



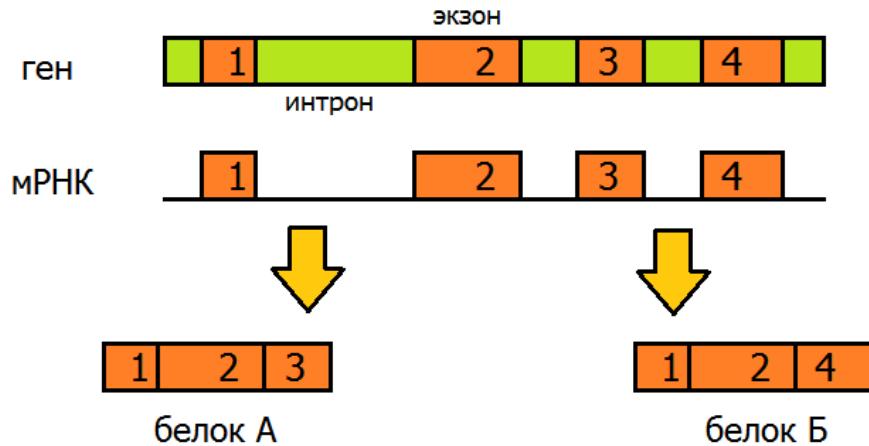
# BIOINFORMATICS SUMMER SCHOOL

## Глоссарий для информатиков

## A

**Аллель** — одна из двух (или более) альтернативных форм гена, каждая из которых характеризуется уникальной последовательностью нуклеотидов. Одна аллель для каждого гена наследуется от каждого из родителей.

**Альтернативный спlicing** — процесс, позволяющий одному гену производить несколько мРНК и, соответственно, белков благодаря разным наборам экзонов.



**Аминокислоты** — органические кислоты, содержащие минимум одну аминогруппу ( $-\text{NH}_2$ ), то есть основную, и одну карбоксильную группу ( $-\text{COOH}$ ), то есть кислотную. Среди многообразия аминокислот только 20 участвуют во внутреклеточном синтезе белков.

**Амплификация** — увеличение числа копий конкретного фрагмента ДНК; может быть в естественных условиях (*in vivo*) или в пробирке (*in vitro*).

**Аннотация генома** — предсказание и нахождение участков, кодирующих РНК и/или белки, регуляторных участков и т.д.

**Антиген** — вещество (обычно белки, реже полисахариды), вызывающее у животных иммунный ответ (образование антител).

**Антитело** — белок (иммуноглобулин), образуемый иммунной системой организма животных в ответ на введение антигена и способный вступать с ним в специфическое взаимодействие.

**Апоптоз** — запрограммированная смерть клетки. Происходит с минимальным вредом для окружающих клеток (в отличие от некроза) и часто является закономерным этапом

жизненного цикла клеток (например, происходит при исчезновении перепонок между пальцами во внутриутробном развитии человека).

**Аутосома** — любая неполовая хромосома. У человека имеется 22 пары аутосом и одна пара половых хромосом.

## Б

**База** — одна пара азотистых оснований нуклеотидов, которые в свою очередь образуют молекулы ДНК и РНК. Единица измерения длины последовательностей ДНК или РНК (например, длина генома бактерий варьируется от 139 до 13000 килобаз). В английском обозначении: **bp** (base pair - пара оснований).

**Библиотека (геномная библиотека)** — коллекция клонов, сделанных из набора случайно сгенерированных перекрывающихся фрагментов ДНК, которые представляют весь геном организма.

**Болезни, сцепленные с полом** — болезни, обусловленные дефектом генов, локализованных в Х- или Y-хромосомах.

## В

**Вектор** — молекула нуклеиновой кислоты (ДНК или РНК), служащая инструментом для введения генетической информации в клетку.

**Вирусы** — инфекционные агенты неклеточной природы, способные в процессе реализации генетической информации, закодированной в их геноме, перестроить метаболизм клетки, направив его в сторону синтеза вирусных частиц. Вирусы состоят из белковой оболочки (капсида), внутри которой находится ДНК или РНК.

**Выравнивание (alignment)** — способ сопоставления последовательностей ДНК, РНК или белков для выявления областей сходства, которые могут быть следствием функциональных, структурных, или эволюционных отношений между последовательностями. Выровненные последовательности нуклеотидов или аминокислот обычно представлены в виде строк в матрице.

AAB24882	ТУНМСQFHCRYVNNHSGEKLYECNERSKAFCSCPShLQCHKRRQIGEKTHEHNQCGKAFPT	60
AAB24881	-----YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK	40
	****: .***: * *;** * :****. ;* *****..	

AAB24882	PSHLQYHERTHTGEGKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ-	116
AAB24881	HSHLQCHKRTHTGEGKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS	98
	***** * :*****:*****:***:***. : .*****:***** : *.: :	

## Г

**Гамета** — зрелая половая клетка с гаплоидным набором хромосом. У человека их 23.

**Гаплоид** — клетка, ткань или организм с одинарным набором генов или хромосом (например, сперматозоиды и яйцеклетки).

**Гаплотип** (сокр. от «гаплоидный генотип») — совокупность аллелей на локусах одной хромосомы, обычно наследуемых вместе. Если же при кроссинговере комбинация аллелей меняется (что происходит очень редко), говорят о возникновении нового гаплотипа. Гаплотип может быть как у одного локуса, так и у целого генома. Генотип определенных генов диплоидной особи состоит из двух гаплотипов, расположенных на двух хромосомах, полученных от матери и отца соответственно.

**Ген** — последовательность нуклеотидов в ДНК, кодирующая определенную РНК, которая обуславливает какую-либо функцию в организме или регулирует транскрипцию другого гена.

**Генетическая карта** — схема расположения структурных генов и регуляторных элементов в хромосоме.

**Генетический код** — соответствие между триплетами (тремя идущими подряд нуклеотидами) в ДНК или РНК и аминокислотами белков.

**Генная терапия** — процедура, направленная на замену, манипулирование или дополнение нефункциональных или неправильно функционирующих генов здоровыми генами.

**Геном** — общая генетическая информация, содержащаяся в генах организма, или генетический состав клетки. Термин «геном» иногда употребляется для обозначения гаплоидного набора хромосом.

**Генотип** — (1) вся генетическая информация организма; (2) генетическая характеристика организма по одному или нескольким изучаемым локусам.

**Гетерозигота** — клетка (или организм), содержащая две различных аллели в конкретном локусе гомологичных хромосом.

**Гибридизация *in situ*** — использование ДНК или РНК зондов для обнаружения присутствия комплементарной последовательности ДНК в клетках.

**Гибридизация ДНК** — процесс образования двуцепочечной ДНК или дуплексов ДНК-РНК в результате взаимодействия комплементарных нуклеотидов по правилу G-C, A-T(U).

**Гомозигота** — клетка (или организм), содержащая две одинаковые аллели в конкретном локусе гомологичных хромосом.

**Гомологичные хромосомы** — хромосомы, одинаковые по набору составляющих их генов.

**Группа сцепления** — все гены, локализованные в одной хромосоме. У человека 25 групп сцепления: 22 аутосомы, две разных половых хромосомы и митохондриальная хромосома.

## Д

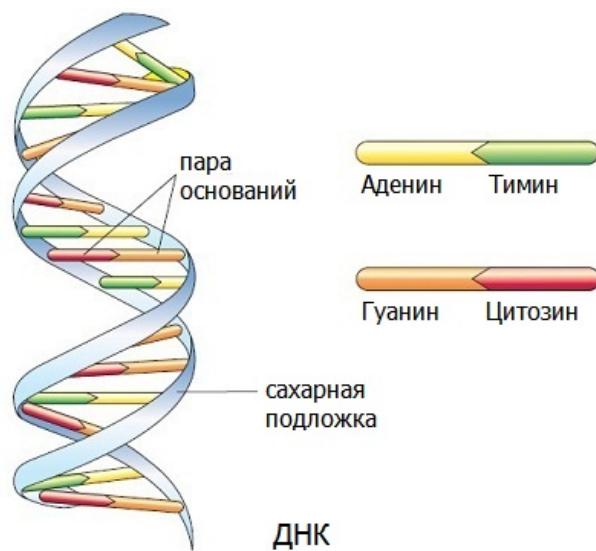
**Делеция** — тип хромосомной мутации, при которой утрачивается участок хромосомы.

**Денатурация** — нарушение пространственной структуры молекулы в результате разрыва внутри- или межмолекулярных нековалентных связей.

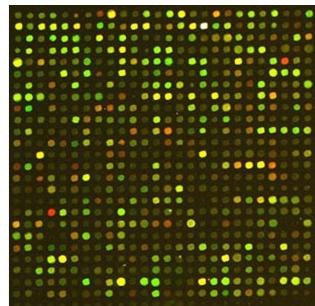
**Диплоидный набор** — полный набор генетического материала, состоящего из парных хромосом, по одному от каждого родительского набора. Большинство клеток животных, кроме половых клеток, имеют диплоидный набор хромосом. Диплоидный набор хромосом человека содержит 46 хромосом.

### ДНК (дезоксирибонуклеиновая кислота)

— молекула, кодирующая генетическую информацию. ДНК является двухцепочечной молекулой, которая удерживается слабыми связями между парами оснований нуклеотидов. Четыре нуклеотида ДНК содержат основания: **аденин (А)**, **гуанин (Г)**, **цитозин (С)**, и **тимин (Т)**. В природе пары оснований образуются только между А и Т или между Г и С.



**ДНК-микрочип** — технология, используемая в молекулярной биологии и медицине. Он состоит из тысяч дезоксирибонуклеотидов (зондов или проб), сгруппированных в виде микроскопических точек и закреплённых на твёрдой подложке. Каждая точка содержит определённую нуклеотидную последовательность, которая может быть коротким участком гена или других функциональных элементов ДНК и используется для гибридизации с цДНК или мРНК. Гибридизация зонда и мишени регистрируется и количественно характеризуется при помощи флюоресценции или хемилюминесценции, что позволяет определять относительное количество нуклеиновой кислоты с заданной последовательностью в образце.



**ДНК-полимераза** — фермент, осуществляющий матричный синтез ДНК.

**Доминирование** — тип наследования, при котором признак, кодируемый одним аллелем, проявляется, а кодируемый другим, парным — нет.

**Дрейф генов** — изменение частот генов в ряду поколений, обусловленное случайными событиями митоза, оплодотворения и размножения.

**Дупликация** — тип мутации, при которой удвоен какой-либо участок ДНК.

## 3

**Зонд генетический** — короткий отрезок ДНК или РНК известной структуры или функции, меченный каким-либо радиоактивным или флуоресцентным соединением.

## И

**Импринтинг** — эпигенетический процесс, при котором экспрессия определённых генов осуществляется в зависимости от того, от какого родителя получены аллели. Наследование признаков, определяемых импринтируемыми генами, происходит не по Менделью.

**Интерфероны** — белки, синтезируемые клетками позвоночных в ответ на вирусную инфекцию, и подавляющие их развитие.

**Инtron** — некодирующий участок гена, который транскрибируется, а затем удаляется из предшественника мРНК при сплайсинге.

## К

**Картирование генов** — определение относительного положения генов в молекуле ДНК (хромосоме или плазмиде) и расстояния в единицах сцепления или физических единицах между ними.

**кДНК** (комплементарная ДНК) — ДНК, синтезируемая по матрице РНК с помощью обратной транскриптазы.

**Кишечная палочка** (лат. *Escherichia coli*, *E. coli*) — распространенная бактерия, которая интенсивно используется в исследованиях, в частности, из-за небольшого размера генома и простоты выращивания в лаборатории.

**Клетка** — элементарная единица строения и жизнедеятельности всех организмов (кроме вирусов), обладающая собственным обменом веществ, способная к самостоятельному существованию, самовоспроизведению и развитию.

**Клон** — группа генетически идентичных клеток, возникших неполовым путем от общего предка.

**Кодон** — тройка расположенных подряд нуклеотидных остатков в ДНК или РНК, кодирующая определенную аминокислоту или являющаяся сигналом окончания трансляции (стоп-кодон).

**Комплементарность** — свойство азотистых оснований образовывать с помощью водородных связей парные комплексы аденин — тимин (или урацил) и гуанин — цитозин при взаимодействии цепей нукleinовых кислот.

**Контиг** — группа из нескольких последовательно соединенных секвенированных участков ДНК.

**Кроссинговер** (от англ. *crossing over* — пересечение) — процесс обмена участками гомологичных хромосом во время мейоза.

## Л

**Линия клеток** — генетически однородные клетки животных или растений, которые можно выращивать *in vitro* в течение неограниченно долгого времени.

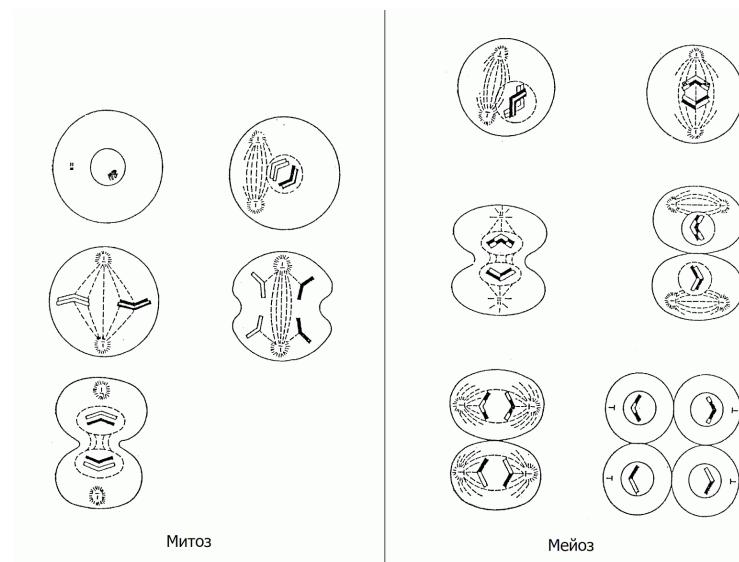
**Локус** — местоположение определённого гена или другой последовательности ДНК на генетической или цитологической карте хромосомы.

## M

**Метаболизм** — совокупность химических реакций, обеспечивающих существование и воспроизведение клетки.

**Мейоз** — деление эукариотической клетки, в результате которого из диплоидных клеток образуются гаплоидные, которые, как правило, являются гаметами.

**Митоз** — непрямое деление клетки, наиболее распространенный способ репродукции эукариотических клеток. Биологическое значение митоза состоит в строго одинаковом распределении хромосом между дочерними ядрами, что обеспечивает образование генетически идентичных дочерних клеток и сохраняет преемственность в ряду клеточных поколений.



**Мутагены** — физические, химические или биологические агенты, увеличивающие частоту возникновения мутаций.

**Мутация** — стойкое (т.е. которое может быть унаследовано потомками данной клетки или организма) преобразование генотипа, происходящее под влиянием внешней или внутренней среды.

## O

**Обратная транскриптаза** — фермент, катализирующий реакцию синтеза ДНК по РНКовой матрице.

**Олигонуклеотид** — цепь, состоящая из нескольких (обычно от 2 до 20) нуклеотидных остатков.

**Онкогены** — гены, чьи продукты обладают способностью трансформировать эукариотические клетки так, что они приобретают свойства опухолевых клеток.

**Оперон** — совокупность совместно транскрибируемых генов, обычно контролирующих родственные биохимические функции.

## П

**Покрытие нуклеотида** — сколько раз нуклеотид считывается в процессе секвенирования. Отсюда можно посчитать, например, среднее покрытие генома.

**Плазмида** — кольцевая или линейная молекула ДНК, реплицирующаяся автономно от клеточной хромосомы.

**Полимеразы** — ферменты, ведущие матричный синтез нуклеиновых кислот.

**Полимеразная цепная реакция (ПЦР)** — экспериментальный метод молекулярной биологии, способ значительного увеличения малых концентраций определённых фрагментов ДНК в биологическом материале (пробе).

**Полиморфизм** — разница в последовательности ДНК среди отдельных лиц, групп или групп населения (например, гены голубого цвета глаз против карего). Иными словами — одновременное существование в популяции разных аллелей одного гена.

**Полипептид** — полимер, состоящий из аминокислотных остатков, связанных пептидными связями.

**Праймер** — короткий фрагмент нуклеиновой кислоты, комплементарный ДНК- или РНК-мишени. Служит затравкой для синтеза комплементарной цепи с помощью ДНК-полимеразы (при репликации ДНК).

**Прокариоты** — организмы, у которых нет клеточного ядра.

## Р

**Рак** — заболевания, при которых клетки начинают бесконтрольно делиться и расти.

**Референс** (референсная сборка, референсный геном; от англ. *reference* — начало отсчёта, эталон, образец) — готовая геномная сборка, представляющая, по возможности, наиболее точный пример генома определенного вида.

**Рекомбинантный белок** — белок, полученный в результате экспрессии с рекомбинантной молекулой ДНК, часто получаемый в кишечной палочке.

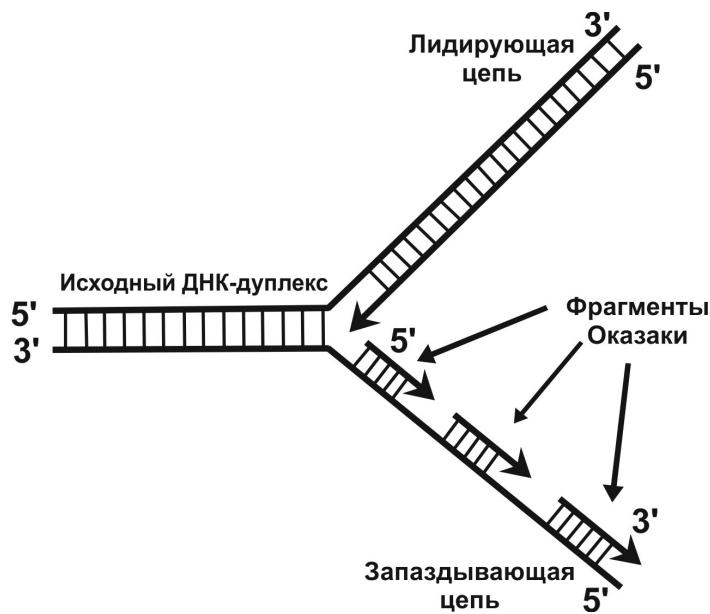
**Рекомбинантная молекула ДНК** — молекула ДНК, в которой содержатся различные гены, соединенные методами генетической инженерии.

**Рекомбинация гомологическая** — обмен генетическим материалом между двумя гомологичными молекулами ДНК.

**Рекомбинация сайт-специфическая** — объединение путем разрыва и слияния двух молекул ДНК или участков одной молекулы, происходящее по определенным сайтам.

**Репарация ДНК** — исправление повреждений молекулы ДНК, восстанавливающее её первоначальную структуру.

**Репликация** — процесс удвоения молекул ДНК или геномных вирусных РНК.



**Ретровирусы** — семейство РНК-содержащих вирусов, заражающих преимущественно позвоночных. Наиболее известный и активно изучаемый представитель — вирус иммунодефицита человека (ВИЧ).

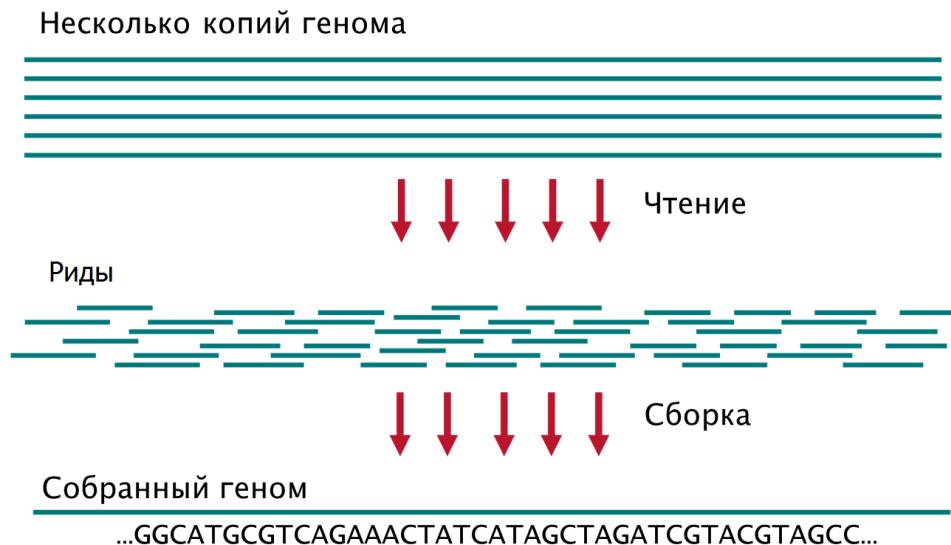
**Рецессивное наследование** — тип наследования признака или болезни, при котором мутантный аллель должен быть унаследован от обоих родителей для проявления признака.

**Риды** (от англ. **read**) — результат работы секвенатора: строки, содержащие прочитанные последовательности нуклеотидов.

## C

**Сайт** — участок молекулы ДНК, белка и т. п.

**Сборка (assembly)** — процесс объединения большого количества коротких фрагментов ДНК или РНК (ридов), полученных в результате секвенирования, в одну или несколько длинных последовательностей (контигов и скаффолдов) с целью восстановления исходных последовательностей.



**Секвенирование** — определение аминокислотной (белки) или нуклеотидной (ДНК / РНК) последовательности.

**Скаффолд** (англ. scaffold) — серия контигов, которые находятся в правильном порядке по отношению к геному, но не обязательно связаны в одну непрерывную последовательность.

**«Снип»** (англ. Single nucleotide polymorphism, **SNP**) — отличия последовательности ДНК размером в один нуклеотид (А, Т, Г или С) в геноме (или в другой сравниваемой последовательности) представителей одного вида или между гомологичными участками гомологичных хромосом.

**Соматические клетки** — клетки тканей многоклеточных организмов, не относящиеся к половым.

**Сплайсинг** — процесс формирования зрелой мРНК или функционального белка путем удаления внутренних частей молекул — инtronов РНК.

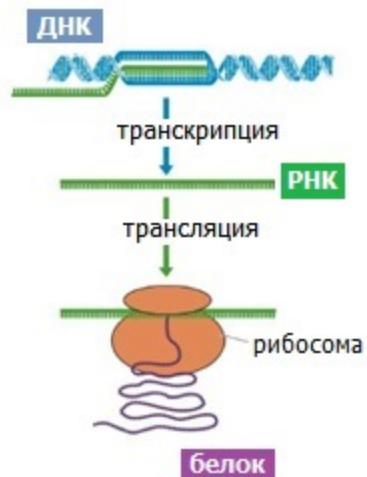
**Стволовые клетки** — недифференцированные (незрелые) клетки, имеющиеся во всех многоклеточных организмах. Стволовые клетки способны самообновляться, образуя новые стволовые клетки, делиться посредством митоза и дифференцироваться в специализированные клетки, т.е. превращаться в клетки различных органов и тканей.

## Т

**Транскрипция** — синтез РНК на ДНК-матрице; осуществляется РНК-полимеразой.

**Транскрипт** — продукт транскрипции, т. е. РНК, синтезированная на данном участке ДНК как на матрице и комплементарная одной из его нитей.

**Транскриптаза обратная** — фермент, синтезирующий по РНК как по матрице комплементарную ей однонитевую ДНК.



**Транскриптом** — совокупность всех транскриптов, синтезируемых одной клеткой или группой клеток, включая мРНК и некодирующие РНК. В отличие от генома, который, как правило, одинаков для всех клеток одной линии, транскриптом может сильно меняться в зависимости от ткани, этапа развития и условий окружающей среды.

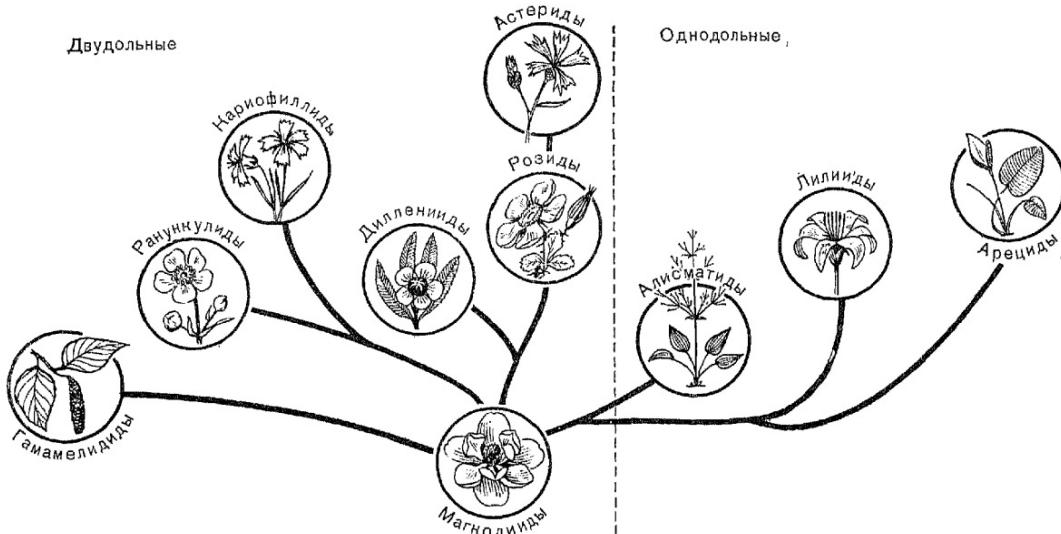
**Трансляция** — процесс синтеза полипептида, определяемый матричной РНК.

**Транспозон** — генетический элемент, реплицируемый в составе репликона и способный к самостоятельным перемещениям (транспозиции) и интеграции в разные участки хромосомной или внекромосомной ДНК.

## Ф

**Фенотип** — внешнее проявление свойств организма, зависящих от его генотипа и факторов окружающей среды.

**Филогенетическое дерево (эволюционное дерево)** — дерево, отражающее эволюционные взаимосвязи между различными видами или другими сущностями, имеющими общего предка.



**Финишированный геном** — полная последовательность ДНК всех хромосом без разрывов, с высоким качеством и низким процентом ошибок.

## Х

**Хромосомы** — структуры клетки, в которых сосредоточена большая часть наследственной информации и которые предназначены для её хранения, реализации и передачи. Иными словами, форма упаковки наследственной информации. Основной составляющей каждой хромосомы является ДНК. У прокариот хромосомная ДНК является кольцевой, и весь геном находится обычно на одной хромосоме. Геномы эукариот состоят из нескольких хромосом, каждая из которых ассоциируется с различными видами белков.

## Ш

**Штамм** — клonalльная по происхождению культура микроорганизмов, изолированная в определенное время в определенном месте и сходная по морфологическим, биохимическим, генетическим и другим признакам.

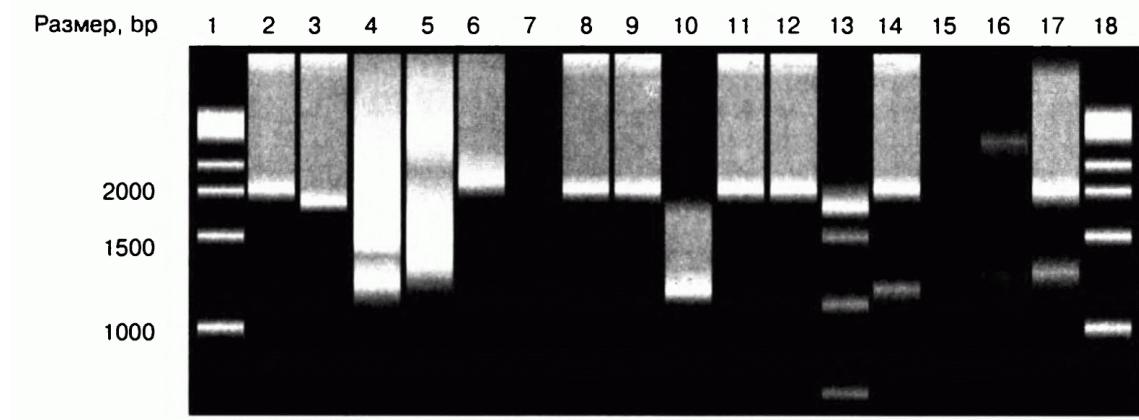
## Э

**Экзон** — сохраняющаяся при сплайсинге часть интронированного гена и несущая генетическую информацию.

**Эпигенетика** (греч. επί — над, выше) — изучение закономерностей эпигенетического наследования — изменения экспрессии генов или фенотипа клетки, вызванных механизмами, не затрагивающими последовательности ДНК.

**Экспрессия гена** — процесс реализации информации, закодированной в гене. Состоит из двух основных стадий — транскрипции и трансляции.

**Электрофорез** — метод для разделения больших молекул (например, белки или фрагменты ДНК) из смеси подобных молекул. Электрический ток пропускают через среду, содержащую смесь, и каждый вид молекул проходит через среду с различной скоростью в зависимости от его электрического заряда и размера.



**Эукариоты** — организмы, клетки которых содержат ядра. Эукариоты включают в себя все организмы, кроме вирусов, бактерий и архей.

## B

**BLAST (Basic Local Alignment Search Tool)** — алгоритм (и одноименный инструмент), используемый для поиска последовательностей в базах данных, оптимальных для локального выравнивания с запрашиваемой строкой. Часто применяется для поиска гомологичных (похожих) генов.

## F

**FASTA format** — текстовый файл для представления нуклеотидных или пептидных последовательностей, в котором нуклеотиды или аминокислоты представлены в виде однобуквенных кодов. Последовательность в формате FASTA начинается с описания в одну строку, а затем строк последовательности. Описание начинается с символа «>» («больше»).

```
>Sequence_1
MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAACKAD
LVSVKVSDDFTIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEERR
>Sequence_2
SATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTINGVKFEELYK
ATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVASKSRDLLR
```

## G

**GC-состав** — процентный состав суммы всех гуанинов (G) и цитозинов (C) по отношению к длине исследуемого участка нуклеиновых кислот.

## I

**In vitro** (лат. «в стекле») — технология выполнения экспериментов, при которой опыты проводятся «в пробирке» — вне живого организма.

**In vivo** (лат. «в (на) живом») — проведение экспериментов на (или внутри) живой ткани при живом организме.

**In silico** (искаж. лат. «в кремнии») — компьютерное моделирование экспериментов.