстояния учитываются частоты транзиций и трансверсий и $d = -\frac{1}{2} \log_e [(1 - 2P - Q)\sqrt{1 - 2Q}]$

Слайд 22.

Одним из важных классов генетических макромолекул являются белки. Это нерегулярные гетерополимеры, состоящие из 20 мономеров – аминокислот. Структура аминокислоты включает основную цепь и боковую группу. Основная цепь одинакова у всех аминокислот. Боковые группы аминокислот различаются по своей химической структуре и свойствам. Именно они определяют физико-химические особенности и взаимодействия аминокислотных остатков. Аминокислоты в белках связаны между собой пептидной связью. Существует несколько уровней организации белка. Первый из них – его аминокислотная последовательность (первичная структура). Вторичная структура это локально упорядоченные участки основной цепи белка. Наиболее часто встречаются два типа вторичных структур – альфа спирали и бета-нити. Вторичная структура образует жесткие сегменты полипептидной цепи, которые укладываются в третичную структуру. Стабильность упаковки пространственной структуры белка обеспечивается взаимодействиями удаленных по первичной последовательности остатков. Необходимо отметить, что белковая глобула гетерогенна и содержит внутреннюю гидрофобную часть (ядро) и полярную оболочку.

Слайд 23.

На данном слайде схематически представлены различные уровни структурной организации белка – от первичной структуры до четвертичной.

Слайд 24.

Замены нуклеотидов кодирующей части ДНК могут приводить к заменам аминокислот в белках. Каким образом такие замены могут повлиять на структуру и функцию белка? Для выяснения этого вначале рассмотрим, какие взаимодействия стабилизируют третичную структуру белка. К наиболее важным взаимодействиям относятся гидрофобные, водородные связи и взаимодействия заряженных групп – солевые мостики. Каков же вклад каждого из этих типов взаимодействий в стабильность белка?

Наиболее важными считаются гидрофобные взаимодействия. Они определяют формирование глобулярной структуры белка таким образом, что неполярные боковые группы остатков располагаются внутри белковой глобулы и становятся недоступными растворителю. Согласно экспериментам по замещению аминокислот в белках, вклад этих взаимодействий в стабильность белковой глобулы составляет около 1.2 ккал/моль на одну группу $-\text{CH}_2-$.

Важную роль при укладке белка в нативную структуру играет также формирование водородных связей между полярными группами аминокислотных остатков. Основная часть водородных связей - это взаимодействия между $-\text{NH}-$ и $-\text{CO}-$ группами основной цепи аминокислот, которые обуславливают формирование вторичной структуры белка. Вклад энергии водородных связей в стабилизацию белка оценивается в среднем как 1.6 ± 1 ккал/моль на одну водородную связь.

Существенный вклад в стабильность белковой глобулы помимо гидрофобных взаимодействий и водородных связей вносят солевые мостики. Это электростатические взаимодействия разноименно заряженных атомов боковых групп остатков. Считается, что пара остатков формирует солевой мостик, если расстояние между заряженными атомами этих остатков не превышает 4 Å. Вклад одного поверхностного солевого мостика в

стабильность белка составляет около 1 ккал/моль. В то же время, нарушение солевого мостика, расположенного внутри глобулы, вносит дестабилизирующий эффект на величину около 3-5 ккал/моль.

В среднем же выигрыш в свободной энергии при образовании нативной структуры оценивается как 1.7 ккал/моль на один остаток. Для сравнения – энергия тепловых флуктуаций при комнатной температуре составляет около 0.6 ккал/моль.

Слайд 25.

Одиночные аминокислотные замены могут вызывать различные изменения в структуре и функции белков. Все замены остатков можно классифицировать как по фенотипическому проявлению, так и по молекулярному механизму действия. В работе Колчанова (Колчанов, 1988) приведена следующая классификация одиночных замен (мутаций).

По фенотипическому проявлению мутации можно разделить на следующие классы:

- 1) нейтральные;
- 2) снижающие эффективность функционирования;
- 3) повышающие эффективность функционирования;
- 4) приводящие к возникновению новой функции.

По молекулярному механизму действия замены остатков можно разделить на три основных группы:

- 1) изменяющие конформационную стабильность белков;
- 2) влияющие на процесс укладки (самоорганизации) полипептидной цепи в нативную глобулу;
- 3) изменяющие функцию белков за счет нарушения локальной конформации функциональных центров;

Слайд 26.

Следует отметить, что эффект замены существенно зависит от структурного контекста аминокислоты. Менее всего влияют на стабильность замены на поверхности белка и замены в неструктурных участках белка (в рентгеноструктурных данных такие остатки обычно имеют высокое значение температурного фактора). Дестабилизирующие замены – такие, что увеличение свободной энергии в результате замены происходит более чем на 1 ккал/моль. Дестабилизирующие замены наблюдаются в гидрофобном ядре, затрагивают формирование солевых мостиков и водородных связей, влияют на конформацию основной цепи (замены Pro, Gly и наоборот).

Слайд 27.

Соэр и соавт. был проведен исчерпывающий анализ влияния мутаций на стабильность Arg репрессора фага P22 (Sauer *et al.*, 1996, Milla *et al.*, 1994), структура которого показана на рисунке. Проанализированы как одиночные замены, так и замены одновременно по нескольким позициям этого белка. В частности проведен эксперимент по одиночным заменам на аланин в 51-й из 53-х позиций Arg репрессора. Из них 25 мутаций отнесены авторами к классу нейтральных, поскольку они существенно не влияли на его стабильность (изменение свободной энергии денатурации белка при этих заменах по абсолютной величине составило менее 1 ккал/моль). Большинство из нейтральных мутаций затрагивали остатки, доступные растворителю и имеющие большие тепловые флуктуации боковых групп в кристаллической структуре нативного белка (высокое значение В-фактора). Около половины проведенных замен приводили к значительному понижению стабильности белка (на величину более чем 1 ккал/моль) и к существенному увеличению скорости его денатурации. Эти замены затрагивали гидрофобное ядро белка, влияли на формирование солевых мостиков и водородных связей. Часть дестабилизирующих замен были заменами глицинов в позициях с ограничениями на

конформацию углов ϕ, ψ основной цепи. Очевидно, что эти замены затрагивают важнейшие детерминанты стабильности и процесса укладки Arg репрессора.

Слайд 28.

В то же время, существуют примеры замен, повышающих стабильность белковой глобулы. Две замены, исследованные в работе Соэр и соавт. (Sauer *et al.*, 1996, Milla *et al.*, 1994; Vershon *et al.*, 1984) повышали стабильность белка на величину около 2.5 ккал/моль. Это были замены пролина в позиции 8 на аланин и лейцин. Анализ димерного комплекса репрессора мутантного белка Pro8->Leu показал, что эта замена приводит к формированию дополнительной водородной связи между β -нитями двух димеров (NH (Leu8) – CO(Tyr14)) – показано на рисунке, что по мнению авторов обуславливает увеличение стабильности белка (Sauer *et al.*, 1996). Интересен факт, что доля одиночных замен повышающих стабильность относительно невелика. Этот вывод также следует из работ Шортла и соотр., которые провели анализ 290 замен аминокислот в последовательности стафилококковой нуклеазы. Из них только 33 привели к относительно небольшому повышению стабильности. Остальные оказались дестабилизирующими в той или иной мере (Shortle *et al.*, 1990, Green *et al.*, 1992; Meeker *et al.*, 1996).

Слайд 29.

Одна из важнейших функций белка в организме – участие в биохимических реакциях в качестве катализаторов. Катализ реакций может осуществляться в несколько стадий и включать связывание с субстратом, образование интермедиатов, продуктов и переходных состояний на протяжении пути реакции. Поэтому функциональная роль остатков, участвующих в катализе, может существенно различаться. Соответственно, замены аминокислотных остатков могут оказывать влияние на разные стадии катализируемых реакций.

В настоящее время установлено, что наиболее важными для катализа являются остатки, которые формируют водородные связи с переходными комплексами субстратов, тем самым, стабилизируя их (Fersht, 1987). Было показано, что большинство замен этих остатков являются понижающими активность белка. Одним из примеров может служить исследование тирозил тРНК синтетазы - фермента, который катализирует аминокислотирование РНК^{Tyr}.

Группой Фершта исследовались различные замены в районе активного сайта этого белка. По данным рентгеноструктурного анализа и мутационных экспериментов было установлено, что в катализе участвуют 14 остатков, которые выполняют различные функции (Fersht, 1987). Остатки Tyr169, Asp78, Gln173 формируют водородные связи с аминогруппой тирозина (α -амино сайт). Остатки Asp176, Tyr34 формируют водородные связи с боковой группой тирозина (сайт специфичности к боковой группе). Остатки Cys35, His48, Thr51 взаимодействуют с рибозой в переходном комплексе. Остатки Lys82, Arg86, Thr40, His45, Lys220, Lys233 образуют сайт взаимодействия с переходным комплексом. Результаты мутационных экспериментов позволили оценить энергетический вклад этих остатков во взаимодействия с субстратами в ходе реакции (табл.3).

Слайд 30.

Тот факт, что в экспериментальных работах большей частью мутации либо мало влияют на стабильность белка, либо понижают ее, согласуются с теорией нейтральной эволюции Кимуры. Эта теория позволяет объяснить, почему в реальных популяциях частота мутантных аллелей является достаточно высокой на протяжении многих поколений. Процесс динамики мутантных генов в стохастической модели Кимуры изображен на рисунке. При этом предполагается, что основная доля мутаций, фиксировавшихся в популяции, это селективно нейтральные мутации. Для *нейтральных* мутаций существуют два основных параметра, определяющих их динамику в популяции – время фиксации

мутации t_{fix} и частота фиксации мутации K , которая равна частоте их возникновения u .
Для адаптивных замен $K=4Nsu$.

Слайд 31.

Основные положения теории молекулярной эволюции, разработанной Кимурой, заключаются в следующем.

- Адаптивные мутации очень редки;
- Деструктивные мутации быстро элиминируются
- Подавляющее большинство фиксировавшихся мутаций селективно нейтральны
- Скорость фиксации замен в генах является постоянной (гипотеза молекулярных часов, Zuckercandle & Poling)
- Функционально важные районы функционируют медленнее, менее важные - быстрее

На рисунке показана схема изменения частот мутаций разного типа до и после естественной селекции.

Слайд 32.

Рассмотрим известные примеры, которые согласуются с нейтральной теорией эволюции. Широко известен факт постоянства скоростей аминокислотных и нуклеотидных замен (этот эффект по аналогии с атомными часами в физике был назван Цукеркендлом и Полингом "молекулярными часами"). На рисунке приведен график зависимости геологического времени расхождения видов (ось Y) и числа накопленных аминокислотных замен (ось X). Видно, что эта зависимость линейна, что отражает постоянство скорости фиксации мутаций. Причем для разных белков эта скорость различна. Для белков, функциональная нагрузка которых считается невысокой (фибринопептиды), эта скорость наибольшая из приведенных. Этот факт хорошо согласуется с положениями Кимуры о том, что скорости фиксации замен постоянны и они выше для позиций (белков), которые несут меньшую функциональную нагрузку.

Слайд 33.

Еще одним интересным фактом может служить сравнение скоростей замен в кодонах по первой и второй позиции (т.н. несинонимические замены), и по третьей позиции, замены в которых слабо влияют на замену соответствующих аминокислот в последовательности белка (синонимические замены). Из приведенной таблицы видно, что скорости несинонимических замен в несколько раз меньше, чем синонимических. Для белков отвечающих за упаковку ДНК в нуклеосомах (гистоны) эта скорость практически равна 0. Т.е. их эволюция подвержена сильным ограничениям. Однако частоты несинонимических замен в них близки к частотам в таких сильно варьируемых белках, как иммуноглобулины.

Слайд 34.

Обратимся теперь к модели аминокислотных замен в белках. В этом случае алфавит описывается набором из 20 символов – канонических аминокислот. Таким образом размерность вектора частот равна 20. Простейшая модель аминокислотных замен на основе матрицы замен 20x20 была предложена Маргарет Дайхофф, в 70х гг. Эта модель и сейчас широко используется при анализе аминокислотных последовательностей. В работе Дайхофф матрица замен M была определена эмпирически на основе сравнения гомологичных белков нескольких семейств. Особенности матрицы Дайхофф таковы.

- Равновесные частоты равны частотам встречаемости аминокислот в последовательностях белков.
- Наиболее часты замены аминокислот на аминокислоты, сходные по физико-химическим свойствам.
- Исходная матрица нормирована на время, эквивалентное 1 замене на 100 позиций (1РАМ), т.е. время измеряется в единицах РАМ.
- Для оценки вероятности замен через время $t=p$ надо матрицу 1РАМ возвести в степень p .

Слайд 35.

В случае эволюции аминокислотных последовательностей рассматриваются матрицы семейства PAM. Матрицы серии PAM-N отражают вероятность замены аминокислот $a_i \rightarrow a_j$ за время эволюции, эквивалентное N PAM единиц. Зависимость между расстоянием PAM и числом замен так не является линейной, как и в случае моделей нуклеотидных замен, рассмотренных ранее. Эта зависимость показана на рисунке. Она близка к линейной на малых временах эволюции (в пределах 30% замен или ~50 PAM).

Слайд 36.

Ключевым шагом при сравнительном анализе последовательностей белков является их множественное выравнивание.

Множественное выравнивание позволяет:

- Выявлять более слабую гомологию белков.
- Обеспечивает информацию для разбиения последовательностей по паттернам дивергенции на подгруппы (подсемейства)
- Позволяет выделять вырожденные функциональные мотивы

Слайд 37.

Рассмотрим множественное выравнивание дегалогеназ и еноил-Со-А гидратаз. На рисунке видно, что последовательности двух типов белков имеют разные наборы (паттерны) аминокислот в позициях выравнивания. Причем существуют позиции, сходные для двух классов (затемнены на рисунке). Такие остатки как правило являются ключевыми для структуры белка. Существуют так же позиции, в которых наборы аминокислот являются специфическими для этих классов белков (выделены серым цветом). Такие остатки определяют специфичность фермента по отношению к лиганду.

Слайд 38.

В некоторых случаях существуют консервативные блоки позиций, замены в которых происходят на аминокислоты, сходные по физико-химическим свойствам. Такая блочная структура соответствует структурно или функционально важным участкам белков.

Вариабельные участки выравнивания как правило содержат много делеций/вставок и соответствуют петлям в третичной структуре белка.

Слайд 39.

Так, например, наиболее консервативные позиции в дегалогеназах формируют активный сайт.

Таким образом, множественное выравнивание позволяет делать определенные предположения о структурной и функциональной нагрузке выровненных белков даже при отсутствии их пространственной структуры.