

Биоинформатика. Белки и ДНК. Почему дети похожи на родителей?

С древних времён люди замечали, что у тигров рождаются тигрята, у птиц — птенцы, у людей — дети. Мало того, дети обычно внешне похожи на родителей — имеют тот же цвет глаз, цвет волос или форму носа. Новорождённый младенец часто не похож ни на мать, ни на отца, но со временем он приобретает черты внешнего сходства с матерью и отцом. Получается, что любой организм уже при рождении (а на самом деле ещё до рождения) «знает», какие у него во взрослом состоянии будут глаза, рост или голос. Значит, вся эта информация уже заложена с рождения, она где-то хранится, а по мере роста организм считывает эту информацию и приобретает те черты, которые ему предписаны наследственной программой.

Где же и как хранится наследственная информация? Как организм считывает её и понимает? Как он использует эту информацию по мере развития и роста? Этим вопросам много веков. Но ответы учёные стали находить только в последние 50 лет.

Всё в мире состоит из отдельных частей — и предметы, и живые организмы. В свою очередь, эти части сами состоят из частей и т. д. Некоторые более крупные части мы видим, а есть такие



Почему дети похожи на родителей?

маленькие, которые наш глаз не способен различить (они в миллионы раз тоньше человеческого волоса). Мы видим, например, что дом сложен из кирпичей, но мы не можем различить, из каких частичек состоит вода в стакане. И в воде, и в кирпичах, из которых построен дом, и в клетках тела человека можно выделить мельчайшие частицы, которые называются *молекулами*. Всё вокруг нас — и живое, и неживое — построено из молекул. В разных веществах молекулы различны. Они очень малы, их невозможно увидеть без специальных приборов. К счастью, такие приборы существуют — они помогают изучать молекулы.



Среди молекул любого живого организма на нашей планете основную часть составляют молекулы особых веществ — **белков**.

Различных белков, т. е. различных видов молекул белков, очень много — несколько миллионов. Например, только в организме человека встречается около 30 000 различных белков. Но при этом молекулы всех белков устроены похожим образом.



Молекула любого белка — это **цепочка (последовательность)**, состоящая из сотен, а иногда и тысяч звеньев. При этом во всех известных молекулах белков **встречается только 20 видов звеньев!** (Можно сказать, что книга жизни написана в 20-буквенном алфавите.)

Звенья белковых цепей называют *аминокислотными остатками*. Каждый аминокислотный остаток имеет своё название и обозначение (одной буквой латинского алфавита), по-



Компьютерная модель молекулы белка. Радужная расцветка позволяет проследить ход звеньев цепочки. Молекула на рисунке увеличена примерно в миллиард раз

этому белки часто описывают словами (последовательностями букв) в 20-буквенном алфавите (см. таблицу на форзаце в конце учебника).

Именно набором белков один организм отличается от другого. Наборы белков у двух людей (у двух кошек, у двух берёз) очень похожи, только мелкие различия определяют, например, разный цвет глаз у разных людей или разную расцветку шерсти двух кошек. Наборы белков у организмов разных видов разные, но чем более родственны эти виды, тем более похожи наборы белков. Например, белки человека и шимпанзе совпадают на 99 % (т. е. различается только одно звено из 100). А у человека и мыши степень сходства около 80 % (различаются 20 звеньев из 100).

Где же хранится наследственная информация?



За хранение и передачу наследственной информации в живых организмах отвечают специальные молекулы — **молекулы ДНК**.

ДНК — это сокращение, полное название — *дезоксирибонуклеиновая кислота*. Молекулы ДНК во всех клетках одного живого организма одинаковы. Но при этом молекулы ДНК разных организмов разные: у каждого человека свои молекулы ДНК, у каждой мышки свои.

Все молекулы ДНК, как и молекулы белков, — это цепочки, но звенья в молекулах ДНК отличаются от звеньев белков.



Звенья ДНК называются **нуклеотидами**. В молекулах ДНК встречается всего 4 вида нуклеотидов.

Молекулы ДНК в клетках живых организмов гораздо длиннее молекул белков. Даже самые короткие молекулы ДНК (ДНК вирусов) содержат сотни тысяч звеньев (нуклеотидов). А ДНК человека содержит около трёх миллиардов нуклеотидов. То есть молекула ДНК — это целая книга, написанная в 4-буквенном алфавите. Нуклеотиды обозначаются латинскими буквами А, С, G и Т.



Молекулы ДНК каждого живого организма **полностью определяют**, какие белки будут в этом организме.

Но как именно это происходит? Иными словами, как в ДНК закодированы (т. е. зашифрованы) белки? На этот вопрос мы ответим позже. А пока займёмся просто шифрованием.

Шифрование

С древних времён люди использовали шифрование для секретной передачи и хранения информации. Шифрование выглядит как увлекательная игра, но преследует серьёзные цели. Шифры используются в военных целях, для передачи секретных сообщений, для хранения тайного знания и во многих других случаях.

Первые зашифрованные сообщения использовались ещё в Древнем Египте. Способ шифрования тогда был очень прост, сейчас он называется «шифрование простой подстановкой»: каждый иероглиф исходного сообщения заменялся в зашифрованном сообщении другим. При этом одинаковые иероглифы заменялись одинаковыми, а разные — разными.

Сегодня существует много способов шифрования и *шифров*. Мы будем пользоваться лишь одним способом, который построен по следующим правилам:

1. Каждая русская буква, а также пробел или знак препинания заменяется последовательностью латинских букв длины 3. Такая последовательность называется **кодом**. При этом используются только четыре латинские буквы — А, С, G, T.
2. Каждый код всегда заменяет (**кодирует**) одну и ту же букву или знак. При этом одна буква или знак необязательно всегда заменяется одним и тем же кодом.
3. При замене букв и знаков кодами порядок букв и знаков не меняется.



Замена каждой буквы её кодом называется **шифрованием**. Обратная замена каждого кода на соответствующую ему букву называется **расшифровкой**.

Для удобства шифрования и расшифровки строят *шифровальную таблицу*, в которой указан код (или все коды, если их несколько) для каждой буквы, для пробела и знаков препинания. Незаполненная часть такой таблицы помещена ниже.

Буква/знак	Код
А	
Б	
В	
Г	
Д	
Е	
Ё	

 **Полный шифр** — это заполненная шифровальная таблица, указывающая соответствие каждой буквы или знака и каждого кода.

Вот примеры шифровок:

Код буквы Я — САТ.

Шифровка слова ТЫ — ССССГГ.

Зашифруем слово ОНИ — получим шифровку АСТАГСТАА.

Закодируем слово КОМПЬЮТЕР — получим шифровку АГГАСТАСГАТССТТСГАСССГАААГТ.

После раскодирования шифровки ГАААААСТСТААССГАГТ-АСТААГАААААГСАГССГГАСС получаем слово ЗАШИФРОВАННЫЙ.

Обратите внимание, что знак переноса нам помогает удобно расположить на странице слова и их шифровки. При этом знак переноса при шифровании и расшифровке не учитывается.

Полный шифр вы построите, решая задачи.

182

Вырежи из вкладыша тетради проектов заготовку шифровальной таблицы и вложи в тетрадь (лучше прикрепить её скрепкой, чтобы потом не выпала). Используя коды из примеров на этой странице выше, заполни в своей таблице все строки, которые сможешь.



Проверь: в таблице должны быть коды для 19 букв.

183

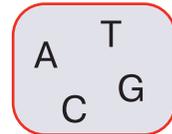
Зашифруй те слова множества F, для шифрования которых в твоей таблице имеются все необходимые коды. Запиши шифровки в тетрадь.

ШАР	БАНАН	ШАРФ	КАРТОШКА
ОНА	ОН	ВЬЮНОК	ДЕДУШКА

F**184**

Сколько различных последовательностей длины 3 можно составить из букв множества M (конечно, буквы могут повторяться)?

Построй дерево перебора вариантов. Можно ли было использовать для шифрования букв русского алфавита не тройки, а пары, составленные из букв множества M? Поясни свой ответ.

**M**

При расшифровке удобно пользоваться обратной шифровой таблицей — *таблицей расшифровки*. В такой таблице в левой колонке выписаны все возможные коды (в словарном порядке), а в правой — буквы или знаки, которые этим кодам соответствуют. В заготовке таблицы расшифровки заполни все строки, которые сможешь.

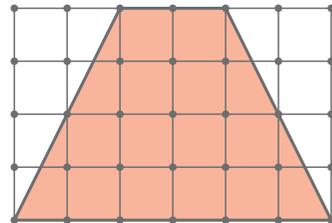
185

Раскодируй шифровки: перепиши их в тетрадь и напиши рядом с каждой шифровкой зашифрованное в ней слово.

A	G	G	A	C	T	G	A	A	A	A						
A	G	G	A	C	T	G	A	A	C	G						
A	T	C	A	G	A	A	G	T	A	A	G	A	A	C	A	T
A	A	A	G	C	C	C	A	G	T	A	A	A	G	G	C	C

186

Нарисуй, как разрезать четырёхугольник на части, чтобы из этих частей можно было собрать прямоугольник на сетке. Нарисуй этот прямоугольник и покажи штриховыми линиями, из каких частей он составлен.



187

Множество В — множество шифровок всех слов из множества А. Запиши для каждого слова его шифровку и заполни пустые клетки шифровальной таблицы и таблицы расшифровки.



Проверь себя — в каждой таблице теперь должны быть коды для всех русских букв.

А

УХ	ЭХ
ЧАС	УЖ
	БАЦ
БАС	
	ЛУГА
ДУГА	
	ЛУЖА
ОБЩИХ	
ОБЪЁМ	

В

CAGCCT	AACAACAC
CCACCT	AACAAAATG
CCACAA	CGCAAAATG
	ACACCAAATAAA
	ATTCCACAAAAA
	ATTCCAAATAAA
ACTAACGCCTAACCT	
ACTAACTCCATAACG	

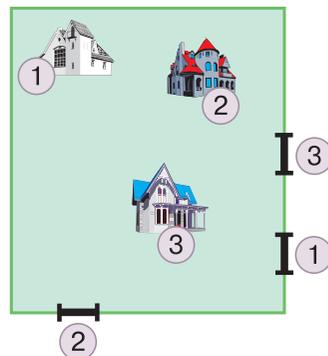
188

Дано зашифрованное предложение. Слова в этой шифровке разделены тройкой символов, кодирующей пробел. Расшифруй и запиши в тетрадь предложение. Дополни таблицы шифровки и расшифровки кодом пробела.

A C G A A A C G A A A C T A A C G C G G
A T T A A A C T A A G T A A A A C G C C A

189

На квадратном участке расположены три дома, а в ограде сделаны три калитки. Проложи дорожку от каждого дома к калитке с тем же номером так, чтобы дорожки не пересекались. Нарисуй схему участка и дорожек в тетради. Дома перерисовывать не надо — достаточно поставить номера.



190

При помощи таблицы расшифровки раскодируй следующие шифровки, запиши получившиеся слова.

АТСААААГТААААСААСТАГГАТГ
 ССГГАААААГГАТААААТТСТТ
 АСГССАГААСГГАААА
 АГСАГААСТСССТССАГААСГ-
 АТТАГААСГССГААСС

191

При помощи шифровальной таблицы зашифруй слова: ПАРОМ, ВОЗДУХ. Теперь, не обращаясь к шифровальной таблице, зашифруй слова: ПАРОВОЗ, ДУХОМ.

192

Пользуясь шифровальной таблицей, зашифруй предложение: ЛЮБЛЮ ГРОЗУ В НАЧАЛЕ МАЯ.

193

Раскодируй зашифрованное предложение.

АТСАГТТААССТАСТАСАТ-
 ААСТААТГАГАААТАСТАС-
 ААГССАТСТААГСАААСТА-
 АГГАААСССАСТААГГАСТАА-
 АГСТААТГАГАААСГСТТ

194

Найди выигрышную стратегию для игры *Двадцать пять*.

Правила игры *Двадцать пять*

Начальная позиция. Число 0.

Возможные ходы. На каждом ходу игрок прибавляет к имеющемуся числу 1, 2, 3 или 4.

Как определить победителя. Игра заканчивается, если позиция оказывается равной 25. Выигрывает тот, кто добавил последнее число.