

Геномное редактирование

2022/23 учебный год

Второй отборочный этап

Наверное, вы знаете, что в октябре 2020 года **Нобелевскую премию по химии** за «разработку метода редактирования генома» присудили Эмманюэль Шарпантье и Дженнифер Даудне.



© Nobel Media. Ill. Niklas Elmehed.

Emmanuelle Charpentier

Prize share: 1/2



© Nobel Media. Ill. Niklas Elmehed.

Jennifer A. Doudna

Prize share: 1/2

Пандемия коронавирусной инфекции стала трудным испытанием для системы здравоохранения. Компании рынка «Хелснет» НТИ внесли значительный вклад в решение задач по ПЦР- и ИФА-диагностике коронавирусной инфекции, разработке новых вакцин. Большую роль в этом сыграли не только специалисты, работающие в молекулярно-биологических лабораториях, но и биоинформатики.

Сквозная технология Национальной технологической инициативы (НТИ) «**Управление свойствами биологических объектов**» основана на достижениях инженерной биологии: генетической инженерии, биотехнологии, молекулярной биологии, биохимии и многих других. Актуальность исследований в области разработки новых инструментов генетической инженерии и геномного редактирования сегодня не вызывает сомнений. Большая часть новых продуктов фармацевтической индустрии, биотехнологий и других отраслей основана на использовании инструментов, разработанных в конце XX века (**генетическая инженерия**), сегодня активно разрабатывают и другие подходы, среди которых особое место занимает **синтетическая биология** и технология **геномного редактирования**.

Задания второго отборочного этапа разработаны для того, чтобы лучше подготовить команды участников к решению задачи финала. Для понимания процессов, происходящих *in vitro* (в пробирке в лаборатории) необходимы знания о физико-химических основах происходящих реакций, особенно в случае участия в них различных ферментов. Для **расчета компонентов реакционной смеси** требуется знать и освоить на практике достаточно простые вычислительные действия.

Сегодня практически любое исследование либо начинается с математического моделирования, либо заканчивается **биоинформатическим анализом** полученных экспериментальных данных. Для такой работы не всегда требуются мощные суперкомпьютеры или вычислительные кластеры, многие действия можно выполнить с использованием онлайн-сервисов и баз данных, а также при помощи доступного свободного программного обеспечения. Ну и, конечно, для понимания сути происходящего *in vivo* и *in vitro*, необходимы знания об основных **молекулярно-генетических процессах**. Задания отборочного трека также посвящены изучению основ языка **Python** и его использования в биоинформатике.

Участникам второго отборочного этапа Олимпиады будут предложены задачи по генетической инженерии и геномному редактированию. Для их решения вам потребуются компетенции *исследователя*, имеющего опыт работы (и выполнения расчетов) в молекулярно-биологической лаборатории, а также *биоинформатика*, владеющего навыками работы в программном пакете UGENE, использованию языка программирования Python и работе в онлайн-базах данных. Замечательно, если каждый участник обладает всеми компетенциями, но мы советуем командам выделить двух «мокрых» биологов, которые на финале планируют работать *в лаборатории* и одного биоинформатика. Для того, чтобы вам было легче сделать выбор, задания второго тура сочетают в разной степени эти две группы компетенций.

Задания для 8–9 классов

Первый блок задач

Задача IV.1.1.1. Приготовление раствора заданной концентрации (1 балл)

Темы: расчеты.

Реакционная смесь для полимеразной цепной реакции (ПЦР) содержит ионы Mg^{2+} в определенной концентрации.

Условие

Рассчитайте массу 2 М раствора $MgCl_2$ (плотность $1,1742 \text{ г/см}^3$), который нужно взять для приготовления 150 мл 2,5 мМ раствора, который будет использован для приготовления реакционной смеси для ПЦР. Ответ приведите в мг, в виде числа, без обозначения «мг», округлите с точностью до целого значения.

Ответ: 220.

Задача IV.1.1.2. Полимеразная цепная реакция (1 балл)

Темы: ПЦР, расчеты.

Полимеразная цепная реакция (ПЦР) — метод молекулярной биологии, позволяющий специфически наработать копии фрагмента ДНК. При выполнении расчетов предполагают, что в каждом цикле ПЦР происходит удвоение специфического фрагмента ДНК.

Условие

Вычислите, на каком цикле ПЦР образуется больше 3 миллиардов копий фрагмента ДНК, если известно, что в реакционную смесь добавили 10 копий матрицы ДНК. Для расчета используйте формулу:

$$N = m \times 2^n.$$

Ответ введите в виде числа, округлите до целого значения.

Список литературы

- Лекция «Полимеразная цепная реакция» (Сергей Седых, ИХБФМ СО РАН, НГУ) — <https://youtu.be/0GLMcUz0044>.
- Лекция «Полимеразная цепная реакция» (Артем Тикунов, ИХБФМ СО РАН, НГУ) — <https://youtu.be/V2qm9jTNRKI>.

Ответ: 28.

Задача IV.1.1.3. Эффективность трансформации бактериальных клеток (2 балла)

Темы: расчеты, трансформация.

В процессе трансформации бактериальные клетки поглощают молекулы ДНК из внешней среды. Для этого используют химически компетентные и электрокомпетентные клетки. Такие клетки после теплового шока или в результате воздействия электрического тока поглощают суперскрученную плазмидную ДНК. Метод трансформации бактериальных клеток используют для доставки конструкций, содержащих ген Cas9 при редактировании бактериальных геномов.

Условие

Произведите подсчет эффективности трансформации, для которой использовали 1 нг плазмиды pCas9 в электрокомпетентные клетки *E. coli* (штамм NEBα). Известно, что после трансформации к клеткам добавили среду SOC до 1 мл, выселили на чашку с селективной средой аликвоту 20 мкл, в результате выросло 70 колоний.

Ответ запишите в КОЕ/мкг плазмиды в виде натурального числа.

Список литературы

- Википедия: Трансформация бактерий — [https://ru.wikipedia.org/wiki/Трансформация_\(генетика\)](https://ru.wikipedia.org/wiki/Трансформация_(генетика)).
- Биомолекула: 12 методов в картинках — <https://biomolecula.ru/articles/12-metodov-v-kartinkakh-gennaia-inzheneriia-chast-ii-instrumenty-i-tekhniki>.

Ответ: 3500000 ± 100000 .

Задача IV.1.1.4. Эффективность трансформации бактериальных клеток (3 балла)

Темы: расчеты, трансформация.

Используя результат, полученный в прошлой задаче, сравните эффективность электропорации и трансформации химически компетентных клеток.

Условие

Во сколько раз электропорация отличается от химической трансформации, если известно, что эффективность последней — $2,5 \cdot 10^5$ КОЕ/мкг плазмиды. (Вычислите отношение эффективности электропорации к трансформации химически компетентных клеток). Ответ запишите в виде натурального числа.

Ответ: 14 ± 1 .

Задача IV.1.1.5. Принцип комплементарности (1 балл)

Темы: молекулярно-генетические процессы.

Последовательность лидирующей цепи ДНК при репликации 3'-GTC CGT GTA CTG-5'. Известно, что ДНК-полимераза движется по матричной цепи в направлении 3' → 5' и синтезирует дочернюю цепь в направлении 5' → 3'.

Условие

Запишите последовательность в направлении 3' → 5' (!), которая будет синтезирована ДНК-полимеразой на приведенном участке согласно принципу комплементарности. В ответе введите только нуклеотиды (ATGC, латинскими буквами).

Список литературы

- Википедия: Комплементарность — [https://ru.wikipedia.org/wiki/Комплементарность_\(биология\)](https://ru.wikipedia.org/wiki/Комплементарность_(биология)).
- Википедия: Репликация — https://ru.wikipedia.org/wiki/Репликация_ДНК.
- Репликация ДНК в Khan Academy — <https://ru.khanacademy.org/science/biology/dna-as-the-genetic-material/dna-replication/a/molecular-mechanism-of-dna-replication>.

Ответ: CAGTACACGGAC.

Задача IV.1.1.6. Свойства генетического кода (1 балл)

Темы: молекулярно-генетические процессы.

Генетический код — правило, согласно которому информация, записанная в ДНК и мРНК преобразуется в последовательность аминокислотных остатков белка. Генетический код обладает некоторыми свойствами.

Условие

Выберите свойства, которыми обладает генетический код.

1. Универсальность.
2. Вырожденность.
3. Двухцепочечность.
4. Триплетность.
5. Специфичность.
6. Вариативность.

Список литературы

- Википедия: Генетический код — https://ru.wikipedia.org/wiki/Генетический_код.
- Учебник «25 лекций по молекулярной биологии» (Г. М. Дымшиц, НГУ) — <https://e-lib.nsu.ru/reader/bookView.html?params=UmVzb3VyY2UtNTk2Mw/cGFnZTAwMDA>.

Ответ: 1, 2, 4, 5.

Задача IV.1.1.7. База данных NCBI (1 балл)

Темы: NCBI.

С помощью базы данных NCBI можно определить, какому виду принадлежит последовательность ДНК.

Условие

С помощью сервиса NCBI определить, какому виду принадлежит последовательность ДНК. Последовательность ДНК можно скачать по ссылке: (<https://disk.yandex.ru/d/mkOKchoBFdXe8w>).

1. *Vulpes vulpes*.
2. *Vulpes zerda*.
3. *Vulpes ferrilata*.
4. *Vulpes corsac*.
5. *Canis familiaris*

Список литературы

- Биомолекула: 12 методов в картинках — <https://biomolecula.ru/articles/metody-v-kartinkakh-sekvenirovanie-nukleinovykh-kislot>.
- Инструмент Nucleotide BLAST — <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.

Ответ: 2.

Задача IV.1.1.8. Протоспейсер и PAM (5 баллов)

На первом этапе работы системы геномного редактирования CRISPR/Cas9 происходит распознавание мотива, прилегающего к протоспейсеру (PAM).

Белок Cas9 из *Streptococcus pyogenes* (SpyCas9) широко используют для редактирования генома, отчасти потому, что он распознает короткую последовательность PAM — 5'-NGG-3', где N — любой нуклеотид. «Дальние родственники» Cas9 демонстрируют значительное разнообразие в PAM последовательностях, например, Cas9 из *Staphylococcus aureus* (SauCas9) распознает последовательность 5'-NNGRRT-3' (где R — A или G), а Cas9 из *Neisseria meningitidis* (Nme2Cas9) распознает 5'-NNNNCC-3'. Ниже приведена последовательность одной из цепей фрагмента ДНК.

5'-AACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTTCGCGCCGAAGAA
CGTTTTTCG-3'.

Условие

Сколько PAM сайтов для Nme2Cas9 и SauCas9 можно найти в фрагменте двуцепочечной ДНК?

Список литературы

- Википедия: CRISPR — <https://ru.wikipedia.org/wiki/CRISPR>.
- Нариман Баттулин (ИЦИГ СО РАН, НГУ) «От геномного редактирования к синтетической биологии» — <https://youtu.be/SD1qAo5YZ94>.
- Чл-корр РАН Дмитрий Жарков (ИХБФМ СО РАН, НГУ) «Геномное редактирование» — <https://youtu.be/NSGvAFnPIkE>.
- Константин Северинов «Редактирование генома» (Сколтех) — https://vk.com/video-46417318_456258673.

Ответ: данная последовательность содержит 4 сайта Nme2Cas9 и 3 сайта SauCas9.

Задача IV.1.1.9. Генетически модифицированные животные (15 баллов)

Умение анализировать научную информацию, работать с литературными источниками, письменно излагать свои мысли — является крайне востребованным навыком для молодого исследователя. Опыт проведения второго отборочного тура НТО в 2019–2022 году показал, что написание короткого текста с развернутым ответом на вопрос оказалось одним из самых трудных и интересных заданий.

Для выполнения творческого задания можно и нужно использовать литературные источники (в том числе, тексты в Интернете), но не стоит копировать большие фрагменты текста. Размер требуемого текста ограничен, помните, что краткость — сестра таланта.

P.S. Эксперты будут читать ваши работы с интересом и удовольствием!

Условие

Напишите текст (эссе) объемом **не более 3000 символов (с учетом пробелов)** на тему «Перспективы практического использования генетически модифицированных животных».

Постарайтесь в ответе привести аргументированную точку зрения. Авторы лучших текстов получают специальные призы от спонсоров и партнеров профиля «Геномное редактирование».

Перед работой над эссе рекомендуем посмотреть лекцию к. б. н. Наримана Баттулина (ИЦИГ СО РАН, НГУ) «Генетически модифицированные животные» — <https://youtu.be/SwQJ8sqN3wc>.

Критерии оценивания

- Длина эссе — не более 3000 символов (с учетом пробелов).
- Правильная аргументация, обоснование ответа.
- Отсутствие биологических ошибок.
- Ссылки на литературные источники.
- Текст написан лично автором, отсутствуют заимствованные фрагменты текста.
- В зачет идет одно лучшее эссе от команды.
- При наличии нескольких одинаковых/стереотипных текстов, будет оцениваться текст, загруженный первым, остальные будут оценены в 0 баллов.
- При совпадении фрагментов текстов у разных команд, будет оценен первый загруженный текст, остальные будут оценены в 0 баллов.
- При превышении рекомендованной длины текста, оцениваться будут первые 3000 символов (с учетом пробелов).
- Список литературы включается (!) состав в 3000 символов.

Второй блок задач

Задача IV.1.2.1. Поиск информации о гене (3 балла)

Условие

TP53 — один из самых известных онкосупрессорных генов человека. Этот ген кодирует транскрипционный фактор, регулирующий клеточный цикл. Мутации в этом гене встречаются во многих злокачественных опухолях человека.

Ген TP53 человека имеет идентификатор NCBI Gene ID (35162/12415/7157/643/14241)¹ и располагается на хромосоме номер _____². Белок, кодируемый этим геном, имеет идентификатор (O15169/P04637/Q9BV47/Q13526/P05067)³ в базе данных UniProt. Изоформа этого белка с длиной 393 аминокислотных остатка имеет молекулярную массу _____⁴ кДа (введите натуральное число).

Ответ: 1 — 7157, 2 — 17, 3 — P04637, 4 — 43.

Задача IV.1.2.2. Поиск информации о мутации (3 балла)

Так как ген TP53 вносит значительный вклад в развитие онкологического заболевания, существует множество тест-систем, направленных на детекцию мутаций в этом гене. Найденные мутации могут служить прогностическими маркерами (предсказывать тяжесть заболевания) или являться показаниями для назначения таргетной терапии. Допустим, при помощи тест-системы установлено, что произошла следующая мутация в гене TP53: замена цитозина на тимин в 7670621 позиции на 17 хромосоме (chr17:7670621 C/T). Идентификатор этой мутации в dbSNP: rs876660285.

Условие

Изучите информацию о данной мутации и выберите верные суждения.

Эта мутация является (синонимичной/миссенс/нонсенс)¹ и приводит к (сохранению аминокислоты/замене глицина на валин/замене аспартата на серин/замене аргинина на лизин/появлению стоп-кодона)². Вероятнее всего, эта мутация (вносит существенный вклад в развитие заболевания/не является патогенной)³.

Ответ: 1 — миссенс, 2 — замене аргинина на лизин, 3 — не является патогенной.

Задача IV.1.2.3. Приготовление растворов (3 балла)

В лабораторной практике часто приходится готовить растворы веществ с заданной концентрацией. Чтобы автоматизировать процесс расчетов для приготовления растворов, можно использовать программирование на языке Python.

Условие

Напишите программу, которая будет вычислять, сколько сухого вещества нужно добавить в раствор для достижения нужной молярной концентрации. При помощи функции `input()` считайте строку, которая будет содержать следующие значения, разделенные одним знаком пробела: молярная масса вещества (г/моль), концентрация (моль/л) и объем раствора (л), который нужно приготовить. Выполните необходимые вычисления и выведите ответ на экран при помощи функции `print()` количество грамм вещества, которое необходимо добавить в раствор.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--------------------------|
| Стандартный ввод |
| 40 1 1 |
| Стандартный вывод |
| 40.0 |

Пример №2

| |
|--------------------------|
| Стандартный ввод |
| 121 0.05 0.5 |
| Стандартный вывод |
| 3.025 |

Пример №3

| |
|--------------------------|
| Стандартный ввод |
| 133 0.3 0.33 |
| Стандартный вывод |
| 13.167 |

Пример №4

| |
|--------------------------|
| Стандартный ввод |
| 22 0.77 0.07 |
| Стандартный вывод |
| 1.1858 |

Пример №5

| |
|--------------------------|
| Стандартный ввод |
| 12564 0.0005 0.31 |
| Стандартный вывод |
| 1.94742 |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 a, b, c = map(float, input().split())
2 answer = round(a*b*c, 15)
3 print(answer)
```

Задача IV.1.2.4. Поиск сайтов рестрикции (3 балла)

При помощи языка программирования Python можно работать с биологическими последовательностями, например искать сайты узнавания рестриктаз в последовательности ДНК.

Условие

Напишите программу, которая будет принимать на вход (через `input()`) две последовательности, разделенные одним знаком пробела: сайт узнавания рестриктазы и последовательность ДНК (всегда в направлении 5'-3'). Выполните проверку на наличие этого сайта узнавания в последовательности и выведите на экран (при помощи `print()`) `True`, если сайт рестрикции есть в последовательности и `False`, если сайта рестрикции нет в последовательности.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|----------------------------------|
| Стандартный ввод |
| AAGCTT CACACATAAGCTTATATCACATACA |
| Стандартный вывод |
| True |

Пример №2

| |
|--|
| Стандартный ввод |
| GGATCC CAACACATGATGATCATCATCATGTGATCTA |
| Стандартный вывод |
| False |

Пример №3

| |
|-------------------------------------|
| Стандартный ввод |
| CCCC ATATATACACACTGAGCAGGAGCATATACA |
| Стандартный вывод |
| False |

Пример №4

| |
|------------------------------------|
| Стандартный ввод |
| AAAAA AATATGAGAGAAAAAAAAAATATACACA |
| Стандартный вывод |
| True |

Пример №5

| |
|--------------------------------|
| Стандартный ввод |
| ATATATA ATATATATATATATATATATAT |
| Стандартный вывод |
| True |

Пример №6

| |
|---|
| Стандартный ввод |
| ATGATGCAT ATCACTAGCTACGATCGATCGATCGACTAGCTACGATCGATCGATCGAT |
| Стандартный вывод |
| False |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 rest, seq = input().split()
2 if seq.find(rest) != -1:
3     print(True)
4 else:
5     print(False)
```

Задача IV.1.2.5. Обратнo-комплементарная цепь ДНК (3 балла)

Обычно работают с последовательностями ДНК, записанными в направлении 5'-3'. Если для такой последовательности найти комплементарную последовательность, то она будет записана в направлении 3'-5', и для работы с ней в направлении 5'-3', эту последовательность необходимо «перевернуть». Такая последовательность называется обратнo-комплементарной последовательностью ДНК.

Условие

Напишите программу, которая будет принимать на вход (через `input()`) последовательность ДНК (в направлении 5'-3'). Найдите для этой последовательности обратнo комплементарную последовательность и выведите ее на экран (в направлении 5'-3') при помощи функции `print()`. Программа должна работать с последовательностями в любом регистре, и выводить ее на экран в том регистре, в котором она была подана на вход.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--------------------------|
| Стандартный ввод |
| ATGTTACATGACG |
| Стандартный вывод |
| CGTCA TGTAACAT |

-
2. При помощи метода геномного редактирования можно «выключать» нежелательные гены.
 3. После проведения генно-инженерных манипуляций требуется провести отбор клонов, которые содержат требуемую вставку.
 4. В эксперименте по геномному редактированию обычно все полученные клетки содержат желаемые изменения, так как работа системы CRISPR/Cas9 намного точнее, чем рестрикция и лигирование.
 5. Для переноса трансгенов из одного организма в другой обычно используют бактериальные плазмиды.
 6. Инструменты геномного редактирования, например, белок Cas9 может быть доставлен в клетки в результате трансформации/трансфекции плазмидной ДНК.
 7. Для доставки трансгенов и генетических конструкций в клетки животных обычно используют «генные пушки».
 8. Большая часть препаратов, уже разрешенных к использованию для генной терапии человека основаны на технологии CRISPR/Cas.
 9. Технологии управления свойствами биологических объектов, основанные на методах генной инженерии, обычно используют при каждом цикле экстракорпорального оплодотворения для отбора эмбрионов с заданными свойствами.
 10. Открытие эндонуклеаз рестрикции позволило разработать технологии получения рекомбинантных ДНК.
 11. Бактерии используют ферменты-рестриктазы для защиты от бактериофагов.
 12. Бактерии используют последовательности CRISPR и белки Cas для защиты от бактериофагов.
 13. Стоимость нескольких единиц активности рестриктазы, необходимой для проведения одного эксперимента, обычно не превышает нескольких сотен рублей (нескольких долларов США).
 14. До распространения ПЦР фрагменты ДНК, необходимые для генно-инженерных манипуляций, обычно нарабатывали в бактериальных плазмидах.
 15. Для отбора трансгенных организмов, содержащих целевой фрагмент ДНК, можно использовать метод ПЦР.

Ответ: 2, 3, 5, 6, 10, 11, 12, 13, 14.

Геномное редактирование

Ген TP53, которому были посвящены задания по биоинформатике, является онкосупрессором. То есть инактивирующие (приводящие к потере функции) мутации в этом гене являются причиной возникновения рака. Но существуют другая группа генов — онкогены. Это гены, активирующие мутации в которых способствуют появлению злокачественных опухолей. Нокаут, или «выключение» таких генов, может являться перспективным методом лечения рака. Также, нокаут потенциальных онкогенов при помощи системы CRISPR/Cas9 может использоваться для изучения их функции и влияния на развитие опухоли.

Один из таких потенциальных новых онкогенов — это ген EGFL6. Существует гипотеза, что этот ген стимулирует рост сосудов в опухоли, что является необходимым условием для ее выживания и роста. Компетенции: 1. Нокаут этого гена на

клеточных линиях рака яичника (SKOV3) позволит проверить гипотезу об участии этого гена в ангиогенезе в опухоли.

В следующих заданиях необходимо изучить систему для нокаута этого гена при помощи РНК-направленной эндонуклеазы Cas9 и ответить на вопросы.

Список литературы

1. Zhu W, Liu C, Lu T, Zhang Y, Zhang S, Chen Q and Deng N (2020) Knockout of EGFL6 by CRISPR/Cas9 Mediated Inhibition of Tumor Angiogenesis in Ovarian Cancer. *Front. Oncol.* 10:1451. doi: 10.3389/fonc.2020.01451 Ссылка: <https://www.frontiersin.org/articles/10.3389/fonc.2020.01451/full>.

Задача IV.1.2.7. Поиск PAM (2 балла)

В качестве мишени для редактирования был выбран следующий участок этого гена: 5'-TGCTGTTTCTTCTCCAGCCGTAGCAGCAGGCCAGTT-3'.

Направляющая РНК (sgRNA) имеет последовательность: 5'-TCTTCTCCAGCCGTAGCAGC-3'. В качестве фермента для редактирования выбран белок Cas9 из *Streptococcus pyogenes* (SpyCas9).

Условие

Данный участок имеет последовательность PAM 5'-_____¹-3'. Этот участок будет узнаваться (ферментом *SpyCas9*/за счет комплиментарного связывания с нРНК)², поэтому последовательность PAM (должна/не должна)³ включаться в состав олигонуклеотида, содержащего нРНК, при его синтезе в лаборатории.

Ответ: 1 — AGG, 2 — ферментом *SpyCas9*, 3 — не должна.

Задача IV.1.2.8. Поиск нокаутящей мутации (2 балла)

Эндонуклеаза Cas9 вносит двуцепочечный разрыв в последовательность ДНК, который является субстратом системы репарации в клетке. При этом могут возникать (а могут и не возникать) различные мутации. И не все из этих мутаций будут гарантированно приводить к нокауту (или «выключению») гена.

Разработанная система для нокаута проверили на шести клеточных линиях, и секвенировали целевую последовательность ДНК после редактирования.

Условие

Выберите из предложенных вариантов те, последовательности которых содержат нокаутящую мутацию.

1. TGCTGTTTCTTCTCCAGCACGTAGCAGCAGGCCAGTT.
2. TGCTGTTTCTTCTCCAGCCGTAGCAGCAGGCCAGTT.
3. TGCTGTTTCTTCTCCAGCCGCAGCAGCAGGCCAGTT.
4. TGCTGTTTCTTCTCCAGCGTAGCAGCAGGCCAGTT.

5. TGCTGTTTCTTCTCCAGACCCCGTAGCAGCAGGCCAGTT.
6. TGCTGTTTCTTCTCCAGCCGTAGCAGGCCAGTT.

Ответ: 1, 4.

Задача IV.1.2.9. Изучение вектора для клонирования (2 балла)

Для доставки системы геномного редактирования в клетки будет использован вектор LentiCRISPRv2. Его последовательность в формате GenBank (.gb) можно скачать по ссылке: <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/244694/51ae3e01-5e94-49b8-a4fb-7fa303a7eb81/addgene-plasmid-52961-sequence-244694.gbk>.

Скачайте вектор LentiCRISPRv2, откройте его в программе UGENE (или аналогичной) и изучите его.

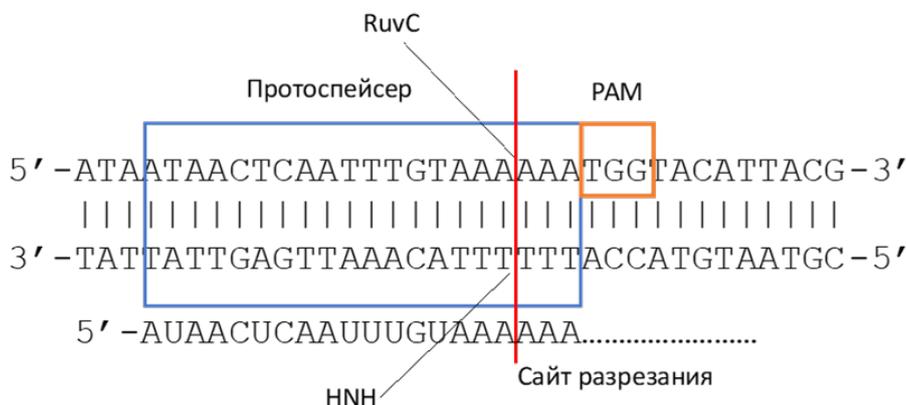
Условие

Длина этого вектора составляет _____¹ пар оснований. Этот вектор содержит (эндонуклеазу Cas9/каркас направляющей РНК (gRNA scaffold)/эндонуклеазу Cas9 и каркас направляющей РНК (gRNA scaffold))² и ген устойчивости к антибиотику (тетрациклин/канамицин/ампициллин/спектромицин/бластицидин)³.

Ответ: 1 — 14873 (14 873; 14,873), 2 — эндонуклеазу Cas9 и каркас направляющей РНК (gRNA scaffold), 3 — ампициллин.

Задача IV.1.2.10. Протоспейсер и РАМ (3 балла)

Несоответствие последовательности ДНК и sgРНК в пределах 8–10 нуклеотидов от РАМ ингибирует связывание такого протоспейсера, а значит и разрезание.



Условие

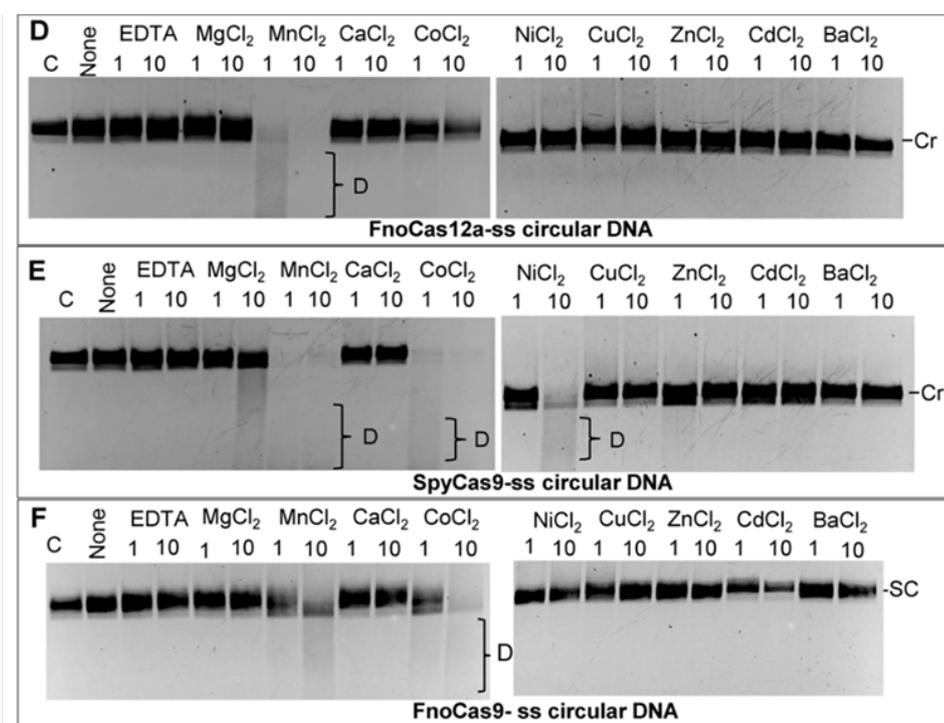
Определите какое теоретическое количество последовательностей ДНК с заменами вне затравочной области будет узнаваться и эффективно разрезаться комплексом

Cas9/sgРНК, если длина затравочной области — 10 п. о., а количество замен пар оснований ≤ 2 .

Ответ: 435 ± 1 .

Задача IV.1.2.11. Активность белков Cas (3 балла)

Без направляющих РНК белки из системы CRISPR/Cas могут неспецифически расщеплять ДНК. Влияние различных ионов двухвалентных ионов металлов на расщепление одноцепочечной кольцевой ДНК с помощью Cas12a и Cas9 из *Francisella tularensis novicida* (FnoCas12a и FnoCas9), Cas9 из *Streptococcus pyogenes* (SpyCas9) показано на рисунке.



Обозначения

- С — контрольный образец без белка.
- EDTA — этилендиаминтетрауксусная кислота.
- None — реакция с белком без добавления ионов металла.
- D — деградация ДНК.
- Cr — кольцевая одноцепочечная ДНК.
- 1 и 10 — концентрация солей, мМ.

Условие

Какой белок (FnoCas12a, FnoCas9 или SpyCas9) разрезает плазмиду в присутствии всех исследуемых двухвалентных ионов? (*FnoCas12a/FnoCas9/SpyCas9/никакой*)¹

Какой белок (FnoCas12a, FnoCas9 или SpyCas9) разрезает суперскрученную плазмиду в присутствии Mg²⁺? (*FnoCas9/FnoCas12a/SpyCas9/никакой*)²

Катион(ы) какого(их) элемента(ов) разрезает(ют) ДНК как при концентрации 10 мМ, так и при 1 мМ? ($Ca^{2+}/Co^{2+}/Mg^{2+}/Mn^{2+}$)³

Ответ: 1 — никакой, 2 — SpyCas9, 3 — Co^{2+} , Mg^{2+} .

Третий блок задач

Задача IV.1.3.1. Поиск ортологов гена EGFL6 при помощи нуклеотидного BLAST (3 балла)

В заданиях второго блока предлагалось изучить систему геномного редактирования CRISPR/Cas9, нацеленную на нокаут гена EGFL6 на клеточных линиях рака молочной железы человека (SKOV3). Однако, изучать этот ген можно не только на клеточных линиях человека, в качестве модельных организмов можно выбрать другие, близкие к человеку организмы, содержащие этот ген. Для того, чтобы найти ортологи (копии этого гена у других организмов) гена EGFL6, можно воспользоваться инструментом NCBI BLAST.

Условие

Используя самый длинный транскрипт, кодируемый геном EGFL6 (идентификатор NM_001167890.2), выполните поиск при помощи BLAST по базе данных nr, с использованием алгоритма megablast, со стандартными настройками.

Самая близкая (имеющая наибольший процент идентичных нуклеотидов) к транскрипту NM_001167890.2 последовательность, не относящаяся к человеку, относится к виду (*Pan paniscus/Pongo abelii/Pan troglodytes/Sapajus apella/Papio anubis*)¹ и имеет идентификатор (*XM_024241039.1/AJ245671.1/NM_001167890.2/XM_003805687.3/BC038587.1*)². В результатах BLAST (*присутствуют/отсутствуют*)³ последовательности, относящиеся к грызунам. В результатах BLAST отсутствуют последовательности, относящиеся к модельному виду *Danio rerio*. Это означает, что (*у этого вида отсутствует ген EGFL6/у этого вида может присутствовать ген EGFL6, однако он менее похож на ген EGFL6 человека, чем те гены, которые представлены в результате/у этого вида может присутствовать ген EGFL6, однако его невозможно найти при помощи алгоритма megablast/у этого вида может присутствовать ген EGFL6, однако его невозможно найти в базе данных nr*)⁴.

Ответ: 1 — *Pan paniscus*, 2 — XM_003805687.3, 3 — присутствуют, 4 — у этого вида может присутствовать ген EGFL6, однако он менее похож на ген EGFL6 человека, чем те гены, которые представлены в результате.

Задача IV.1.3.2. Поиск ортологов гена EGFL6 при помощи аминокислотного BLAST (2 балла)

В предыдущем задании был выполнен поиск ортологов гена EGFL6 при помощи нуклеотидного BLAST. Но существует также версия BLAST для аминокислотных последовательностей.

Условие

Используя последовательность белка, кодируемого геном EGFL6 (идентификатор NP_001161362.1) выполните поиск при помощи blast по базе данных nr с использованием алгоритма blastp, со стандартными настройками. Сравните результат с тем, который был получен на предыдущем шаге.

Количество отрядов класса млекопитающие, для которых были найдены последовательности при помощи нуклеотидного blast = _____¹, при помощи белкового blast = _____². Это значит, что более эволюционно далекие гомологи позволяет найти (*нуклеотидный blast/аминокислотный blast*)³, потому что (*нуклеотидные/аминокислотные*)⁴ замены в ходе эволюции накапливаются медленнее.

Ответ: 1 — 4, 2 — 6, 3 — аминокислотный blast, 4 — аминокислотные.

Задача IV.1.3.3. Поиск направляющей РНК (5 баллов)

При помощи программирования на Python можно автоматизировать процесс подбора направляющих РНК для геномного редактирования с использованием системы CRISPR/Cas9.

Условие

Напишите программу, которая принимает на вход последовательность ДНК (в направлении 5'–3'). В этой последовательности найдите РАМ для эндонуклеазы Cas9 из *Streptococcus pyogenes* (SpyCas9). Затем найдите последовательность направляющей РНК (20 нуклеотидов к 5'-концу от РАМ, не включая РАМ) и выведите ее на экран (в направлении 5'–3').

Тестирование будет проводиться на языке Python версии 3.10. В последовательности, подаваемой на вход, будет ровно 1 РАМ.

Примеры

Пример №1

| |
|--|
| Стандартный ввод |
| АТАСАТАСАТАGACAGATCATCAGACGACGGACATACATAGACATG |
| Стандартный вывод |
| АТАGACAGATCATCAGACGA |

Пример №2

| |
|---|
| Стандартный ввод |
| АТАТGАТАGАТАСАТGАСАТGАСАGGАТАСАСАТАСАGАСАТG |
| Стандартный вывод |
| TGATAGATACATGACATGAC |

Примеры

Пример №1

| |
|----------------------|
| Стандартный ввод |
| ATAGACAGATCATCAGACGA |
| Стандартный вывод |
| True |

Пример №2

| |
|----------------------|
| Стандартный ввод |
| TGATAGATACATGACATGAC |
| Стандартный вывод |
| False |

Пример №3

| |
|----------------------|
| Стандартный ввод |
| GAGACACTGATCACATCGTA |
| Стандартный вывод |
| True |

Пример №4

| |
|----------------------|
| Стандартный ввод |
| GCGCGACGATGACACGGGCA |
| Стандартный вывод |
| False |

Пример №5

| |
|-----------------------|
| Стандартный ввод |
| AAAAAAAAAAAAAAAAAAAAA |
| Стандартный вывод |
| False |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 k = input().lower()
2 res = (k.count('g') + k.count('c')) / len(k) * 100
3 print(40 <= res <= 60)
```

Задача IV.1.3.5. Поиск и фильтрация нРНК (5 баллов)

Условие

Напишите программу, которая будет принимать на вход последовательность ДНК (в направлении 5'–3'). В этой последовательности необходимо найти все направляющие РНК длиной 20 нуклеотидов для работы с эндонуклеазой SruCas9 и сохранить их в список. Произведите фильтрацию найденных направляющих РНК по ГЦ-составу и сохраните только те, у которых ГЦ-состав находится в диапазоне 40–60%. Сохраните все подходящие последовательности нРНК в список, отсортируйте его по возрастанию (при помощи функции `sorted()`) и выведите на экран при помощи функции `print()`.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--|
| Стандартный ввод |
| ATCATACACACATACACATGTACAGGTATGCCGACATATACAGGTGCATTGGGACA |
| Стандартный вывод |
| ['ACAGGTATGCCGACATATAC', 'CCGACATATACAGGTGCATT', 'GCCGACATATACAGGTGCAT'] |

Пример №2

| |
|---|
| Стандартный ввод |
| ACGACGTACGATCGATCASCATCGACGGTATACACATTACA |
| Стандартный вывод |
| ['TACGATCGATCASCATCGA'] |

Пример №3

| |
|--|
| Стандартный ввод |
| ATATATGCCGACGACGACGTGACGATCGATCGATCGAGCTGGTACACG |
| Стандартный вывод |
| ['TGACGATCGATCGATCGAGC'] |

Пример №4

| |
|--|
| Стандартный ввод |
| ATATATGCCGACGACGACGACGACGACGTGGTGTACACTTGGTACACG |
| Стандартный вывод |
| ['ACGACGACGTGGTGTACACT'] |

Пример №5

| |
|---|
| Стандартный ввод |
| ATAGCASCATGAGCAGCAGCGCTAGTATATACGAGCAGGTGAGGCATATACAGAC |
| Стандартный вывод |
| ['CAGCGCTAGTATATACGAGC', 'CTAGTATATACGAGCAGGTG'] |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 seq = input()
2 PAM = ['AGG', 'TGG', 'CGG', 'GGG']
3 nRNA = []
4 pos_all = []
5 nRNA_f = []
6 for i in range(4):
7     PAM_ex = PAM[i]
8     if seq.count(PAM_ex) != 0:
9         pos = [j for j in range(len(seq)) if seq.startswith(PAM_ex, j)]
10        for k in pos:
11            nRNA.append(seq[k-20:k])
12 for i in nRNA:
13     GC = (i.count('G') + i.count('C'))/20
14     if GC >= 0.4 and GC <= 0.6:
15         nRNA_f.append(i)
16 print(sorted(nRNA_f))
```

Геномное редактирование

Во втором блоке вы изучили систему для нокаута гена EGFL6 на клеточных линиях человека: как устроен вектор для клонирования, какую последовательность PAM будет узнавать эндонуклеаза Cas9 и какой должен быть результат редактирования в случае успешного нокаута.

В третьем блоке заданий вам предстоит выполнить рестрикционный анализ плазмиды со вставкой, выполнить ПЦР *in silico* и проанализировать результаты электрофореза.

Ссылки для скачивания последовательностей векторов

1. Исходная плаزمида lentiCRISPRv2: <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/244694/51ae3e01-5e94-49b8-a4fb-7fa303a7eb81/addgene-plasmid-52961-sequence-244694.gbk>.
2. Плазмида lentiCRISPRv2 со вставкой: <https://disk.yandex.ru/d/b2Q6q40NE62TWA>.
3. SibEnzyme маркер длин ДНК 1 kb: <http://sibenzyme.com/product/1-kb-13-fragmentov-ot-0-25-do-10-kb/>.

Задача IV.1.3.6. Рестрикционный анализ плазмиды со вставкой (2 балла)

В предыдущем блоке вы выполнили клонирование олигонуклеотида направляющей РНК в вектор lentiCRISPRv2. На данном этапе необходимо проверить наличие вставки в плазмиде.

Часто для проверки наличия вставки в плазмиде используют рестрикционный анализ. Для этого плазмиду обрабатывают рестриктазами и наносят продукты рестрикции на агарозный гель. В случае наличия вставки образуются другие продукты

рестрикции, чем в плазмиде без вставки, и их можно отличить от исходных по длине продукта(ов) на электрофорезе.

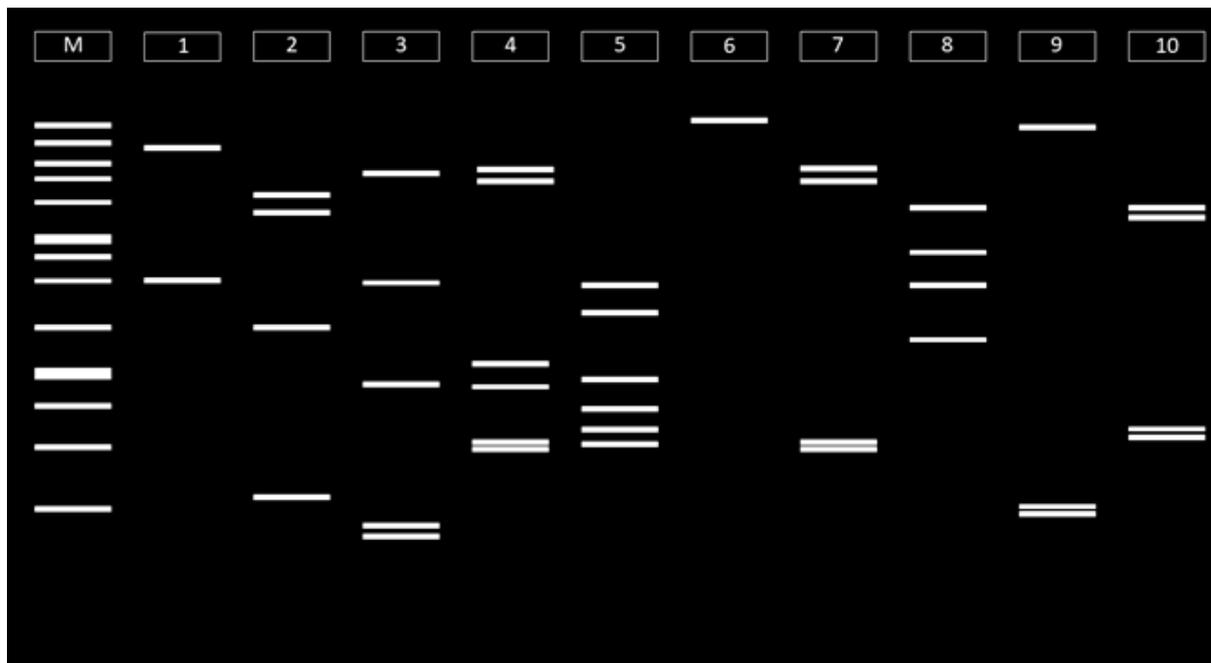
Условие

В лаборатории имеются следующие рестриктазы: *FspI*, *SpeI*, *XcaI*, *XhoI*, *HindIII*, *PvuI*. Только одна из них позволит отличить плазмиду со вставкой от плазмиды без вставки на электрофорезе. Определите, какая рестриктаза подходит по условиям задачи, введите ее название.

Ответ: *HindIII*.

Задача IV.1.3.7. Анализ результатов электрофореза (3 балла)

Плазмиду со вставкой обработали выбранной рестриктазой и нанесли продукты рестрикции на агарозный гель. На рисунке изображена схема результата электрофореза.



На дорожку 1 нанесен 1 kb маркер (SibEnzyme), на остальные дорожки продукты рестрикции различных плазмид различными рестриктазами. Только в одной из колонок присутствует плазида lentiCRISPRv2 со вставкой олигонуклеотида направляющей РНК.

Условие

Определите дорожку, на которую нанесена плазида, lentiCRISPRv2 и введите ее номер (целое число).

Ответ: 7.

Задача IV.1.3.8. Дизайн праймеров (3 балла)

Условие

На предыдущем этапе вы выполнили проверку на наличие вставки в плазмиде при помощи рестрикционного анализа. Вы выбрали плазмиду с нужной длиной фрагментов, образовавшихся после рестрикции. Это гарантирует наличие вставки в плазмиде, но не гарантирует, что вставка имеет нужную последовательность. Иногда в ходе синтеза олигонуклеотида могут возникнуть ошибки, которые могут привести к изменению его последовательности (замена или инсерция/делеция нуклеотидов). Поэтому необходимо выполнить еще одну проверку, выполнив секвенирование участка со вставкой олигонуклеотида по Сэнгеру. Для того, чтобы выполнить секвенирование по Сэнгеру, на первом этапе нужно выполнить ПЦР, и для этого было предложено 5 пар праймеров:

- F: 5'-CGTATGTTCCCATAGTAACG-3',
R: 5'-CACCTCGTCCACGATGTTGC-3'.
- F: 5'-GAAGGTGGAGAGAGACAGAGA-3',
R: 5'-CTAGGCACCGGATCAATTGCC-3'.
- F: 5'-TCGATTAGTGAACGGATCGG-3',
R: 5'-CTTGCTCTGGTCGAAGAAAATCT-3'.
- F: 5'-TTTATTACAGGGACAGCAGAGAT-3',
R: 5'-GCACCGACTCGGTGCCACTTTT-3'.
- F: 5'-TTTTACAAAGTGCGCGAGATCAAC-3',
R: 5'-AGCTGTTTCTGCTCATTATCCT-3'.

Выберите, какая из предложенных пар праймеров удовлетворяет следующим критериям:

- Амплифицирует участок со вставкой.
- Имеет длину ампликона в диапазоне 100–1000 нуклеотидов.
- Имеет ГЦ-состав в диапазоне 40–60%.
- Разница в температуре плавления между праймерами не более 3 градусов.

Выполните *in silico* ПЦР с выбранной парой праймеров в программе UGENE и введите в ответ длину ампликона

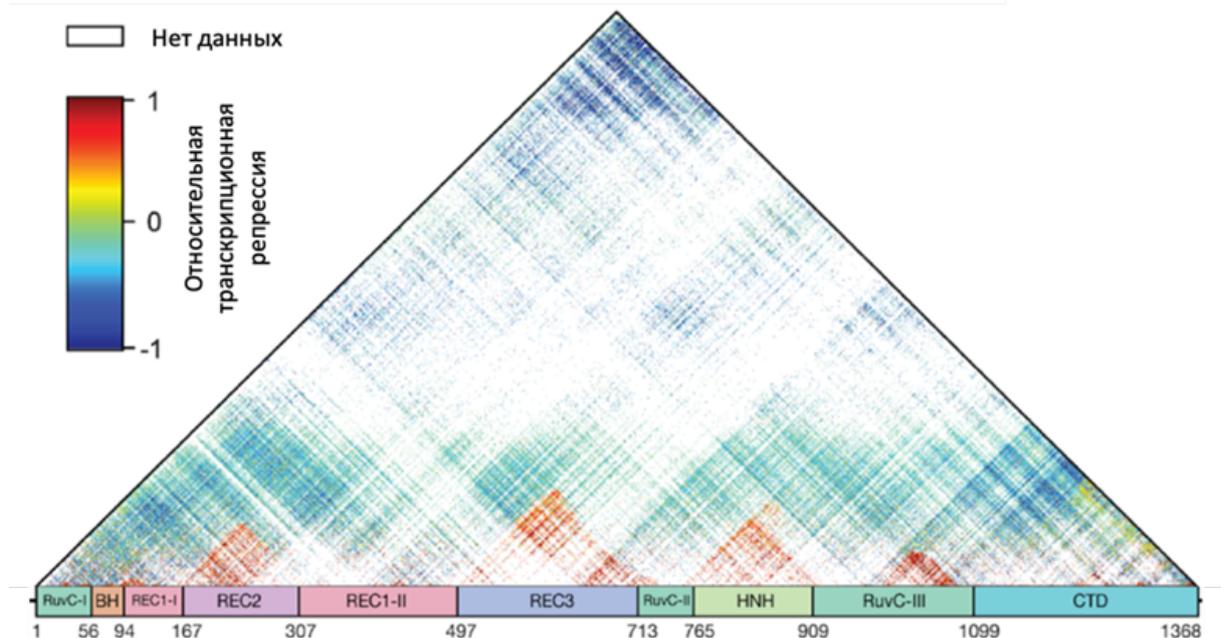
Ответ: 744.

Задача IV.1.3.9. Мертвый Cas и «тайна пирамид» (6 баллов)

В недавнем исследовании ученые задались вопросом: можно ли уменьшить dCas9 («мертвый» Cas9, длина — 1368 а. о.), сохранив его важную функцию — связывание с ДНК посредством sgРНК?

В результате был получен белок Δ dCas9, уменьшенный на 36%, и способный связывать ДНК в экспериментах *in vitro* и *in vivo*. В этом задании вам предлагается исследовать делеции в варианте Δ dCas9.

На рисунке представлена треугольная диаграмма, каждая точка которой предоставляет информацию об относительной репрессии системы для детекции делеций и размер делеции.



Темно-синим обозначены варианты, которые не репрессируют систему, и не связываются с ДНК. Темно-красным — варианты, которые репрессируют систему и связываются с ДНК. Размер делеций и расположение их в доменах можно определить с помощью проведения из любой точки диаграммы отрезков, параллельных боковым сторонам треугольника, к его основанию. На основании этих результатов исследователи создали вариант Δ dCas9, с несколькими делециями.

Условие

Определите, в каких доменах (RuvC-I, BH, REC1-I, REC2, REC1-II, REC3, RuvC-II, HNH, RuvC-III, CTD) произвели делеции больше 56 а. о., чтобы максимально уменьшить dCas9 из *Streptococcus pyogenes*, и ответьте на вопросы.

Условием задачи удовлетворяет $(1/2/3/4/5/6)^1$ делеция(и,ий). Домен/часть домена, в котором присутствует самая маленькая делеция (*RuvC-I/BH/REC1-I/REC2/REC1-II/REC3/RuvC-II/HNH/RuvC-III/CTD*)².

Нуклеазный домен (*RuvC-I/BH/REC1-I/REC2/REC1-II/REC3/RuvC-II/HNH/RuvC-III/CTD*)³ не разделен на части вдоль последовательности белка. Распознающий домен _____⁴ разделен на части вдоль последовательности белка. Домен, который ответственен за связывание с PAM, имеет размер _____⁵ остатков аминокислот. В домене (*RuvC-I/BH/REC1-I/REC2/REC1-II/REC3/RuvC-II/HNH/RuvC-III/CTD*)⁶ обнаружена самая большая делеция.

Ответ: 1 — 4, 2 — RuvC-III, 3 — HNH, 4 — REC1, 5 — 269, 6 — REC3.

Задача IV.1.3.10. Шпильки, петли и Cas9 (6 баллов)

Вторичная структура sgРНК в комплексе с Cas9 имеет следующие элементы, начиная с 5'-конца последовательности:

- спейсер длиной 20 нуклеотидов;
- последовательность повтор-антиповтор (шпилька с внутренней петлей);
- шпилька №1 с выпетливанием;
- линкер — одноцепочечный участок РНК, длиной 5 и больше нуклеотидов;
- шпилька №2;
- шпилька №3.

Первичная структура sgРНК:

5'-GGAAAUUAGGUGCGCUUGGCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU
AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU-3'.

Смоделируйте вторичную структуру sgРНК (структура №1) из ее первичной последовательности с помощью биоинформатической платформы RNAfold или UNAFold¹. Извлеките вторичную структуру sgРНК (структура №2) из рентгеноструктурных данных комплекса Cas9/sgРНК (PDB ID: 4O08), используя сервис RNAPdbee².

Необходимые ссылки

- <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>.
- <http://www.unafold.org/mfold/applications/rna-folding-form.php>.
- <http://rnadbее.cs.put.poznan.pl>.

Условие

Сравните вторичные структуры sgРНК (структура №1 и №2) и выберите верные утверждения.

1. В обеих структурах присутствует последовательность повтор-антиповтор (шпилька с внутренней петлей).
2. Только в структуре №2 присутствует последовательность повтор-антиповтор (шпилька с внутренней петлей).
3. В обеих структурах присутствует шпилька №1 с выпетливанием.
4. Только в структуре №2 присутствует шпилька №1 с выпетливанием.
5. В обеих структурах присутствует линкер.
6. Линкер имеется только в структуре №2.
7. В обеих структурах присутствует шпилька №2.
8. Шпилька №2 имеется только в структуре №2.
9. В обеих структурах присутствует шпилька №3.
10. Шпилька №3 имеется только в структуре №2.
11. В структуре №1 спейсер связан с РНК.
12. В структуре №2 спейсер связан с РНК.
13. В структуре №1 спейсер связан с ДНК.
14. В структуре №2 спейсер связан с ДНК.

¹Выберите структуру с минимальной энергией (ΔG).

²Для всех онлайн-платформ используйте параметры по умолчанию.

Ответ: 1, 4, 6, 7, 9, 11, 14.

Задания для 10–11 классов

Первый блок задач

Задача IV.2.1.1. Приготовление разведений растворов (1 балл)

Темы: расчеты.

Концентрацию компонентов в растворе обозначают различными способами. Широко используют количественные характеристики, например, г/л, моль/л (М), % и другие. При приготовлении растворов для нанесения образцов на гель или при расчете компонентов смеси для ферментативных реакций часто используют кратные растворы (2х, 4х, 5х, 10х). Например, для приготовления 100 мл однократного водного раствора (1х), нужно взять 50 мл двукратного раствора (2х) и добавить 50 мл дистиллированной воды.

Условие

Вычислите, в какой объем реакционной смеси нужно добавить 15 мкл шестикратного (6х) буферного раствора, чтобы в итоговой смеси буферный раствор стал двукратным (2х)? Ответ введите в виде натурального целого числа без «мкл».

Ответ: 30.

Задача IV.2.1.2. Приготовление реакционной смеси (1 балл)

Темы: расчеты.

Согласно протоколу реакционная смесь для рестрикции должна содержать $7,553 \cdot 10^{11}$ молекул плазмиды pUC19 (длина 2686 пар нуклеотидов). В лаборатории имеется раствор с концентрацией pUC19 1 мкг/мкл. При вычислениях используйте молекулярную массу одной пары нуклеотидов 660 г/моль.

Условие

Вычислите, какой объем раствора pUC19 необходимо добавить в пробирку с реакционной смесью. Ответ введите в виде натурального целого числа без «мкл», округлите до ближайшего целого по правилам округления.

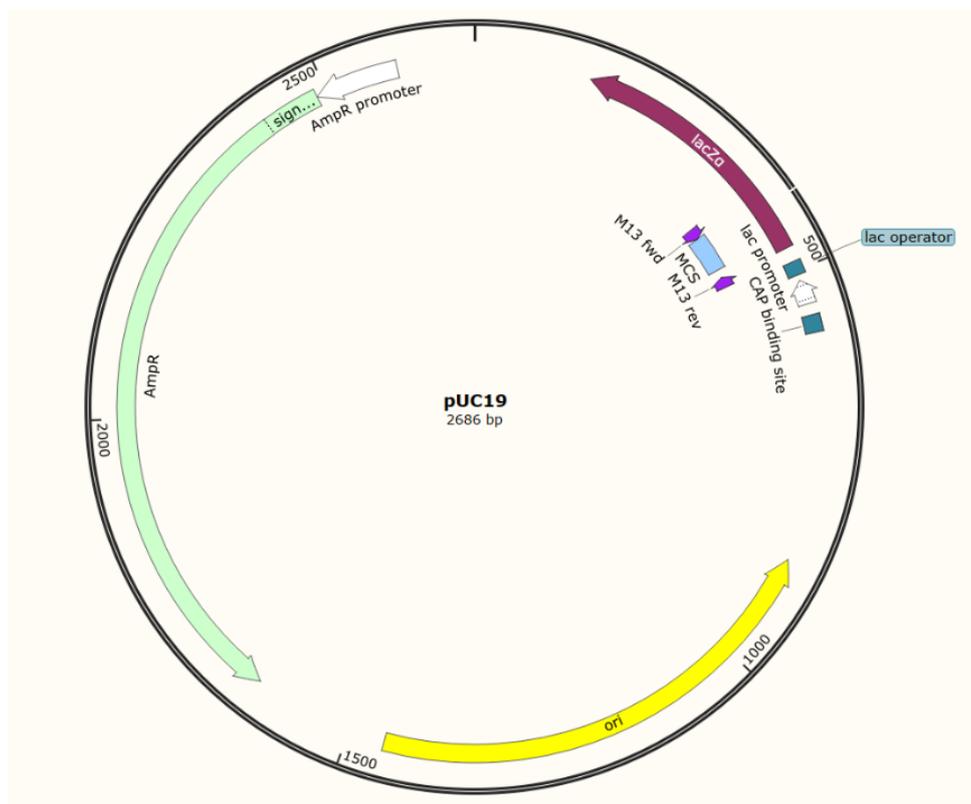
Ответ: 2.

Задача IV.2.1.3. Определение числа фрагментов ДНК после рестрикции (1 балл)

Темы: рестрикция.

В результате гидролиза плазмиды pUC19 (длина 2686 п. н.) рестриктазами TsoI и PciI получены некоторые фрагменты. Файл с последовательностью плазмиды можно

скачать по ссылке: https://disk.yandex.ru/d/W_0IIny-I9n5ug.



Используя программу UGENE, карту плазмиды pUC19, проведите анализ фрагментов ДНК, которые образуются в результате рестрикции.

Условие

Определите разницу длин более длинного и более короткого фрагментов плазмиды, образовавшихся после рестрикции двумя ферментами. Введите ответ в виде числа.

Ответ: 102.

Задача IV.2.1.4. Спектрофотометрический анализ ДНК (5 баллов)

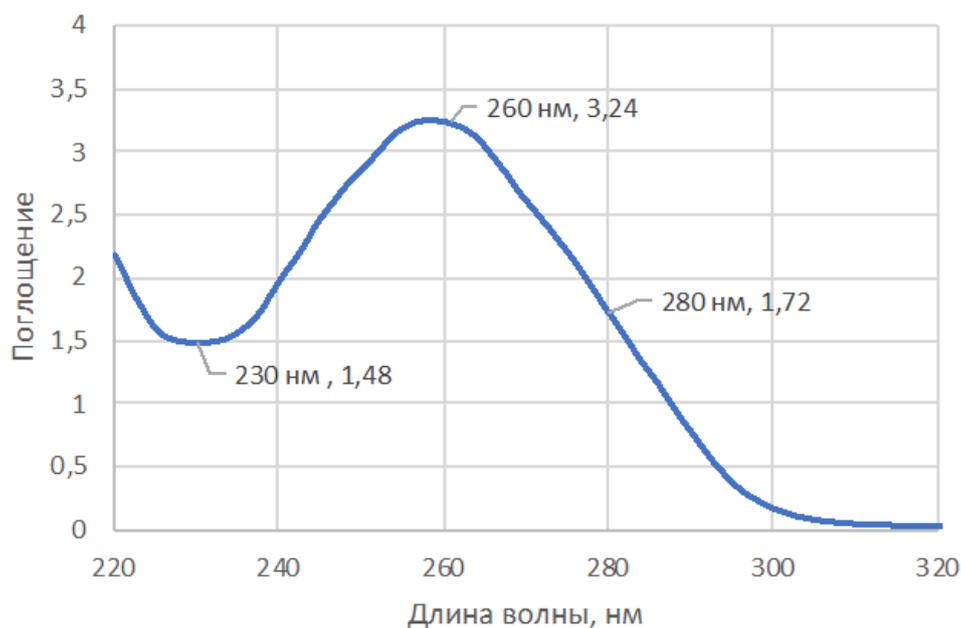
Темы: расчеты.

Соотношение оптической плотности при 260/280 нм используют для оценки чистоты ДНК. Соотношение $\sim 1,8$ считают характерным для препарата ДНК, не содержащего примесей. Если отношение заметно ниже, это может указывать на присутствие примесей белков, фенола или других загрязняющих веществ, которые сильно поглощают на длине волны 280 нм.

Другим показателем «чистоты» ДНК является соотношение оптической плотности при 260/230 нм. Значения 260/230 для «чистой» нуклеиновой кислоты часто выше, чем соответствующие значения 260/280. Ожидаемые значения 260/230 обычно находятся в диапазоне 2,0–2,2. Если отношение заметно ниже ожидаемого, это может указывать на присутствие загрязняющих веществ, которые поглощают при

230 нм (фенол, изотиоцианат и другие органические соединения) и используются при выделении и очистке ДНК.

Спектр поглощения ДНК приведен на рисунке.



Условие

Рассчитайте концентрацию плазмиды pCas9 (Addgene, #42876) в нг/мкл, если известно, что для 1 единицы поглощения при 260 нм (A_{260}) концентрация ДНК равна 50 мкг/мл. Переведите эту концентрацию в наномолярную (нМ), зная среднюю молярную массу пары нуклеотидов — 660 г/(моль·п. о.). Вычислите соотношения A_{260}/A_{280} и A_{260}/A_{230} . Заполните пропуски.

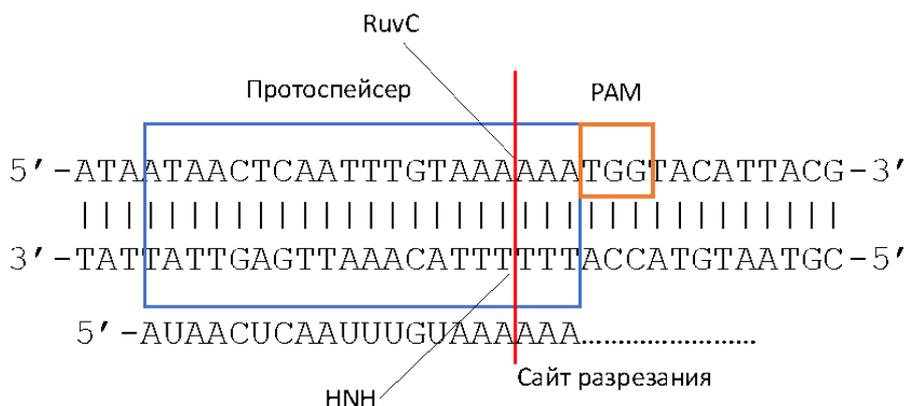
Концентрация плазмиды в нг/мкл ($16/26/62/126/162/216$)¹. Концентрация плазмиды в нМ ($16/26/62/126/162/216$)². Соотношение A_{260}/A_{280} ($1.66/1.88/2.19/0.53/0.46/0.93/1.13$)³. Соотношение A_{260}/A_{230} ($1.66/1.88/2.19/0.53/0.46/0.93/1.13$)⁴.

Ответ: 1 — 162, 2 — 26, 3 — 1.88, 4 — 2.19.

Задача IV.2.1.5. Геномное редактирование (1 балл)

Темы: CRISPR.

Сравнительная стабильность гетеродуплекса в затравочной области критична для узнавания: несоответствие последовательности ДНК и sgРНК в пределах 8–10 нуклеотидов от РАМ ингибирует связывание такого протоспейсера.



Условие

Определите какое теоретическое количество последовательностей ДНК с заменами будет узнаваться и эффективно разрезаться комплексом Cas9/sgРНК, если длина затравочной области — 8 п. о., а количество замен пар оснований ≤ 2 .

Список литературы

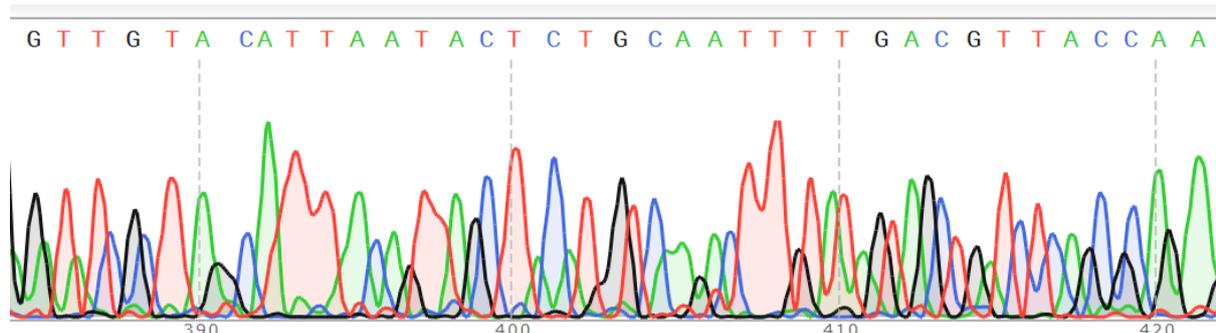
- Википедия: CRISPR — <https://ru.wikipedia.org/wiki/CRISPR>.
- Нариман Баттулин (ИЦИГ СО РАН, НГУ) «От геномного редактирования к синтетической биологии» — <https://youtu.be/SD1qAo5YZ94>.
- Чл-корр РАН Дмитрий Жарков (ИХБФМ СО РАН, НГУ) «Геномное редактирование» — <https://youtu.be/NSGvAFnPIkE>.
- Константин Северинов «Редактирование генома» (Сколтех) — https://vk.com/video-46417318_456258673.

Ответ: 630.

Задача IV.2.1.6. Анализ результатов секвенирования по Сэнгеру (1 балл)

Темы: Сэнгер.

Результаты секвенирования ДНК по Сэнгеру можно представить в виде «секвенограммы», пики которой соответствуют последовательности нуклеотидов в анализируемой молекуле.



Условие

Используя инструмент Nucleotide BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), определите организм, к которому относится данный фрагмент ДНК. Выберите варианты ответа, относящиеся к таксономическому (систематическому) положению.

1. Род *Pseudomonas*.
2. *Escherichia coli*.
3. Класс *Gamma*proteobacteria.
4. *Bacillus subtilis*.
5. Грамположительная бактерия.
6. Грамотрицательная бактерия.

Ответ: 1, 3.

Задача IV.2.1.7. Полимеразная цепная реакция (1 балл)

Темы: ПЦР.

Полимеразную цепную реакцию используют для амплификации специфического фрагмента ДНК. В задании приведены суждения, часть из которых относится к ПЦР с детекцией продуктов по конечной точке (в агарозном геле с использованием бромистого этидия).

Условие

Выберите все варианты ответа, подходящие для данного способа детекции продуктов ПЦР.

1. Реакционная смесь содержит термостабильную ДНК-полимеразу.
2. Реакционная смесь содержит термостабильную обратную транскриптазу.
3. Реакционная смесь содержит термостабильную обратную полимеразу.
4. Реакционная смесь содержит раствор интеркалирующего красителя EtBr.
5. Реакционная смесь содержит раствор интеркалирующего красителя SYBR.
6. Реакционная смесь содержит раствор АТФ и другие рибоолигонуклеотиды.
7. В результате каждого цикла ПЦР происходит удвоение специфического фрагмента ДНК.
8. Реакционная смесь отрицательного контрольного образца не содержит ДНК-полимеразу.
9. При детекции по конечной точке обычно используют температуру отжига праймеров 72 °С.
10. Для того, чтобы снизить количество неспецифических продуктов, используют «горячий старт».

Список литературы

- Рекомендации по постановке ПЦР от Евроген — <http://evrogen.ru/kit-use-r-manuals/Evrogen-PCR-recommendation.pdf>.

-
- Биомолекула: 12 методов в картинках: полимеразная цепная реакция — <https://biomolecula.ru/articles/metody-v-kartinkakh-polimeraznaia-tsepnaia-reaktsiia>.
 - Лекция «Полимеразная цепная реакция» — <https://youtu.be/OGLMcUz0044>.

Ответ: 1, 7, 10.

Задача IV.2.1.8. Генетически модифицированные животные (15 баллов)

Умение анализировать научную информацию, работать с литературными источниками, письменно излагать свои мысли — является крайне востребованным навыком для молодого исследователя. Опыт проведения второго отборочного тура НТО в 2019–2022 году показал, что написание короткого текста с развернутым ответом на вопрос оказалось одним из самых трудных и интересных заданий.

Для выполнения творческого задания *можно и нужно использовать литературные источники* (в том числе, тексты в Интернете), но не стоит копировать большие фрагменты текста. Размер требуемого текста ограничен, помните, что краткость — сестра таланта.

P.S. Эксперты будут читать ваши работы с интересом и удовольствием!

Условие

Напишите текст (эссе) объемом **не более 3000 символов (с учетом пробелов)** на тему «Зачем нужно изучать древнюю ДНК?».

Постарайтесь в ответе привести аргументированную точку зрения. Авторы лучших текстов получают специальные призы от спонсоров и партнеров профиля «Геномное редактирование».

Перед работой над эссе рекомендуем посмотреть лекцию Александра Пилипенко (ИЦИГ СО РАН, НГУ) «О Нобелевской премии за исследование эволюции человека методами палеогенетики» — <https://youtu.be/-1rsoB0EF8U>.

Критерии оценивания

- Длина эссе — не более 3000 символов (с учетом пробелов).
- Правильная аргументация, обоснование ответа.
- Отсутствие биологических ошибок.
- Ссылки на литературные источники.
- Текст написан лично автором, отсутствуют заимствованные фрагменты текста.
- В зачет идет одно лучшее эссе от команды.
- При наличии нескольких одинаковых/стереотипных текстов, будет оцениваться текст, загруженный первым, остальные будут оценены в 0 баллов.
- При совпадении фрагментов текстов у разных команд, будет оценен первый загруженный текст, остальные будут оценены в 0 баллов.
- При превышении рекомендованной длины текста, оцениваться будут первые 3000 символов (с учетом пробелов).

-
- Список литературы включается (!) состав в 3000 символов.

Второй блок задач

Задача IV.2.2.1. Поиск информации о гене (2 балла)

RB1 — это онкосупрессорный ген, мутации в котором встречаются при многих онкологических заболеваниях человека. Этот ген кодирует белок, который ингибирует чрезмерное деление клеток.

Условие

Изучите этот ген в базах данных и выберите верные суждения.

Ген RB1 имеет идентификатор Gene ID _____¹. Этот ген экспрессируется в (*органах ЖКТ/органах дыхательной системы/нервной системе/половых органах/во всех перечисленных органах*)². Одна из изоформ этого гена имеет транскрипт с идентификатором (*NM_001206524.2/NM_005104.4/NP_001012270.1/NM_001407165.1/NP_001394094.1*)³. Идентификатор белка, кодируемого этим геном, в базе данных UniProt SwissProt — _____⁴. Этот белок локализуется в клеточном компартменте (*цитозоль/эндоплазматический ретикулум/ядро/митохондрии*)⁵. В базе данных Ensembl этот ген имеет идентификатор _____⁶. Делеция участка 48301356–48320412 на 13 хромосоме приведет к (*полной потере гена RB1/потере 1 и 2 экзона гена RB1/потере первых 4 экзонов гена RB1/не затронет ген RB1*)⁷.

Ответ: 1 — 5925; 2 — во всех перечисленных органах; 3 — NM_001407165.1; 4 — P06400, A8K5E3, P78499, Q5VW46, Q8IZL4; 5 — ядро; 6 — ENSG00000139687, 7 — потере первых 4 экзонов гена RB1.

Задача IV.2.2.2. Поиск информации о мутации (3 балла)

Условие

Так как ген RB1 вносит значительный вклад в развитие онкологического заболевания, существует множество тест-систем, направленных на детекцию мутаций в этом гене. Найденные мутации могут служить прогностическими маркерами (предсказывают тяжесть заболевания) или являться показаниями для назначения таргетной терапии.

Допустим, при помощи тест-системы вы получили следующую мутацию в гене RB1: замена цитозина на аденин в 48379599 позиции на 13 хромосоме. Изучите мутацию и выберите верные суждения.

Данная мутация является (*сионимичной/миссенс/нонсенс*)¹ и приводит к (*сохранению аминокислоты/замене изолейцина на глутамат/замене валина на фенилаланин/замене серина на метионин/появлению стоп-кодона*)² в _____³ позиции в белке. По данным базы данных ClinVar, эта мутация является (*патогенной/безвредной*)⁴.

Ответ: 1 — нонсенс, 2 — появлению стоп-кодона, 3 — 446, 4 — патогенной.

Знакомство с Python

Вторая группа заданий знакомит с установкой Python в дистрибутиве Anaconda Individual Edition. Вам необходимо написать простые скрипты в Spyder, который устанавливается вместе с Anaconda, и решить более сложные задачи.

Обязательно рекомендуем посмотреть на канале «Практическая биоинформатика» плейлист «Введение в Python» (<https://www.youtube.com/playlist?list=PL8nkomsB9uV0k5hqS9RQ8-TeMXH3P3D8b>).

Для тех, кто не знаком с языком программирования Python и средой Anaconda, составители заданий рекомендуют начать со следующих материалов.

- Курс «Программирование на Python» (первые два модуля)
- Установка Anaconda — <https://www.youtube.com/watch?v=D6o8CjF1DhE> (можно смотреть до 2:30 мин).
- Введение в Spyder — <https://www.youtube.com/watch?v=IT2qjsc4ThM>.
- «Учите питон» — <http://pythontutor.ru/>.

Задача IV.2.2.3. Приготовление реакционной смеси (2 балла)

В лабораторном протоколе реакции лигирования в состав реакционной смеси входят следующие компоненты:

- ДНК-лигаза — 1 мкл;
- буферный раствор — 2 мкл;
- фрагмент ДНК — 0,5 мкг/мкл;
- плаزمиды — 0,1 мкг/мкл;
- вода — до 20 мкл.

При этом необходимо каждый раз рассчитывать, какой объем фрагмента и плазмиды нужно добавлять в реакционную смесь. Этот процесс можно автоматизировать, написав скрипт на языке Python.

Условие

Напишите программу, которая будет принимать (через `input()`) строку, содержащую следующие значения, разделенные одним знаком пробела: имеющаяся концентрация фрагмента и имеющаяся концентрация плазмиды (в мкг/мкл). Программа должна вычислить объем растворов фрагмента и плазмиды, которые нужно добавить в реакционную смесь (в мкл), и округлить их до второго знака после запятой при помощи функции `round()`. Если концентрации растворов достаточно для того, чтобы приготовить реакционную смесь, программа выводит на экран (при помощи `print()`) 3 числа (разделенные одним знаком пробела): объем раствора с фрагментом, объем раствора с плазмидой, объем воды (необходимый, чтобы довести объем всей реакционной смеси до 20 мкл). В противном случае, программа выводит сообщение: «невозможно приготовить реакционную смесь при данных концентрациях растворов».

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--------------------------|
| Стандартный ввод |
| 0.98 1.23 |
| Стандартный вывод |
| 10.2 1.63 5.17 |

Пример №2

| |
|---|
| Стандартный ввод |
| 0.11 2.14 |
| Стандартный вывод |
| невозможно приготовить реакционную смесь при данных концентрациях растворов |

Пример №3

| |
|--------------------------|
| Стандартный ввод |
| 1.14 1.61 |
| Стандартный вывод |
| 8.77 1.24 6.99 |

Пример №4

| |
|--------------------------|
| Стандартный ввод |
| 5.17 8.12 |
| Стандартный вывод |
| 1.93 0.25 14.82 |

Пример №5

| |
|--------------------------|
| Стандартный ввод |
| 0.88 0.99 |
| Стандартный вывод |
| 11.36 2.02 3.62 |

Пример №6

| |
|---|
| Стандартный ввод |
| 0.01 0.02 |
| Стандартный вывод |
| невозможно приготовить реакционную смесь при данных концентрациях растворов |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 f, p = map(float, input().split())
2 a = round(10 / f, 2)
3 b = round(2 / p, 2)
4 c = round(20 - (3 + a + b), 2)
5 if c < 0:
6     print('невозможно приготовить реакционную смесь при данных концентрациях
7     ↪ растворов')
8 else:
9     print(a, b, c)
```

Задача IV.2.2.4. Поиск сайтов рестрикции (2 балла)

При помощи языка программирования Python можно работать с биологическими последовательностями, например искать сайты узнавания рестриктаз в последовательности ДНК.

Условие

Напишите программу, которая будет принимать на вход (через `input()`) две последовательности, разделенные одним знаком пробела: сайт узнавания рестриктазы и последовательность ДНК (всегда в направлении 5'-3'). Выполните проверку на наличие этого сайта узнавания в последовательности и выведите на экран (при помощи `print()`) количество сайтов рестрикции в последовательности.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|---|
| Стандартный ввод |
| AAGCTT ATATCATCTGATCATCATGATСТААGCTTTACTCTACTGTAGTATC |
| Стандартный вывод |
| 1 |

Пример №2

| |
|--|
| Стандартный ввод |
| CATATG AATACACCATATGATATGATGTACCATATGTATATTGACCATATG |
| Стандартный вывод |
| 3 |

Пример №3

| |
|---|
| Стандартный ввод |
| GGATCC ATTGATGATGATGATCATCACAТATGATGATCATCATTTTATACTGTATCAC |
| Стандартный вывод |
| 0 |

Пример №4

| |
|--|
| Стандартный ввод |
| GGATCC ATTGATGATGATGATCATCACTATGATGATCATCATGGATCCTTTATACTGTATCAC |
| Стандартный вывод |
| 1 |

Пример №5

| |
|--|
| Стандартный ввод |
| CCCC ATATATATATATGATGATGCGACTATATTATGT |
| Стандартный вывод |
| 0 |

Пример №6

| |
|--|
| Стандартный ввод |
| ATATA AGTGATGTATATAAGAGAGTCSACATATAGGCGCGGATATAGCGCGCATATAGCGCGC |
| Стандартный вывод |
| 4 |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 b, word = input().split()
2 c = word.count(b)
3 print (c)
```

Задача IV.2.2.5. Поиск сайтов рестрикции 2 (3 балла)

Существуют рестриктазы с более сложными сайтами узнавания. Они могут узнавать последовательности ДНК, содержащие сразу несколько различных нуклеотидов. Например, сайт узнавания рестриктазы Hru188I выглядит следующим образом: TCNGA. То есть в третьей позиции этого сайта может стоять любой нуклеотид (рестриктаза узнает сайты TCAGA, TCTGA, TCCGA, TCGGA). Сайт узнавания рестриктазы EaeI имеет последовательность YGGCCR, где Y — это нуклеотид C или T, а R — это нуклеотид G или A. Для обозначения сразу нескольких различных нуклеотидов используют дополнительные буквенные обозначения, с которыми можно ознакомиться по ссылке <https://ru.wikipedia.org/wiki/Нуклеотиды>.

Условие

Напишите программу, которая будет принимать на вход (через `input()`) две последовательности, разделенные одним знаком пробела: сайт узнавания рестриктазы и последовательность ДНК (всегда в направлении 5'-3'). Сайт узнавания может содержать (а может и не содержать) дополнительные буквенные коды, обозначающие

сразу несколько нуклеотидов. Выполните проверку наличия данного сайта узнавания (во всех возможных его вариациях) в последовательности. При помощи функции `print()` выведите на экран `True`, если такой сайт узнавания есть в последовательности и `False` в противном случае.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|----------------------------------|
| Стандартный ввод |
| TCNGA ATACATCAGAAATATCACCAATACCG |
| Стандартный вывод |
| True |

Пример №2

| |
|--|
| Стандартный ввод |
| YGGCCR ATATATСТАСТАТATGATTATCATGTGTATCAT |
| Стандартный вывод |
| False |

Пример №3

| |
|---|
| Стандартный ввод |
| YGGCCR GTATACACATGТАСТАТCGGCCGCATACATTA |
| Стандартный вывод |
| True |

Пример №4

| |
|------------------------------------|
| Стандартный ввод |
| TCNGA AAAAAAATTTTTTTTCCCCCGGGGGGGG |
| Стандартный вывод |
| False |

Пример №5

| |
|--------------------------|
| Стандартный ввод |
| YYYYY ATCTCTCTTCAA |
| Стандартный вывод |
| True |

Пример №6

| |
|--|
| Стандартный ввод |
| WWW GCGCGCCGCGCGCGCGCGCGATGCGCAGGTGGCA |
| Стандартный вывод |
| False |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 import re
2 site, dna = input().split()
3 letrs = {
4     'M':r'(A|C)',
5     'R' : r'(A|G)',
6     'W' : r'(A|T)',
7     'S' : r'(C|G)',
8     'Y' : r'(C|T)',
9     'K' : r'(G|T)',
10    'V' : r'(A|C|G)',
11    'H' : r'(A|C|T)',
12    'D' : r'(A|G|T)',
13    'B' : r'(C|G|T)',
14    'N' : r'(A|C|G|T)',
15    'A': r'A',
16    'C':r'C',
17    'G':r'G',
18    'T':r'T'
19 }
20 pat = ''
21 for i in site:
22     pat += letrs[i]
23 answ = bool(re.search(pat, dna))
24 print(answ)
```

Задача IV.2.2.6. Поиск сайтов рестрикции — поиск для конкретных рестриктаз (2 балла)

Исследователь в лаборатории работает в основном с рестриктазами XhoI, NdeI, SpeI, EcoRI, HindIII, BamHI. Данная задача проверяет, содержит ли последовательность сайты узнавания этих рестриктаз.

Условие

Напишите программу, которая будет принимать на вход (через `input()`) последовательность и проверяет, есть ли в ней сайты узнавания перечисленных выше рестриктаз. Выведите на экран (в алфавитном порядке, по возрастанию) все рестриктазы, сайты которых есть в последовательности. Если последовательность не содержит сайты узнавания ни одной рестриктазы, выведите на экран.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--|
| Стандартный ввод |
| AGTGATGTATATAAGAGACATATGGTCACATATAGGCGCGGATATAGCGGCATATAGGATCCGCGCGCTC GAGC |
| Стандартный вывод |
| BamHI NdeI XhoI |

Пример №2

| |
|--|
| Стандартный ввод |
| AGTGATGAATTCGTATATAAGAGAGTCACATATAGGCGCGGATATAGCGGCATATAGCGCGC |
| Стандартный вывод |
| EcoRI |

Пример №3

| |
|--|
| Стандартный ввод |
| AATACTCASSAATAGСТААТАСГАТСГАТСГААТСГАСТАГСТАСГАТСГАТАТАСАСГАТСГААТАСАТ |
| Стандартный вывод |
| False |

Пример №4

| |
|--|
| Стандартный ввод |
| AATACTCASSAATAGСТААТАСГАТСГАТСГААТСГААГСТТАСТАГСТАСГАТСГАТАТАСАСГАТСГААТ АСАТ |
| Стандартный вывод |
| HindIII |

Пример №5

| |
|--|
| Стандартный ввод |
| AATACTCASSAATAGСТААТАСГАТСГАТСГААТСГААГСТТАСТАГСТАСГАТСГАТАТАСАСГАТСГААТ AGGATCCCAT |
| Стандартный вывод |
| BamHI HindIII |

Пример №6

| |
|---|
| Стандартный ввод |
| АТСАСТАГСТАСГАТСГАТСГАТСГАСТАГСТАСГАТСГАТСГАТСГАТ |
| Стандартный вывод |
| False |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 b = {'XhoI': "CTCGAG",
2     'NdeI': "CATATG",
3     'SpeI': "ACTAGT",
4     'EcoRI': "GAATTC",
5     'HindIII': "AAGCTT",
6     'BamHI': "GGATCC"}
7 a = input()
8 if x := " ".join(sorted([x for x in b if b[x] in a])):
9     print(x)
10 else:
11     print('False')
```

Секвенирование

Метод секвенирования по Сэнгеру был предложен в 1977 году Нобелевским лауреатом по химии Фредриком Сэнгером. Методы высокопроизводительного секвенирования (Next Generation Sequencing — NGS) появились намного позже, в начале 21 века. Сегодня наиболее используемой платформой NGS является секвенирование на платформе Illumina, именно этот инструмент широко использовал Нобелевский лауреат 2022 по физиологии и медицине Сванте Паабо и его коллеги для работы с древней ДНК.

Список литературы

- Артем Тикунов «Метод Сэнгера» — <https://youtu.be/hBicxm20n2g>.
- Александр Пилипенко (ИЦИГ СО РАН) «О Нобелевской премии за исследование эволюции человека методами палеогенетики» — <https://youtu.be/-1rsoBOEF8U>.
- Биомолекула: «Секвенирование нуклеиновых кислот» — <https://biomolecula.ru/articles/metody-v-kartinkakh-sekvenirovanie-nukleinovykh-kisl-ot>.

Задача IV.2.2.7. Метод Сэнгера и NGS (2 балла)

Условие

Соотнесите варианты ответов, соответствующие NGS и секвенированию по Сэнгеру.

| | Метод Сэнгера | NGS |
|---|---------------|-----|
| Секвенированию всегда предшествует стадия приготовления библиотек | | |
| Для амплификации фрагмента/ов ДНК используют ПЦР | | |
| Использование смеси дНТФ и ддНТФ на одной из стадий для обрыва цепи | | |

| | Метод Сэнгера | NGS |
|--|---------------|-----|
| Процесс определения последовательности ДНК в реальном времени связан с работой ДНК-полимеразы (sequencing by synthesis) | | |
| Возможно определение последовательности одного фрагмента ДНК одновременно в двух направлениях | | |
| При секвенировании используют чипы | | |
| При секвенировании используют капиллярный электрофорез | | |
| Данный метод был использован в проекте «Геном человека» | | |
| Возможно определение последовательности фрагмента ДНК длиной до 600-1000 п. н. | | |
| Возможно одновременное определение последовательности фрагментов ДНК длиной много тысяч и десятков тысяч пар нуклеотидов | | |
| Для обработки результатов секвенирования используют сложные биоинформатические инструменты | | |
| Перед секвенированием к фрагменту ДНК «пришивают» адаптеры | | |
| Данный метод использовали для идентификации ранее неизвестного вируса SARS-CoV-2 в январе 2020 года | | |

Ответ.

| | Метод Сэнгера | NGS |
|--|---------------|-----|
| Секвенированию всегда предшествует стадия приготовления библиотек | | + |
| Для амплификации фрагмента/ов ДНК используют ПЦР | + | + |
| Использование смеси дНТФ и ддНТФ на одной из стадий для обрыва цепи | + | |
| Процесс определения последовательности ДНК в реальном времени связан с работой ДНК-полимеразы (sequencing by synthesis) | | + |
| Возможно определение последовательности одного фрагмента ДНК одновременно в двух направлениях | | + |
| При секвенировании используют чипы | | + |
| При секвенировании используют капиллярный электрофорез | + | |
| Данный метод был использован в проекте «Геном человека» | + | |
| Возможно определение последовательности фрагмента ДНК длиной до 600-1000 п. н. | + | |
| Возможно одновременное определение последовательности фрагментов ДНК длиной много тысяч и десятков тысяч пар нуклеотидов | | + |
| Для обработки результатов секвенирования используют сложные биоинформатические инструменты | | + |
| Перед секвенированием к фрагменту ДНК «пришивают» адаптеры | | + |
| Данный метод использовали для идентификации ранее неизвестного вируса SARS-CoV-2 в январе 2020 года | | + |

Геномное редактирование

Ген RB1, которому были посвящены задания по биоинформатике, является онкосупрессором. Инактивирующие (приводящие к потере функции) мутации в этом гене являются причиной возникновения рака. Но существуют и другая группа генов – онкогены. Это гены, активирующие мутации в которых способствуют появлению злокачественных опухолей. Нокаут или «выключение» таких генов, является перспективным методом лечения рака. Также нокаут потенциальных онкогенов при помощи системы CRISPR/Cas9 может использоваться для изучения их функции и влияния на развитие опухоли.

Один из потенциальных онкогенов – ген UBR5. Механизм, по которому UBR5 влияет на инициацию и прогрессирование рака еще не до конца изучен. Однако, в различных исследованиях показано, что этот ген часто амплифицируется в опухолях яичников и молочной железы. Также высокая экспрессия гена UBR5 связана с плохим прогнозом по выживаемости для пациентов с раком молочной железы. Таким образом, нокаут этого гена на клеточной линии рака молочной железы (4T1) позволит проверить гипотезу об участии этого гена в развитии и прогрессировании опухоли.

В следующих заданиях вам необходимо изучить систему для нокаута этого гена при помощи РНК-направленной эндонуклеазы Cas9 и ответить на вопросы.

Ссылки для скачивания последовательностей векторов

1. psPAX2 — <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/324820/6f428005-0d8c-4978-bf23-880e8a636c97/addgene-plasmid-12260-sequence-324820.gbk>.
2. pMD2.G — <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/352465/54963f23-ba10-4ee9-ac0e-4352722ea663/addgene-plasmid-12259-sequence-352465.gbk>.
3. lentiCRISPRv2 — <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/244694/51ae3e01-5e94-49b8-a4fb-7fa303a7eb81/addgene-plasmid-52961-sequence-244694.gbk>.

Список литературы

- Shearer RF, Iconomou M, Watts CK, Saunders DN. Functional Roles of the E3 Ubiquitin Ligase UBR5 in Cancer. Mol Cancer Res. 2015 Dec;13(12):1523-32. doi: 10.1158/1541-7786.MCR-15-0383. Epub 2015 Oct 13. PMID: 26464214. — <https://pubmed.ncbi.nlm.nih.gov/26464214/>.
- Юрий Тарасов. Lentивирусные векторы: как они стали лучшими векторами для терапии ex vivo. Биомолекула — <https://biomolecula.ru/articles/lentivirusnye-vektory-kak-oni-stali-luchshimi-vektorami-dlia-terapii-ex-vivo>.

Задача IV.2.2.8. Доставка системы геномного редактирования в клетки (2 балла)

Условие

Для доставки системы геномного редактирования в клетки будет использована лентивирусная система, состоящая из трех векторов — psPAX2, рMD2.G и lentiCRISPRv2. Скачайте последовательности этих векторов, изучите их в программе UGENE и ответьте на вопросы.

Вектор psPAX2 является (оболочечной/упаковочной/трансферной/содержащей Cas9 и нРНК)¹ плазмидой и кодирует белки, необходимые для (сборки вириона/проникновения в клетку/редактирования генов/внедрения трансгена в геном)². Вектор рMD2.G является (оболочечной/упаковочной/трансферной/содержащей Cas9 и нРНК)³ плазмидой и кодирует белки, необходимые для (сборки вириона/проникновения в клетку/редактирования генов/внедрения трансгена в геном)⁴. Вектор LentiCRISPRv2 является (оболочечной/упаковочной/трансферной/содержащей Cas9 и нРНК)⁵ плазмидой и кодирует белки, необходимые для (сборки вириона/проникновения в клетку/редактирования генов/внедрения трансгена в геном)⁶.

Ответ: 1 — упаковочной, 2 — сборки вириона, 3 — оболочечной, 4 — проникновения в клетку, 5 — содержащей Cas9 и нРНК, 6 — редактирования генов.

Задача IV.2.2.9. Поиск направляющей РНК (3 балла)

Условие

В качестве мишени для редактирования был выбран следующий участок гена UBR5: 5'-GTACAGCCAGACAGATTGGAATTGGGTA AACCTGATAATAATGATGG-3'. Посчитайте, сколько всего направляющих РНК длиной не менее 20 нуклеотидов можно подобрать на этот участок (при использовании фермента SpyCas9). В ответ введите целое число.

Ответ: 4.

Задача IV.2.2.10. Изучение участка для редактирования (3 балла)

Условие

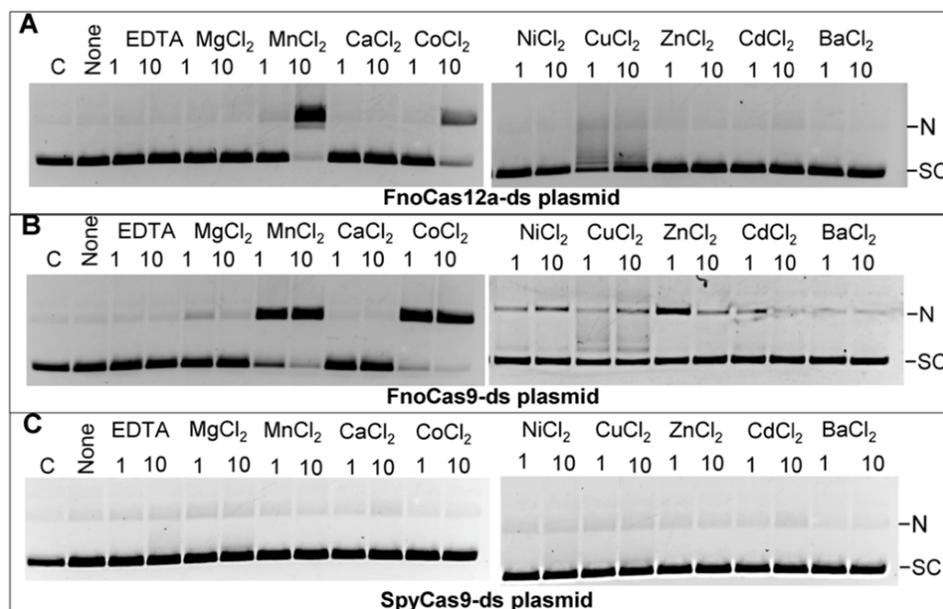
Изучите таргетный участок гена UBR5, в который будет внесена мутация, и выберите верные суждения.

Этот участок находится на экзоне номер _____¹ с идентификатором (ENSE00002109064/ENSE00000981200/ENSE00000981202/ENSE00003660785/ENSE00000981205/ENSE00000981207)² и затрагивает домен в белке (цинковые пальцы/полиаденилат-связывающий домен/HECT домен/ни один из вышеперечисленных)³.

Ответ: 1 — 4, 2 — ENSE00003660785, 3 — ни один из вышеперечисленных.

Задача IV.2.2.11. Активность белков Cas (3 балла)

Без направляющих РНК белки из системы CRISPR/Cas могут неспецифически расщеплять ДНК. Влияние различных ионов двухвалентных ионов металлов на расщепление суперскрученной двухцепочечной ДНК с помощью Cas12a и Cas9 из *Francisella tularensis novicida* (FnoCas12a и FnoCas9), Cas9 из *Streptococcus pyogenes* (SpyCas9) показано на рисунке.



Обозначения

- С — контрольный образец, без белка Cas.
- EDTA — этилендиаминтетрауксусная кислота.
- None — реакция с белком без добавления ионов металла.
- N — релаксированная форма плазмиды.
- SC — суперскрученная плазида.
- 1 и 10 — концентрация солей, мМ.

Условие

Проанализируйте результаты и ответьте на вопросы.

Какой белок (FnoCas12a, FnoCas9 или SpyCas9) разрезает плазмиду в присутствии всех исследуемых двухвалентных ионов? (*FnoCas12a/FnoCas9/SpyCas9/никакой*)¹

Какой белок (FnoCas12a, FnoCas9 или SpyCas9) разрезает суперскрученную плазмиду в присутствии Mg²⁺? (*FnoCas9/FnoCas12a/SpyCas9/никакой*)²

Катион какого элемента разрезает ДНК при концентрации 10 мМ, но не при 1 мМ? (*Ca²⁺/Co²⁺/Mg²⁺/Mn²⁺*)³

Ответ: 1 — никакой, 2 — FnoCas9, 3 — Co²⁺.

Задача IV.2.2.12. Ортологи Cas (3 балла)

Механизм действия Cas9 начинается с распознавания мотива, прилегающего к протоспейсеру (PAM). Cas9 из *Streptococcus pyogenes* (SpyCas9) широко используют для редактирования генома отчасти потому, что он распознает короткую последовательность PAM — 5'-NGG-3' (N — любой нуклеотид). Ортологи Cas9 демонстрируют значительное разнообразие в PAM последовательностях, например Cas9 из *Campylobacter jejuni* (CjeCas9) распознает 5'-NNNNRYAC-3' (R — A или G, Y = C или T), а Cas9 из *Geobacillus stearothermophilus* (GeoCas9) распознает 5'-NNNNCRAA-3' и 5'-NNNNGMAA-3' (M — A или C). Ниже представлена одна из цепей фрагмента ДНК. 5'-AACTGGATCTACACAGCGGTACGATCCTTGCGAGTTTTTCGCGCCGAAATACGTTTTTCG-3'.

Условие

1. Сколько PAM сайтов для CjeCas9 можно найти в фрагменте двуцепочечной ДНК?
2. Сколько PAM сайтов для GeoCas9 можно найти в фрагменте двуцепочечной ДНК?

Ответ: 1 — 4, 2 — 7.

Третий блок задач

Задача IV.2.3.1. Поиск ортологов гена UBR5 у приматов (4 балла)

Для исследования гена UBR5, которому были посвящены задания по геномному редактированию во втором блоке, в качестве модельных организмов можно использовать приматов. Но для того, чтобы изучать этот ген на приматах, сначала необходимо найти его ортологи у этой группы организмов. Поиск ортологов можно выполнить при помощи сервиса BLAST в NCBI.

Условие

Используя транскрипт с идентификатором NM_001282873.2 гена UBR5, выполните поиск при помощи алгоритма blastn по базе данных refseq_gna. Выполните поиск в таксоне приматы (taxid:9443), исключив при этом человека (taxid:9606). В поле максимальное количество таргетных последовательностей выберите 100. За совпадение нуклеотидов +2 очка, за несовпадение -3. За открытие гена штраф -5, за расширение -2.

Больше всего ортологов нашлось для вида (*Theropithecus gelada/Mandrillus leucophaeus/Macaca mulatta/Pan troglodytes/Pan paniscus/Gorilla gorilla gorilla*)¹. Идентификатор последовательности с наименьшим e-value (XM_032760159.1/XM_016959749.1/XM_011731652.2/XM_032760156.1/XM_025393634.1/XM_050800629.1)². Идентификатор последовательности, имеющей наибольший процент идентичных нуклеотидов (XM_034966444.1/XM_016959749.1/XM_034966441.1/XM_008975199.3/

XM_025393634.1/XM_032760159.1)³. Эта последовательность, имеющая наибольший процент идентичных нуклеотидов, относится к виду (*Lemur catta*/*Pan paniscus*/*Pan troglodytes*/*Carlito syrichta*/*Cebus imitator*/*Pongo abelii*)⁴.

Ответ: 1 — *Pan troglodytes*, 2 — XM_025393634.1, 3 — XM_016959749.1, 4 — *Pan troglodytes*.

Задача IV.2.3.2. Поиск ортологов гена UBR5 у растений (6 баллов)

Вряд ли растительные модели подойдут для изучения влияния онкогена человека на развитие онкологического заболевания. Однако, поиск ортологов гена у растений позволит более детально изучить этот ген. Например, найти в нем эволюционно-консервативные участки.

Условие

Используя белок, который кодируется транскриптом NM_001282873.2 гена UBR5, выполните поиск при помощи алгоритма psi-blast по базе данных nr в высших растениях (taxid: 3193).

В качестве матрицы замен выберите наиболее подходящую из предложенных: PAM30, PAM70, PAM250. В поле максимальное количество таргетных последовательностей выберите 500. За открытие гэпа штраф −14, за расширение −2.

Для BLAST при таких условиях наиболее подходит матрица замен (*PAM30/PAM70/PAM250*)¹. Среди найденных последовательностей, к классу однодольных растений относят _____², к роду *Arabidopsis* _____³ последовательности(-тей). Выравниваниями покрыт(а,о) (*вся последовательность белка UBR5/начало белка/середина белка/конец белка*)⁴, и это означает, что наиболее эволюционно-консервативный участок находится (*на N-конце белка/в середине белка/на C-конце белка*)⁵. Этот домен является (*убиквитин-протеин лигазным доменом/кадгериновым доменом/доменом цинковые пальцы/AP2/ERF доменом*)⁶.

Ответ: 1 — PAM250, 2 — 145, 3 — 3, 4 — конец белка, 5 — на C-конце белка, 6 — убиквитин-протеин лигазным доменом.

Задача IV.2.3.3. Поиск самой длинной открытой рамки считывания (3 балла)

Открытой рамкой считывания (ORF) называют последовательность ДНК, способную кодировать белок. Открытая рамка считывания всегда начинается со стартового кодона и заканчивается стоп-кодоном.

Условие

Напишите программу, которая будет находить все возможные открытые рамки считывания в последовательности и выводить на экран самую длинную из них. На вход подается последовательность ДНК в направлении 5′-3′. На выходе должна

быть последовательность самой длинной открытой рамки считывания, также в направлении 5'-3'. Если открытых рамок считывания нет, выведите на экран False. Используйте стандартные старт и стоп кодоны, поиск открытых рамок считывания производите только в прямом направлении (5'-3').

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--|
| Стандартный ввод |
| АТАТСТАТСТГАТСАТАТГАССГСАТАСГГАГТАССГТГААТАТАСАСАТ |
| Стандартный вывод |
| АТГАССГСАТАСГГАГТАССГТГА |

Пример №2

| |
|--|
| Стандартный ввод |
| АГСССГАСССАТАТГСССАТСАТТТТАСААТАТГСССССАТААТТАААССАТТГАСАТСА |
| Стандартный вывод |
| АТГСССАТСАТТТТАСААТАТГСССССАТААТТАА |

Пример №3

| |
|------------------------------|
| Стандартный ввод |
| АТТТАТСАСАТАААСАТГСГАТААГТСА |
| Стандартный вывод |
| False |

Пример №4

| |
|--------------------------|
| Стандартный ввод |
| АТГАСАСАСССАТАССАТГА |
| Стандартный вывод |
| АТГАСАСАСССАТАССАТГА |

Пример №5

| |
|--|
| Стандартный ввод |
| ГСАТАТСААТГАТТГСАТСАТГАССГТАААСААТТАСГТААТЦГГСАС |
| Стандартный вывод |
| АТГАТТГСАТСАТГАССГТАААСААТТАСГТАА |

Пример №6

| |
|------------------------------|
| Стандартный ввод |
| АТТТТТТТТТТТСССССГГГГГГГАААА |
| Стандартный вывод |
| False |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 dna = input()
2 import re
3 start = 'ATG'
4 stop = r'(TAA)|(TAG)|(TGA)'
5
6 def find_longest(read):
7     for i in range(len(read)):
8         if read[i] == start:
9             for j in range(len(read) - 1, i, -1):
10                if re.match(stop, read[j]):
11                    ans = ''.join(read[i:j+1])
12                    return ans
13     return ''
14
15 fin_ans = ''
16 for x in range(3):
17     mid_ans = find_longest([dna[i:i+3] for i in range(x, len(dna), 3)])
18     if len(mid_ans) > len(fin_ans):
19         fin_ans = mid_ans
20
21 if fin_ans:
22     print(fin_ans)
23 else:
24     print(False)
```

Задача IV.2.3.4. Трансляция открытой рамки считывания (3 балла)

Условие

Напишите программу, на вход которой подается последовательность открытой рамки считывания (в направлении 5'-3', от старт-кодона до стоп-кодона включительно). Эту последовательность необходимо транслировать в белок (при помощи стандартного генетического кода) и вывести на экран последовательность белка (в направлении от N-конца к C-концу).

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--------------------------|
| Стандартный ввод |
| ATGACCGCATACGGAGTACCGTGA |
| Стандартный вывод |
| MTAYGVP |

Пример №2

| |
|--------------------------------------|
| Стандартный ввод |
| ATGCCCATCATTTTTACAATATGCCCCCATAATТАА |
| Стандартный вывод |
| МРІІFTICPHN |

Пример №3

| |
|--------------------------|
| Стандартный ввод |
| ATGACACACCCCATACCATGA |
| Стандартный вывод |
| МТНPIР |

Пример №4

| |
|------------------------------------|
| Стандартный ввод |
| ATGATTCGCATCATGACCGTAAACAATTACGТАА |
| Стандартный вывод |
| MIRIMTVTIT |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 seq = input()
2 d = {'TTT': 'F', 'TTC': 'F', 'TTA': 'L',
3     'TTG': 'L', 'CTT': 'L', 'CTC': 'L',
4     'CTA': 'L', 'CTG': 'L', 'ATT': 'I',
5     'ATC': 'I', 'ATA': 'I', 'ATG': 'M',
6     'GTT': 'V', 'GTC': 'V', 'GTA': 'V',
7     'GTG': 'V', 'TCT': 'S', 'TCC': 'S',
8     'TCA': 'S', 'TCG': 'S', 'CCT': 'P',
9     'CCC': 'P', 'CCA': 'P', 'CCG': 'P',
10    'ACT': 'T', 'ACC': 'T', 'ACA': 'T',
11    'ACG': 'T', 'GCT': 'A', 'GCC': 'A',
12    'GCA': 'A', 'GCG': 'A', 'TAT': 'Y',
13    'TAC': 'Y', 'CAT': 'H', 'CAC': 'H',
14    'CAA': 'Q', 'CAG': 'Q', 'AAT': 'N',
15    'AAC': 'N', 'AAA': 'K', 'AAG': 'K',
16    'GAT': 'D', 'GAC': 'D', 'GAA': 'E',
17    'GAG': 'E', 'TGT': 'C', 'TGC': 'C',
18    'TGG': 'W', 'CGT': 'R', 'CGC': 'R',
19    'CGA': 'R', 'CGG': 'R', 'AGT': 'S',
20    'AGC': 'S', 'AGA': 'R', 'AGG': 'R',
21    'GGT': 'G', 'GGC': 'G', 'GGA': 'G',
22    'GGG': 'G'
23 }
24 stop_lst = ['TAA', 'TAG', 'TGA']
25 list_ams = []
26 for i in range(0, len(seq), 3):
27     codon = seq[i:i+3]
```

```

28     if codon not in stop_lst:
29         am = d.get(codon)
30         list_ams.append(am)
31 p_seq = ''.join(list_ams)
32 print(p_seq)

```

Задача IV.2.3.5. Поиск делеций в выравнивании (3 балла)

Допустим, вам часто приходится анализировать парное выравнивание последовательностей и находить в нем делеции. Делеции обозначаются пропусками (символ -) в выравниваемой последовательности.

Пример: CAGA--ATACA. Данная последовательность имеет делецию длиной 4 нуклеотида, с 5 по 8 позицию в последовательности. Эту делецию можно записать в следующем виде: del5-8.

Условие

Напишите программу, которая будет принимать на вход последовательность после выравнивания. Эта последовательность может содержать (а может и не содержать) делецию. Делеция обозначается знаком минус (-). Найдите координаты начала и конца этой делеции и выведите их на экран в следующем виде:

del [координата_начала] - [координата_конца].

Если делеция имеет длину 1 нуклеотид, то выведите ее на экран в следующем виде:

del [координата_делеции].

Если последовательность не содержит делеций, выведите False.

Тестирование будет проводиться на языке Python версии 3.10.

Примеры

Пример №1

| |
|--------------------------|
| Стандартный ввод |
| CAGA----ATACA |
| Стандартный вывод |
| del5-8 |

Пример №2

| |
|--------------------------|
| Стандартный ввод |
| ATATGAG-ACATA |
| Стандартный вывод |
| del8 |

Пример №3

| |
|--------------------------------------|
| Стандартный ввод |
| ACATAGTACGATCGTGCTAGCTAAATGCATCGATCG |
| Стандартный вывод |
| False |

Пример №4

| |
|---|
| Стандартный ввод |
| AC---ATAGTACGATCGTGCTAGCTAAATGCATCGATCG |
| Стандартный вывод |
| del3-5 |

Пример №5

| |
|-----------------------------------|
| Стандартный ввод |
| TAGTACGATCGAACATATATGCTAGCTAAATGC |
| Стандартный вывод |
| False |

Пример №6

| |
|-----------------------------------|
| Стандартный ввод |
| C----ATAGTACGATCGTGCTAGCTAAATGCAT |
| Стандартный вывод |
| del2-5 |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 import re
2 def find_del(seq):
3     pat = '-+'
4     return re.search(pat, seq)
5 ans = find_del(input())
6
7 if not ans:
8     print(False)
9 elif ans.start() != ans.end() - 1:
10    print('del' + str(ans.start()+ 1) + '-' + str(ans.end()))
11 else:
12    print('del' + str(ans.end()))
```

Задача IV.2.3.6. Поиск frameshift мутации (3 балла)

Результат успешного нокаута гена при помощи системы геномного редактирования CRISPR/Cas9 — это внесение мутации, нарушающей рамку считывания (frameshift).

Условие

Напишите функцию `is_frameshift`, которая будет принимать на вход два аргумента: референсную последовательность ДНК и последовательность ДНК мутанта, выровненную на референс (может содержать пропуски, обозначенные символом `-`). Функция должна проверить, содержит ли последовательность мутацию, нарушающую рамку считывания и вернуть `True`, если содержит и `False`, если не содержит.

Тестирование будет проводиться на языке *Python* версии 3.10. Функцию необходимо только создать (название обязательно должно быть таким, как в условии!). Вызывать ее или печатать что-либо на экран не нужно. Аргументы, перечисленные в *Input*, будут переданы в функцию автоматически при проверке.

Примеры

Пример №1

| |
|-----------------------------|
| Стандартный ввод |
| CAGAACTAATACA CAGA----ATACA |
| Стандартный вывод |
| True |

Пример №2

| |
|---------------------------------|
| Стандартный ввод |
| ATATGAGTTCACATA ATATGAG---ACATA |
| Стандартный вывод |
| False |

Пример №3

| |
|-------------------------------|
| Стандартный ввод |
| ATAGTGTСATСATС ATAGTGTСATСATС |
| Стандартный вывод |
| False |

Пример №4

| |
|--|
| Стандартный ввод |
| ATATCACCSCSSCATATACAC ATATCAC-----ATATACAC |
| Стандартный вывод |
| False |

Пример №5

| |
|---|
| Стандартный ввод |
| TATATACACATCGATCGATCGATCG TATATACACA-CGATCGATCGATCG |
| Стандартный вывод |
| True |

Пример №6

| |
|---|
| Стандартный ввод |
| ATCGATCGATCGATCGATCGATCGAT ATCGATCGATCGATCG-----TCGAT |
| Стандартный вывод |
| True |

Пример программы-решения

Ниже представлено решение на языке Python 3.

```
1 import re
2 def is_frameshift(seq, mut):
3     pat = '-+'
4     delet = re.search(pat, mut)
5     if not delet:
6         return False
7     if (delet.end() - delet.start()) % 3 == 0:
8         return False
9     else:
10        return True
```

Геномное редактирование

Во втором блоке вы решали задачи, посвященные использованию системы для но-каута гена UBR5 на клеточных линиях человека: как доставить систему геномного редактирования в клетку при помощи системы CRISPR/Cas9, что из себя представляет участок для редактирования и сколько всего на него можно подобрать направляющих РНК.

В третьем блоке заданий вам предлагается выбрать наиболее подходящую нРНК для внесения двуцепочечного разрыва в этот участок ДНК, выполнить *in silico* клонирование этой нРНК в плазмиду и подобрать рестриктазы для анализа результатов клонирования.

Ссылки для скачивания последовательностей векторов

1. lentiCRISPRv2 — <https://media.addgene.org/snapgene-media/v1.7.9-0-g88a3305/sequences/244694/51ae3e01-5e94-49b8-a4fb-7fa303a7eb81/addgene-plasmid-52961-sequence-244694.gbk>.

Последовательность каркаса нРНК (*sgRNA scaffold*)

5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT
GAAAAAGTGGCACCGAGTCGGTGTCTTTT-3

Инструменты для дизайна нРНК

1. CCTop: <https://cctop.cos.uni-heidelberg.de:8043/index.html>.
2. RNAfold: <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>.

Задача IV.2.3.7. Дизайн направляющей РНК (2 балла)

В качестве мишени для редактирования был выбран следующий участок гена UBR5 (его вы изучили в предыдущем блоке): 5'-GTACAGCCAGACAGATTGGAATTGGGTTAAACCTGATAATAATGATGG-3.

Условие

Подберите наиболее подходящий таргетный участок для нРНК, используя следующие критерии (приведены в порядке убывания значимости):

- отсутствие офф-таргетов (не должно быть комплементарности с участком экзона другого гена с количеством мисметчей < 5);
- правильная вторичная структура (три шпильки, как можно более свободный 5'-конец);
- как можно более высокая предсказанная эффективность.

Самый подходящий таргетный участок для нРНК имеет последовательность (введите последовательность в направлении 5'-3', без РАМ): ____¹. Для того, чтобы использовать данную систему для геномного редактирования с вектором lentiCRISPRv2, необходимо заказать в лаборатории синтез олигонуклеотида, который будет содержать (таргетный участок + РАМ/таргетный участок без РАМ/таргетный участок и каркас нРНК (sgRNA scaffold))².

Ответ: 1 — TTACCCAAATTCCTGTC, 2 — таргетный участок без РАМ.

Задача IV.2.3.8. Клонирование олигонуклеотида в вектор (2 балла)

Для клонирования полученного на предыдущем шаге олигонуклеотида в вектор lentiCRISPRv2 будет использоваться рестриктаза Esp3I. Вектор lentiCRISPRv2 необходимо обработать рестриктазой Esp3I и выделить более длинный фрагмент, который соответствует линеаризованному вектору с липкими концами.

К олигонуклеотиду необходимо на этапе синтеза добавить липкие концы, которые будут комплементарны липким концам, образовавшимся в векторе lentiCRISPRv2 после обработки рестриктазой Esp3I.

Условие

Выполните описанные выше процедуры *in silico*. Откройте вектор lentiCRISPRv2 в программе UGENE и выполните *in silico* рестрикцию с рестриктазой Esp3I. Определите, какие липкие концы образуются в результате рестрикции.

Создайте файл с последовательностью олигонуклеотида, полученного на предыдущем шаге, и откройте его в программе UGENE. Создайте из этой последовательности фрагмент с необходимыми липкими концами. Выполните лигирование полученных фрагментов и введите в поле ответ длину получившегося вектора.

Ответ: 13012.

Задача IV.2.3.9. Рестрикционный анализ плазмиды со вставкой (2 балла)

Условие

Для проверки наличия вставки в плазмиде можно выполнить рестрикционный анализ. Для этого плазмиду обрабатывают парой рестриктаз и наносят результат рестрикции на агарозный гель. В случае наличия вставки образуются другие продукты рестрикции, чем в случае исходной плазмиды и их можно отличить по длине продукта на электрофорезе.

Из предложенных рестриктаз (или пар рестриктаз) выберите те, которые позволят отличить плазмиду со вставкой от исходной плазмиды на агарозном геле.

1. Esp3I.
2. EcoRI + XhoI.
3. HindIII.
4. NdeI + EcoRV.
5. PciSI + PvuI.
6. SpeI.

Ответ: 1, 3, 4, 5.

Задача IV.2.3.10. Анализ результатов эксперимента (2 балла)

Целью всего эксперимента по нокауту гена UBR5 было изучить, влияет ли этот ген на развитии и прогрессировании опухоли. Предположим, что все этапы эксперимента успешно завершены: выполнен дизайн нРНК, вставка клонирована в вектор, клеточную линию рака молочной железы (4T1/wt) трансформировали вектором и получили новую клеточную линию с нокаутированным геном UBR5 (4T1/ubr5-).

Чтобы подтвердить гипотезу о том, что эта клеточная линия будет меньше метастазировать и медленнее прогрессировать, эти клетки были подсажены мышши линии BALB/c и спустя 28 дней был произведен измерения выросшей опухоли для клеточной линии с функциональным геном UBR5 (4T1/wt) и для клеточной линии с нокаутированным геном UBR5 (4T1/ubr5-). Были получены следующие результаты:

- 4T1/wt: 1231 мм³, 1051 мм³, 987 мм³, 1317 мм³, 1156 мм³;
- 4T1/ubr5-: 371 мм³, 154 мм³, 216 мм³, 21 мм³, 429 мм³.

Условие

Проверьте гипотезу о том, что нокаут гена привел к уменьшению размеров опухоли при помощи t-теста Уэлша (Welch's t-test).

Значения t-критерия для данного эксперимента равно _____¹ соответствующее ему значение p-value > 0,05/< 0,05/< 0,01/< 0,01/< 0,001². Это значит, что (мы можем отвергнуть нулевую гипотезу о том, что нокаут гена UBR5 не влияет на размер опухоли (при уровне значимости $p < 0,05$)/мы не можем отвергнуть нулевую гипотезу о том, что нокаут гена UBR5 не влияет на размер опухоли (при

уровне значимости $p < 0,05$)³.

Ответ: 1 — 9,6, 2 — $< 0,001$, 3 — мы можем отвергнуть нулевую гипотезу о том, что нокаут гена UBR5 не влияет на размер опухоли (при уровне значимости $p < 0,05$).

Задача IV.2.3.11. dCas9 и его более короткие версии (5 баллов)

В недавнем исследовании ученые задались вопросом: можно ли уменьшить dCas9, сохранив одну из его функций — связывание с ДНК посредством sgРНК? В результате был получен белок Δ dCas9, который уменьшили на 36%, и он был способен связываться с ДНК в экспериментах *in vitro* и *in vivo*. В этом задании вам предлагается исследовать делеции в варианте Δ dCas9.



На рисунке представлена схема конструкций, содержащие dCas9 («мертвый» Cas9 длиной 1368 а. о.) из *Streptococcus pyogenes* и его вариант Δ dCas9 с делециями (длина 874 а. о.).

Условие

Определите, в каких доменах, какого размера и сколько делеций было введено в вариант Δ dCas9. Учитывайте, что для внесения делеций использовали метод, с помощью которого дополнительно вставлялись последовательности ДНК, кодирующие два аминокислотных остатка — XS, где X — это неизвестный а. о., S — серин.

Список доменов dCas9 и их частей: RuvC-I, BH, REC1-I, REC2, REC1-II, REC3, RuvC-II, HNH, RuvC-III, CTD.

Размер делеций рассчитывайте с учетом вставки двух а. о., например:

AAAAAAAA--XSAAAAAAAA Δ dCas9.

AAAAAAAAEEEEEEAAAAAAAA dCas9.

Размер делеции: 4.

Аминокислотная последовательность конструкции 6xHis-TEV site-4x SV40 NLS- Δ dCas9-2x SV40 NLS:

```
MKSSHHHHHHNENLYFQSNATPKKKRKVGGSPKKRKVGGSPKKRKVGGSPKKRKVGIHGVPAAATMDKKYSIG
LAIGTNSVGVAVITDEYKVP SKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRKRNRI CYL
QEIFSNEMAKVDDSFHRL EESFLVEEDKKHERHPIFGNIVDEVAUYHEKYPTIYHLRKKLVDSTDKADRLRIYL
ALAHMIKFRGHFLIEGDLNPDNSTDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEK
YKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
MTNFDKNLTSQKAQVSGQGDSLHEHIANLAGSPA IKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQK
GQKNSRERMKRIEEGIKELASDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI
REVKVITLKSKLVSDFRKDFQFYK VREINNYHHAHDAYLNAVVG TALIKKYPKLESEFVASTVRKVL SMPQVNI
VKKTEVQTTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEK GKSKKLKSVKELLGITI
```

MERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLYLASHY
EKLKGSPEDENEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLNDKVL SAYNKHDKPIREQAENIIHLFTLT
NLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSIITGLYETRIDLSQLGGDGSPPKKRKKVEDPKKKRQVDTQS
PKGS

Введите число делеций: _____¹. Размер самой большой делеции (введите число) — _____² а. о. Размер самой маленькой делеции (введите число) — _____³ а. о. В домене (*RuvC-I/BH/REC1-I/REC2/REC1-II/REC3/RuvC-II/HNH/RuvC-III/CTD*)⁴ есть делеция длиннее самой маленькой делеции, но короче 110 а. о. В домене (*RuvC-I/BH/REC1-I/REC2/REC1-II/REC3/RuvC-II/HNH/RuvC-III/CTD*)⁵ есть делеция длиннее 110 а. о., но короче самой большой делеции. Во второй делеции X (в составе XS) соответствует аминокислота (введите одну заглавную латинскую букву) — _____⁶. В первой делеции X в составе XS соответствует аминокислота — _____⁷. Какая из двух мутаций, характерных для dCas9, сохранилась в ΔdCas9? Ответ запишите в виде A*B, где A — остаток аминокислоты в Cas9, * — порядковый номер аминокислоты, B — остаток аминокислоты в dCas9 (используйте латинские символы в верхнем регистре) — _____⁸.

Ответ: 1 — 4, 2 — 204, 3 — 70, 4 — HNH, 5 — REC2, 6 — A, 7 — T (thr, треонин, тре, threonine), 8 — D10A.

Задача IV.2.3.12. Знакомство с PyMol (5 баллов)

Условие

Проанализируйте sgРНК (38 шт) из 33 различных структур в комплексе с Cas9. Сделайте выравнивание в PyMOL с помощью функции `extra_fit`. Разделите на три группы sgРНК на основании различий в выравнивании трехмерной структуры спейсера. Заполните пропуски.

NB! Названия файлов (например, 4zt0_B) состоят из четырехзначного PDB ID (например, 4zt0), и название цепи из структуры комплекса, которая кодирует sgРНК (например, B): 4oo8_B, 4oo8_E, 4un3_A, 4un4_A, 4un5_A, 4zt0_B, 4zt0_D, 4zt9_B, 4zt9_D, 5b2r_A, 5b2s_A, 5b2t_A, 5f9r_A, 5fq5_A, 5fw1_A, 5fw2_A, 5fw3_A, 5vzl_B, 5vw1_B, 5xbl_B, 5y36_B, 6aeb_A, 6aeb_E, 6aeg_A, 6ai6_B, 6ifo_C, 6ifo_D, 6k3z_A, 6k4p_A, 6k4q_A, 6k57_A, 6k4s_A, 6k4u_A, 6mcb_B, 6mcc_B, 6o0x_B, 6o0y_B, 6o0z_B.

Список литературы

- Protein Data Bank — <https://www.rcsb.org>.
- PyMOL — <https://pymol.org/2/>.
- Ссылка на папку с файлами — <https://disk.yandex.ru/d/yUmbNx3bsQcECA>.
Данная папка также содержит ключ для регистрации учебной версии PyMOL.

Введите число структур в самой маленькой группе — _____¹. Введите число структур в самой большой группе — _____². С помощью криоэлектронной микроскопии получено _____³ структур. Введите число структур, которые содержат удлиненную последовательность повтор-антиповтор — _____⁴. Введите число структур, в которых не удалена шпилька с 3'-конца — _____⁵. Что объединяет

группу со средним количеством sgРНК? (*Наличие ДНК в структуре с комплексом с Cas9/Отсутствие ДНК в структуре с комплексом с Cas9/Удлиненная последовательность повтор-антиповтор/Укороченная последовательность повтор-антиповтор/С 3'-конца удалена шпилька/С 3'-конца не удалена шпилька*)^б.

Ответ: 1 — 1, 2 — 26, 3 — 7, 4 — 4, 5 — 16, 6 — отсутствие ДНК в структуре с комплексом с Cas9.