

ЗАДАНИЯ
практического тура заключительного этапа XXXV Всероссийской
олимпиады школьников по биологии. 2018-2019 уч. год.
11 класс БИОИНФОРМАТИКА
Время выполнения задания 60 минут

Секвенирование и анализ последовательностей нуклеиновых кислот является ключевой областью современной биоинформатики. Откройте папку с номером Вашей группы на рабочем столе, откройте рисунки А-С для выполнения заданий 1-2, затем запустите на вашем рабочем компьютере программу Gene Runner и выполните с ее помощью задания 3-4.

Задание 1. Секвенирование (13 баллов).

Развитие биоинформатики в 1990-х годах и успех проекта «Геном человека» были связаны с технологией автоматического секвенирования по Сенгеру. Принцип этого метода прост и очевиден: на основе очищенной одноцепочечной матрицы ДНК с помощью ДНК-полимеразы достраивают комплементарную цепь, причем в реакционной смеси присутствуют как обычные дезоксирибонуклеотидтрифосфаты, так и нуклеотиды-терминаторы, которые содержат обычные азотистые основания, трифосфат, флуоресцентную часть и дидезоксирибозу, которая не позволяет присоединиться следующему нуклеотиду. Таким образом, с вероятностью около 0,1% растущая нить ДНК может прекратить рост в случайном месте, причем последний ее нуклеотид будет специфически флуоресцентно помечен, «цвет» флуоресценции зависит от азотистого основания. После этого проводят электрофорез, разделяя нити ДНК по длине, и на выходе из капилляра для электрофореза измеряют флуоресценцию молекул, тем самым определяя последовательность меченых нуклеотидов.

Другим методом является секвенирование путем синтеза. Одноцепочечная матрица ДНК связывается с поверхностью чипа, далее ДНК-полимераза начинает пошагово наращивать ее с помощью флуоресцентно-меченых нуклеотидов (каждый своего цвета). За один шаг присоединяется один нуклеотид, присоединению нескольких подряд мешает флуоресцентная группа. В конце шага флуоресценцию сканируют, после чего флуоресцентную группу удаляют, на следующем шаге растущая нить может удлиниться еще на один нуклеотид.

Третий современный метод – это pH-метрическое пирофосфатное секвенирование. Одноцепочечная матрица ДНК связывается с поверхностью бусины, попадающей в лунку чипа, далее ДНК-полимераза присоединяет очередной нуклеотид, отщепляя при этом пирофосфат. Пирофосфат понижает pH в лунке, что регистрируется микро-pH-метром. Отдельные нуклеотидтрифосфаты подаются последовательными волнами, включение в растущую цепь возможно только с волной комплементарного нуклеотида, однако можно присоединить несколько одинаковых нуклеотидов сразу, что понизит pH еще сильнее.

Откройте рисунки А, В и С в папке с номером вашей группы (tif-файлы в Paint, gif-файл в браузере). Подпишите на Листе Ответов методы, которыми были секвенированы последовательности на рисунках, запишите сами последовательности (учитывая задание 2). Чем дидезоксирибоза, используемая в нуклеотидах-терминаторах, отличается от рибозы?

Задание 2. Выравнивание и выбор дерева (9 баллов).

Последовательности А, В и С – это участки из середины гомологичных кодирующих последовательностей ДНК одного и того же гена, найденные у трех

штаммов вирусов. У четвертого вируса в этом месте находится последовательность D: NNTgCTTAGCCATATgANN, в которой произошла инсерция одного нуклеотида, нарушающая функцию того же самого гена (N – обозначение случайного нуклеотида). Выровняйте последовательности A, B, C и запишите их на Листе ответов одну под другой, так чтобы идентичные нуклеотиды были друг над другом и над теми же нуклеотидами в последовательности D. Если последовательности перерываются со сдвигом, в их начале и конце нужно дописать случайные нуклеотиды N (от 0 до 3). Найдите попарное число различий между выравненными последовательностями, используйте весовые коэффициенты: для транзиции 1, для трансверсии 2, для инсерции/делеции 5. Какое из показанных на Листе ответов дерево соответствует принципу максимальной парсимонии (дерево имеет наименьшее количество мутационных событий, требуемых для построения его топологии)?

Задание 3. Поиск рамки считывания (9 баллов).

Создайте в программе Gene Runner три файла с последовательностями A, B и C соответственно (комбинация кнопок [Ctrl/Alt/N](#) либо путь меню [File → New → Nucleic acid sequence](#)), набирая нуклеотиды на клавиатуре, при этом не забудьте добавить случайные нуклеотиды в начало и конец последовательности, как в задании 2. Сохраните Ваши файлы под названием X-A, X-B, X-C, где X – это номер Вашей группы (комбинация кнопок [Ctrl/Alt/A](#) либо путь меню [File → Save As](#)). Найдите в последовательностях A, B и C все 6 возможных рамок считывания (комбинация кнопок [Ctrl/M](#) и затем [Enter](#) либо путь меню [Analysis → Nucleic acid → Open Reading Frames](#) и затем [Enter](#)) и запишите их на Листе ответов. Если в рамке считывания есть стоп-кодон (обозначаются в Gene Runner знаками @, & или # красного цвета), такая рамка называется закрытой. Сравнивая рамки считывания гомологичных последовательностей A, B и C, определите, какая из них соответствует кодирующей последовательности гена, поставьте на Листе ответов рядом с ними букву «К», рядом с закрытыми рамками поставьте букву «З». Почему последовательность D стала нефункциональна? Попробуйте восстановить исходную кодируемую аминокислотную последовательность у штамма D на этом участке до инсерции. Для того чтобы изменить последовательность, разблокируйте ее для редактирования, нажав [Edit→Unlock](#).

Задание 4. Анализ вторичной структуры (9 баллов)

ДНК, также как и РНК, способна формировать различные вторичные структуры – «шпильки», «глазки» и димеры. Эти структуры могут присутствовать также в молекуле мРНК и играть важную роль в регуляции экспрессии гена, а также могут осложнять проведение полимеразной цепной реакции (искусственной репликации ДНК) из-за взаимодействий между праймерами и последующей неспецифической амплификации. Для анализа вторичных структур в Gene Runner нужно выделить последовательность и нажать комбинацию кнопок [Ctrl/L](#), либо путь меню [Analysis → Oligo...](#) Появится окошко, в котором будут показаны шпильки «Hairpin loops», димеры «Dimers», глазки «Vulge loops» и шпилечные димеры «Internal loops». Для каждой из последовательностей A-D укажите на листе ответов количество этих элементов. Представьте себе, что Вам нужно написать праймеры для специфической амплификации генов A-D (чтобы праймер подходил только для одной последовательности). На каком из концов праймера требование комплементарности и отсутствия вторичных структур будут критически важным и почему? Обозначьте стрелочками снизу под последовательностями A-D из задания 2 направление и границы специфических для них праймеров (формат 5' → 3').

УСПЕХОВ НА ДРУГИХ КАБИНЕТАХ ПРАКТИЧЕСКОГО ТУРА!

Задание 1. (по 1 баллу за каждый метод, по 3 балла за каждую последовательность)

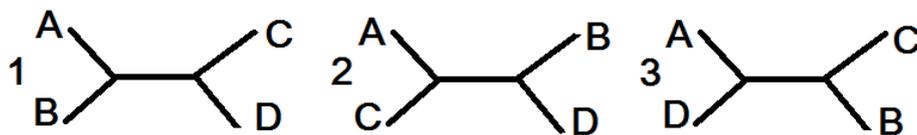
Посл.	Метод	Сиквенс (используйте буквы «A C g T», не путайте гуанин и цитозин)
A		
B		
C		

Дидезоксирибоза _____ (1 балл)

Задание 2. (по 1 баллу за каждое парное различие и дерево, различия можно подчеркивать)

A	
B	
C	
D	5'-N N T g C T T A g C C A T A T g A N N-3'

Пара	Число различий
A-B	
A-C	
A-D	
B-C	
B-D	
C-D	



Укажите число эволюционных событий для каждого дерева, обведите кружком № наиболее парсимоничного дерева.

Задание 3. (по 2 балла за 6 рамок последовательности, комплементарные рамки имеют знак «←→»)

Рамка	Последовательность A	Последовательность B	Последовательность C
Рамка 1			
Рамка 2			
Рамка 3			
Рамка -1			
Рамка -2			
Рамка -3			

Почему ген D стал нефункционален? _____ (1 балл)

Исходная последовательность аминокислот D _____ (2 балла)

Задание 4. По 1 баллу за верно перечисленные вторичные структуры каждой последовательности, по 1 баллу за каждый верно подобранный праймер.

Последовательность	Шпильки	Димеры	Глазки	Шпил. димеры
A				
B				
C				
D				

На каком из концов праймера требование комплементарности и отсутствия вторичных структур будут критически важным и почему? _____ (1 балл)